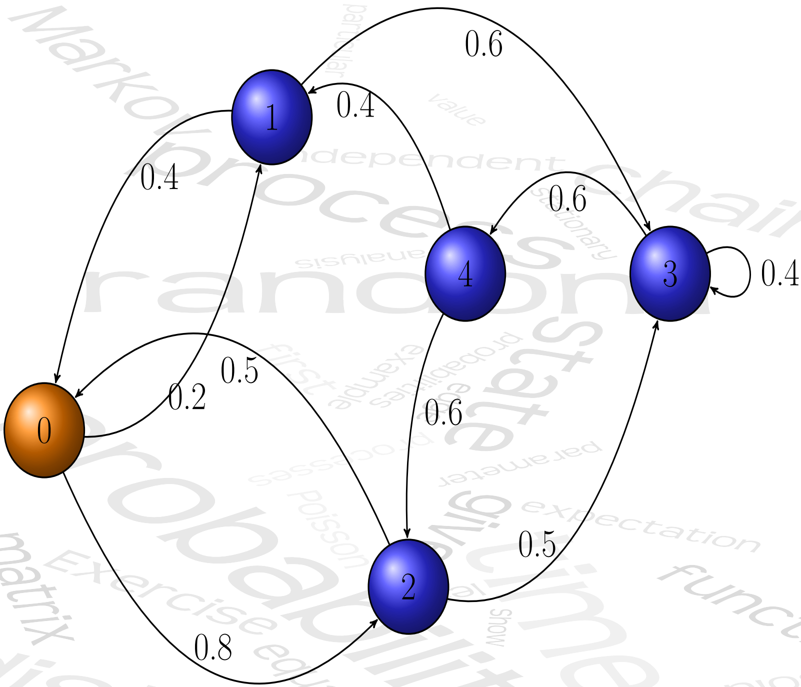
Nicolas Privault

Notes on
Markov Chains



This version: May 4, 2024
https://personal.ntu.edu.sg/nprivault/indext.html

Preface

Stochastic and Markovian modeling are of importance to many areas of science including physics, biology, engineering, as well as economics, finance, and social sciences. This text is an undergraduate-level introduction to the Markovian modeling of time-dependent randomness in discrete and continuous time, mostly on discrete state spaces. The emphasis is put on the understanding of concepts by examples and elementary derivations. Some more advanced topics on Poisson stochastic integrals are also covered.

The book is mostly self-contained except for its main prerequisites, which consist in a knowledge of basic probabilistic concepts. This includes random variables, discrete distributions (essentially binomial, geometric, and Poisson), continuous distributions (Gaussian and gamma) and their probability density functions, expectation, independence, and conditional probabilities, some of which are recalled in the first chapter. Such basic topics can be regarded as belonging to the field of "static" probability, *i.e.* probability without time dependence, as opposed to the contents of this text which is dealing with random evolution over time.

Our treatment of time-dependent randomness revolves around the important technique of first step analysis for random walks, branching processes, and more generally for Markov chains in discrete and continuous time, with application to the computation of ruin probabilities and mean hitting times. In addition to the treatment of Markov chains, a brief introduction to martingales is given in discrete time. This provides a different way to recover the computations of ruin probabilities and mean hitting times which have been presented in the Markovian framework. Spatial Poisson processes on abstract spaces are also considered without any time ordering, with the inclusion of some recent results on deviation inequalities and moment identities for Poisson stochastic integrals.

There already exist many textbooks on stochastic processes and Markov chains, including, e.g., Çinlar (1975), Karlin and Taylor (1998), Bosq and Nguyen (1996), Ross (1996), Norris (1998), Durrett (1999), Grimmett and Stirzaker (2001), Jones and Smith (2001), Steele (2001), Medhi (2010). In comparison with the existing literature, which is sometimes dealing with structural properties of stochastic processes *via* a more compact and abstract treatment, the present book tends to emphasize elementary and explicit calculations instead of quicker arguments that may shorten the path to the solution, while being sometimes difficult to reproduce by undergraduate students.

The text also includes 172 exercises and 40 longer problems whose solutions are completely worked out. Some of the exercises have been influenced by Çinlar (1975), Karlin and Taylor (1998), Ross (1996), Jones and Smith (2001), Medhi (2010), and other references, while a number of them are original, and their solutions have been derived independently. Clicking on an exercise number inside the solution section will send the reader to the original problem text inside the file. Conversely, clicking on a problem number sends the reader to the corresponding solution, however this feature should be used with caution.

The *problems*, which are longer than the exercises, are based on various topics of application, such as cookie-excited random walks (Problem 3.19), phase-type distributions (Problem 4.14), the Wright-Fisher model in population genetics (Problem 5.32), pattern recognition (Problem 5.33), probabilistic automata (Problem 5.36), time reversibility (Problem 7.32), the PageRank™ algorithm (Problem 7.34), mixing times and convergence to equilibrium (Problem 7.35), and the Ising model (Problem 7.36).

Some theorems whose proofs are technical, as in Chapters 7 and 9, have been quoted from Karlin and Taylor (1998) and Bosq and Nguyen (1996). The initial manuscript has benefited from numerous questions, comments and suggestions from undergraduate students in Stochastic Processes at the Nanyang Technological University (NTU) in Singapore. I also thank Yung-Hsiang Huang and Moti Ben-Ari for many corrections and suggestions.

The cover graph represents a 5-state discrete-time irreducible aperiodic Markov chain with transition matrix

$$P = \begin{bmatrix} 0 & 0.2 & 0.8 & 0 & 0 \\ 0.4 & 0 & 0 & 0.6 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.4 & 0.6 & 0 & 0 \end{bmatrix}.$$

This pdf file contains both internal and external links, 126 figures and 9 tables, including 15 animated Figures 3.1, 3.5, 3.6, 4.3, 4.6, 4.7, 7.7, 7.8, 8.9, S.13, S.15 and S.14, that may require the use of Acrobat Reader for viewing on the complete pdf file. It also includes 7 Python codes on pages 140, 141, 297, 333, and 342 and 34 R codes, *e.g.* on pages 28, 31, 140, 141, 271, 297, 310, 310, 333 and 342.

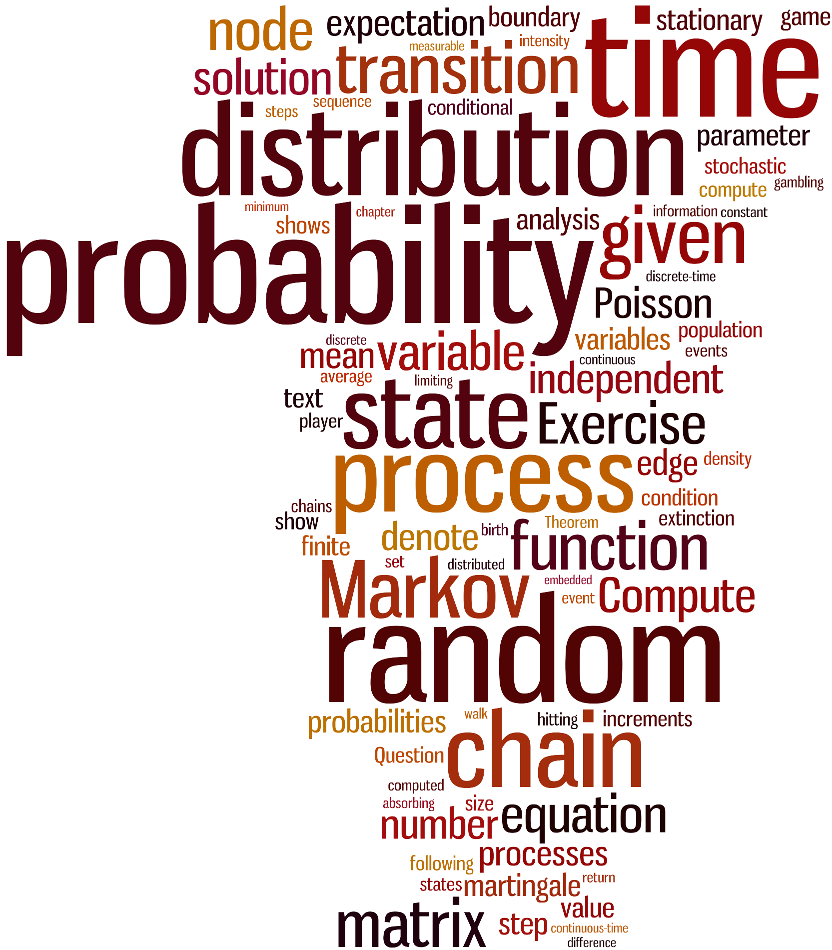
Nicolas Privault
May 2024

Contents

List of Figures

* Animated figures (work in Acrobat Reader).

List of Tables


node
expectation
boundary
stationary
game
measurable
intensity
solution
transition
time
steps
sequence
conditional
distribution
parameter
stochastic
compute
gambling
minimum
chapter
shows
analysis
information
constant
probability
given
discrete-time
Poisson
discrete
mean
variable
variables
population
events
continuous
average
limiting
independent
text
player
state
Exercise
process
edge
density
condition
chains
show
Theorem
extinction
finite
denote
birth
function
set
distributed
Markov
embedded
event
Compute
random
probabilities
walk
hitting
increments
Question
chain
computed
absorbing
size
number
equation
following
processes
states
martingale
return
value
matrix
step
continuous-time
difference

Introduction

A stochastic* process is a mathematical tool used for the modeling of time-dependent random phenomena. Here, the term "stochastic" means random and "process" refers to the time-evolving status of a given system. Stochastic processes have applications to multiple fields, and can be useful anytime one recognizes the role of randomness and unpredictability of events that can occur at random times in an *e.g.* physical, biological, or financial system.

For example, in applications to physics one can mention phase transitions, atomic emission phenomena, etc. In biology the time behavior of live beings is often subject to randomness, at least when the observer has only access to partial information. This latter point is of importance, as it links that the notion of randomness to the concept of information: what appears random to an observer may not be random to another observer equipped with more information. Think, for example, of the observation of the apparent random behavior of cars turning at a crossroad versus the point of view of car drivers, each of whom are acting according to their own decisions. In finance the importance of modeling time-dependent random phenomena is quite clear, as no one can make definite predictions for the future moves of risky assets. The concrete outcome of random modeling lies in the computation of *expectations* or *expected values*, which often turn out to be more useful than the probability values themselves. An average or expected lifetime, for example, can be easier to interpret than a (small) probability of default. The long term statistical behavior of random systems, which also involves the estimation of expectations, is a related issue of interest.

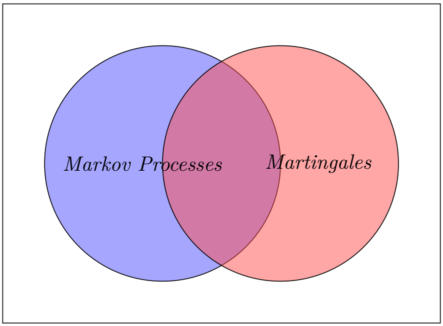
Basically, a stochastic process is a time-dependent family $(X_t)_{t\in T}$ of random variables, where t is a time index belonging to a parameter set or time scale T. In other words, instead of considering a single random variable X, one considers a whole family of random variables $(X_t)_{t\in T}$, with the addition of another level of technical difficulty. The time scale T can be finite (*e.g.*

* From the Greek "$\sigma\tau\acute{o}\chi o\varsigma$" (stokhos), meaning "guess", or "conjecture".

$T = \{1, 2, \ldots, N\}$) or countably infinite (*e.g.* $T = \mathbb{N} = \{0, 1, 2, \ldots\}$) or even uncountable (*e.g.* $T = [0, 1]$, $T = \mathbb{R}_+$). The case of uncountable T corresponds to continuous-time stochastic processes, and this setting is the most theoretically difficult. A serious treatment of continuous-time processes would actually require additional background in *measure theory*, which is outside of the scope of this text. Measure theory is the general study of measures on abstract spaces, including probability measures as a particular case, and allows for a rigorous treatment of integrals *via* integration in the Lebesgue sense. The Lebesgue integral is a powerful tool that allows one to integrate functions and random variables under minimal technical conditions. Here we mainly work in a discrete-time framework that mostly does not require the use of measure theory.

That being said, the definition of a stochastic process $(X_t)_{t \in T}$ remains vague at this stage since virtually *any* family of random variables could be called a stochastic process. In addition, working at such a level of generality without imposing any structure or properties on the processes under consideration could be of little practical use. As we will see later on, stochastic processes can be classified into two main families:

Stochastic Processes



- *Markov (1856-1922) Processes.*

Roughly speaking, a stochastic process $(X_s)_{s \geqslant 0}$ is a Markov (1909) process when its statistical behavior after time t can be recovered from the value X_t of the process at time t. In particular, the values X_s of the process at times $s \in [0, t)$ have no influence on this behavior as long as the value of X_t is known.

- *Martingales.*

Originally, a martingale is a strategy designed to win repeatedly in a game of chance. In mathematics, a stochastic process $(X_t)_{t \in \mathbb{R}_+}$ is a martingale if the best possible estimate at time s of its future value X_t at time $t > s$ is simply given by X_s. This requires the careful definition of a "best possible estimate", and for this we need the tool of conditional expectation which relies on estimation in the mean-square sense. Martingale are useful in physics and finance, where they are linked to the notion of equilibrium.

Time series of order greater than one form another class of stochastic processes that may have neither the Markov property nor the martingale property in general.

The outline of this text is as follows. After reviewing in Chapter 1 the probabilistic tools required in the analysis of Markov chains, we consider simple gambling problems in Chapter 2, due to their practical usefulness and to the fact that they only require a minimal theoretical background. Next, in Chapter 3 we turn to the study of discrete-time random walks with infinite state space, which can be defined as stochastic processes with independent increments, without requiring much abstract formalism. In Chapter 4 we introduce the general framework of Markov chains in discrete time, which includes the gambling process and the simple random walk of Chapters 2 and 3 as particular cases. In the subsequent Chapters 5, 6 and 7, Markov chains are considered from the point of view of first step analysis, which is introduced in detail in Chapter 5. The classification of states is reviewed in Chapter 6, with application to the long-run behavior of Markov chains in Chapter 7, which also includes a short introduction to the Markov chain Monte Carlo method. Branching processes are other examples of discrete-time Markov processes which have important applications in life sciences, *e.g.* for population dynamics or the control of disease outbreak, and they are considered in Chapter 8. Then in Chapter 9, we deal with Markov chains in continuous time, including Poisson and birth and death processes. Martingales are considered in Chapter 10, where they are used to recover in a simple and elegant way the main results of Chapter 2 on ruin probabilities and mean exit times for gambling processes. Spatial Poisson point processes, which can be defined on an abstract space without requiring an ordered time index, are presented in Chapter 11. The end of that chapter also includes some recent results on moments and deviation inequalities for Poisson stochastic integrals, whose proofs should be accessible to advanced undergraduates. Reliability theory is an important engineering application of Markov chains, and it is reviewed in Chapter 12. All stochastic processes considered in this text have a discrete state space and discontinuous trajectories. Brownian motion, which is the first main example of a stochastic process having continuous trajectories in continuous time, remains outside the scope of this book.

We close this introduction with examples of simulations for the random paths of some stochastic *Lévy processes*, which are processes with independent and stationary increments. The graphs shown below are for illustration purposes only. Indeed, apart from the Poisson process of Figure 0.4, most of these processes are however not within the scope of this text, and they are presented as an incentive to delve deeper into the field of stochastic processes.

Examples

i) Standard Brownian motion, $d = 1$.

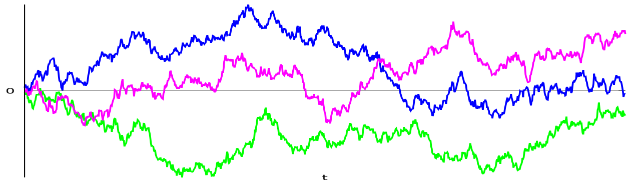


Fig. 0.1: Sample trajectories of Brownian motion.

ii) Standard Brownian motion, $d = 3$.

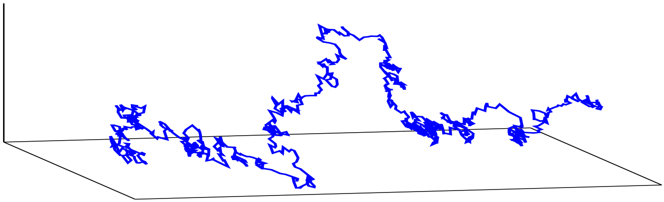


Fig. 0.2: Sample trajectory of a 3-dimensional Brownian motion.

iii) Drifted Brownian motion, $d = 1$.

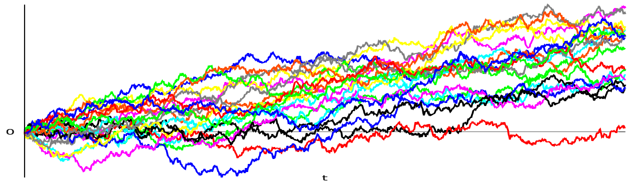


Fig. 0.3: Twenty paths of a drifted Brownian motion, $d = 1$.

iv) The Poisson process, $d = 1$.

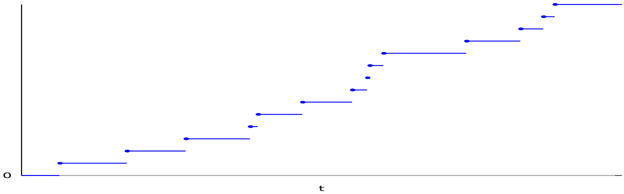


Fig. 0.4: Sample trajectory of a Poisson process.

v) Compound Poisson process, $d = 1$.

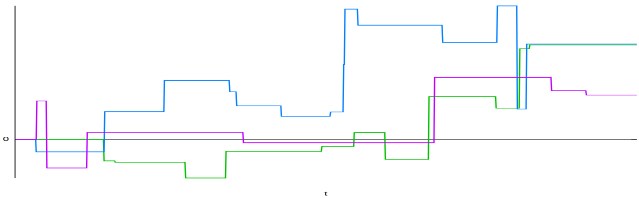


Fig. 0.5: Sample trajectories of a compound Poisson process.

vi) Gamma process, $d = 1$.

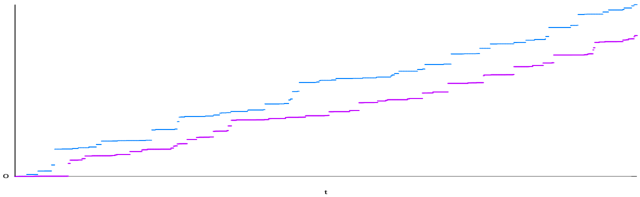


Fig. 0.6: Sample trajectories of a gamma process.

vii) Stable process, $d = 1$.

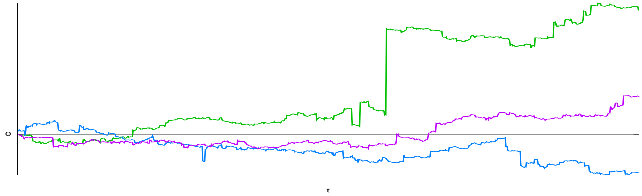


Fig. 0.7: Sample trajectories of a stable process.

viii) Cauchy process, $d = 1$.

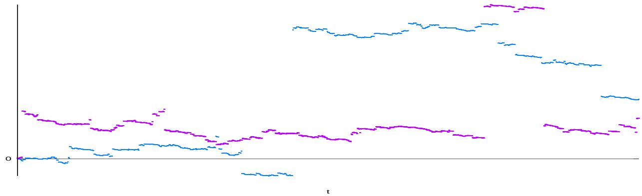


Fig. 0.8: Sample trajectories of a Cauchy process.

ix) Variance gamma process, $d = 1$.

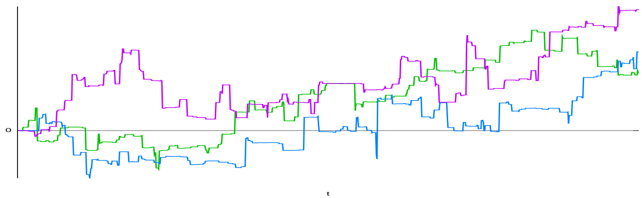


Fig. 0.9: Sample trajectories of a variance gamma process.

x) Inverse Gaussian process, $d = 1$.

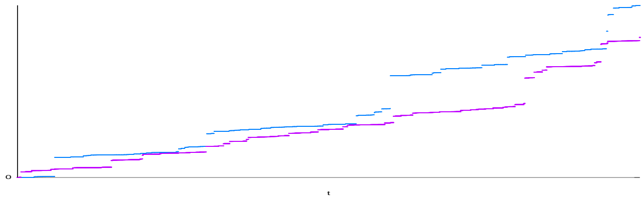


Fig. 0.10: Sample trajectories of an inverse Gaussian process.

xi) Negative Inverse Gaussian process, $d = 1$.

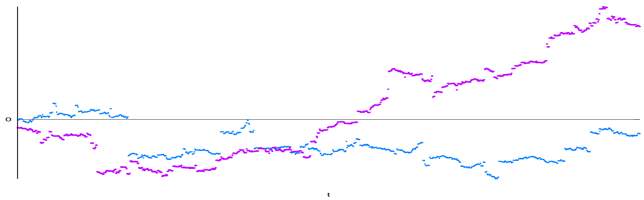


Fig. 0.11: Sample trajectories of a negative inverse Gaussian process.

Chapter 1
Probability Background

In this chapter, we review a number of basic probabilistic tools that will needed for the study of stochastic processes in the subsequent chapters. We refer the reader to *e.g.* Pitman (1999), Jacod and Protter (2000), Devore (2003), for additional background on probability theory.

1.1 Probability Sample Space and Events

We will need the following notation coming from set theory. Given A and B to abstract sets, "$A \subset B$" means that A is contained in B, and in this case, $B \setminus A$ denotes the set of elements of B which do not belong to A. The property that the element ω belongs to the set A is denoted by "$\omega \in A$", and given two sets A and Ω such that $A \subset \Omega$, we let $A^c = \Omega \setminus A$ denote the *complement* of A in Ω. The finite set made of n elements $\omega_1, \ldots, \omega_n$ is denoted by $\{\omega_1, \ldots, \omega_n\}$, and we will usually distinguish between the element ω and its associated singleton set $\{\omega\}$.

A probability sample space is an abstract set Ω that contains the possible outcomes of a random experiment.

Examples

i) Coin tossing: $\Omega = \{H, T\}$.

ii) Rolling one die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

iii) Picking one card at random in a pack of 52: $\Omega = \{1, 2, 3, \ldots, 52\}$.

iv) An integer-valued random outcome: $\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}$.

In this case the outcome $\omega \in \mathbb{N}$ can be the random number of trials needed until some event occurs.

v) A nonnegative, real-valued outcome: $\Omega = \mathbb{R}_+$.

In this case the outcome $\omega \in \mathbb{R}_+$ may represent the (nonnegative) value of a continuous random time.

vi) A random continuous parameter (such as time, weather, price or wealth, temperature, ...): $\Omega = \mathbb{R}$.

vii) Random choice of a continuous path in the space $\Omega = \mathcal{C}(\mathbb{R}_+)$ of all continuous functions on $\mathbb{R}_+$.

In this case, $\omega \in \Omega$ is a function $\omega : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and a typical example is the graph $t \longmapsto \omega(t)$ of a stock price over time.

Product spaces:

Probability sample spaces can be built as product spaces and used for the modeling of repeated random experiments.

i) Rolling two dice: $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.

In this case a typical element of Ω is written as $\omega = (k, l)$ with $k, l \in \{1, 2, 3, 4, 5, 6\}$.

ii) A finite number n of real-valued samples: $\Omega = \mathbb{R}^n$.

In this case the outcome ω is a vector $\omega = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with n components.

Note that to some extent, the more complex Ω is, the better it fits a practical and useful situation, *e.g.* $\Omega = \{H, T\}$ corresponds to a simple coin tossing experiment while $\Omega = \mathcal{C}(\mathbb{R}_+)$ the space of continuous functions on $\mathbb{R}_+$ can be applied to the modeling of stock markets. On the other hand, in many cases and especially in the most complex situations, we will *not* attempt to specify Ω explicitly.

Events

An event is a collection of outcomes, which is represented by a subset of Ω. In what follows we consider collections of events, called *σ-algebras* (or *σ-fields*), according to the following definition.

Definition 1.1. *A collection $\mathcal{G}$ of events is a σ-*algebra* provided that it satisfies the following conditions:*

(i) $\emptyset \in \mathcal{G}$,
(ii) *For all countable sequences* $(A_n)_{n\geqslant 1}$ *such that* $A_n \in \mathcal{G}$, $n \geqslant 1$, *we have* $\bigcup_{n\geqslant 1} A_n \in \mathcal{G}$,
(iii) $A \in \mathcal{G} \Longrightarrow (\Omega \setminus A) \in \mathcal{G}$,

where $\Omega \setminus A := \{\omega \in \Omega \ : \ \omega \notin A\}$.

Note that Properties (ii) and (iii) above also imply the stability of σ-algebras under intersections, as

$$\bigcap_{n\geqslant 1} A_n = \left(\bigcup_{n\geqslant 1} A_n^c \right)^c \in \mathcal{G}, \qquad (1.1.1)$$

for all countable sequences $A_n \in \mathcal{G}$, $n \geqslant 1$.

The collection of all events in Ω will often be denoted by $\mathcal{F}$. The empty set $\emptyset$ and the full space Ω are considered as events but they are of less importance because Ω corresponds to "any outcome may occur" while $\emptyset$ corresponds to an absence of outcome, or no experiment.

In the context of stochastic processes, two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ such that $\mathcal{F} \subset \mathcal{G}$ will refer to two different amounts of information, the amount of information associated to $\mathcal{F}$ being here lower than the one associated to $\mathcal{G}$.

The formalism of σ-algebras helps in describing events in a short and precise way.

Examples

i) Let $\Omega = \{1, 2, 3, 4, 5, 6\}$.

The event $A = \{2, 4, 6\}$ corresponds to

"the result of the experiment is an even number".

ii) Taking again $\Omega = \{1, 2, 3, 4, 5, 6\}$,

$$\mathcal{F} := \{\Omega, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$$

defines a σ-algebra on Ω which corresponds to the knowledge of parity of an integer picked at random from 1 to 6.

Note that in the set-theoretic notation, an event A is a subset of Ω, *i.e.* $A \subset \Omega$, while it is an element of $\mathcal{F}$, *i.e.* $A \in \mathcal{F}$. For example, we have $\Omega \supset \{2,4,6\} \in \mathcal{F}$, while $\{\{2,4,6\},\{1,3,5\}\} \subset \mathcal{F}$.

iii) Taking

$$\mathcal{G} := \{\Omega, \emptyset, \{2,4,6\}, \{2,4\}, \{6\}, \{1,2,3,4,5\}, \{1,3,5,6\}, \{1,3,5\}\} \supset \mathcal{F},$$

defines a σ-algebra on Ω which is bigger than $\mathcal{F}$ and includes the parity information contained in $\mathcal{F}$, in addition to information on whether the outcome of the experiment is equal to 6 or not.

iv) Take

$$\Omega = \{H,T\} \times \{H,T\} = \{(H,H),(H,T),(T,H),(T,T)\}.$$

In this case, the collection $\mathcal{F}$ of all possible events is given by

$$\begin{aligned} \mathcal{F} = \{ & \emptyset, \{(H,H)\}, \{(T,T)\}, \{(H,T)\}, \{(T,H)\}, \\ & \{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \{(H,T),(T,T)\}, \\ & \{(T,H),(T,T)\}, \{(H,T),(H,H)\}, \{(T,H),(H,H)\}, \\ & \{(H,H),(T,T),(T,H)\}, \{(H,H),(T,T),(H,T)\}, \\ & \{(H,T),(T,H),(H,H)\}, \{(H,T),(T,H),(T,T)\}, \Omega\}. \end{aligned} \tag{1.1.2}$$

Note that the set $\mathcal{F}$ of all events considered in (1.1.2) above has altogether

$1 = \binom{n}{0}$ event of cardinality 0,

$4 = \binom{n}{1}$ events of cardinality 1,

$6 = \binom{n}{2}$ events of cardinality 2,

$4 = \binom{n}{3}$ events of cardinality 3,

$1 = \binom{n}{4}$ event of cardinality 4,

with $n = 4$, for a total of

$$16 = 2^n = \sum_{k=0}^{4} \binom{4}{k} = 1 + 4 + 6 + 4 + 1$$

events. The collection of events

$$\mathcal{G} := \{\emptyset, \{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \Omega\}$$

defines a sub σ-algebra of $\mathcal{F}$, which corresponds to the restricted information "the results of two coin tossings are different".

Exercise: Write down the set of all events on $\Omega = \{H, T\}$.

Note also that (H, T) is different from (T, H), whereas $\{(H, T), (T, H)\}$ is equal to $\{(T, H), (H, T)\}$.

In addition, we will distinguish between the *outcome* $\omega \in \Omega$ and its associated *event* $\{\omega\} \in \mathcal{F}$, which satisfies $\{\omega\} \subset \Omega$.

1.2 Probability Measures

Definition 1.2. *A* probability measure *is a mapping* $\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$ *that assigns a probability* $\mathbb{P}(A) \in [0, 1]$ *to any event* $A \in \mathcal{F}$*, with the properties*

a) $\mathbb{P}(\Omega) = 1$*, and*

b) $\mathbb{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \sum_{n\geqslant 1} \mathbb{P}(A_n)$*, whenever* $A_k \cap A_l = \emptyset$*,* $k \neq l$*.*

Property (b) above is named the *law of total probability*. It states in particular that we have

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$$

when the subsets $A_1, \ldots, A_n$ of Ω are disjoint, and

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{1.2.1}$$

if $A \cap B = \emptyset$. We also have the *complement rule*

$$\mathbb{P}(A^c) = \mathbb{P}(\Omega \setminus A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A).$$

When A and B are not necessarily disjoint we can write

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$

which extends to arbitrary families of events $(A_i)_{i\in I}$ indexed by a finite set I as the inclusion-exclusion principle

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{J\subset I} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{j\in J} A_j\right), \tag{1.2.2}$$

and

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \sum_{I\subset J}(-1)^{|I|+1}\mathbb{P}\left(\bigcup_{i\in I} A_i\right). \quad (1.2.3)$$

The triple

$$(\Omega, \mathcal{F}, \mathbb{P}) \quad (1.2.4)$$

is called a *probability space*, and was introduced by A.N. Kolmogorov (1903-1987). This setting is generally referred to as the *Kolmogorov framework*.

A property or event is said to hold $\mathbb{P}$-*almost surely* (also written $\mathbb{P}$-*a.s.*) if it holds with probability equal to one.

Example

Take

$$\Omega = \{(T,T),(H,H),(H,T),(T,H)\}$$

and

$$\mathcal{F} = \{\emptyset, \{(T,T),(H,H)\}, \{(H,T),(T,H)\}, \Omega\}.$$

The *uniform* probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is given by setting

$$\mathbb{P}(\{(T,T),(H,H)\}) := \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\{(H,T),(T,H)\}) := \frac{1}{2}.$$

In addition, we have the following convergence properties.

1. Let $(A_n)_{n\in\mathbb{N}}$ be a *non-decreasing* sequence of events, *i.e.* $A_n \subset A_{n+1}$, $n \geqslant 0$. Then we have

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n). \quad (1.2.5)$$

2. Let $(A_n)_{n\in\mathbb{N}}$ be a *non-increasing* sequence of events, *i.e.* $A_{n+1} \subset A_n$, $n \geqslant 0$. Then we have

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n). \quad (1.2.6)$$

Theorem 1.3. *Borel-Cantelli Lemma. Let $(A_n)_{n\geqslant 1}$ denote a sequence of events on $(\Omega, \mathcal{F}, \mathbb{P})$, such that*

$$\sum_{n\geqslant 1}\mathbb{P}(A_n) < \infty.$$

Then we have

$$\mathbb{P}\left(\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k\right) = 0,$$

i.e. the probability that A_n occurs infinitely many times occur is zero.

1.3 Conditional Probabilities and Independence

We start with examples.

Consider a population $\Omega = M \cup W$ made of a set M of men and a set W of women. Here the σ-algebra $\mathcal{F} = \{\Omega, \emptyset, W, M\}$ corresponds to the information given by gender. After polling the population, *e.g.* for a market survey, it turns out that a proportion $p \in [0,1]$ of the population declares to like apples, while a proportion $1-p$ declares to dislike apples. Let $A \subset \Omega$ denote the subset of individuals who like apples, while $A^c \subset \Omega$ denotes the subset individuals who dislike apples, with

$$p = \mathbb{P}(A) \quad \text{and} \quad 1-p = \mathbb{P}(A^c),$$

e.g. $p = 60\%$ of the population likes apples. It may be interesting to get a more precise information and to determine

- the relative proportion $\dfrac{\mathbb{P}(A \cap W)}{\mathbb{P}(W)}$ of women who like apples, and
- the relative proportion $\dfrac{\mathbb{P}(A \cap M)}{\mathbb{P}(M)}$ of men who like apples.

Here, $\mathbb{P}(A \cap W)/\mathbb{P}(W)$ represents the probability that a randomly chosen woman in W likes apples, and $\mathbb{P}(A \cap M)/\mathbb{P}(M)$ represents the probability that a randomly chosen man in M likes apples. Those two ratios are interpreted as *conditional probabilities*, for example $\mathbb{P}(A \cap M)/\mathbb{P}(M)$ denotes the probability that a given individual likes apples *given that* he is a man.

For another example, suppose that the population Ω is split as $\Omega = Y \cup O$ into a set Y of "young" people and another set O of "old" people, and denote by $A \subset \Omega$ the set of people who voted for candidate A in an election. Here it can be of interest to find out the relative proportion

$$\mathbb{P}(A \mid Y) = \frac{\mathbb{P}(Y \cap A)}{\mathbb{P}(Y)}$$

of young people who voted for candidate A.

Definition 1.4. *Given any two events $A, B \subset \Omega$ with $\mathbb{P}(B) \neq 0$, we call*

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

the probability of A given B, *or* conditionally to B.

Remark 1.5. *We note that if* $\mathbb{P}(B) = 1$ *we have* $\mathbb{P}(A \cap B^c) \leqslant \mathbb{P}(B^c) = 0$, *hence* $\mathbb{P}(A \cap B^c) = 0$, *which implies*

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \cap B),$$

and $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

We also recall the following property:

$$\begin{aligned}\mathbb{P}\left(B \cap \bigcup_{n\geqslant 1} A_n\right) &= \mathbb{P}\left(\bigcup_{n\geqslant 1} (B \cap A_n)\right) \\ &= \sum_{n\geqslant 1} \mathbb{P}(B \cap A_n) \\ &= \sum_{n\geqslant 1} \mathbb{P}(B \mid A_n)\mathbb{P}(A_n) \\ &= \sum_{n\geqslant 1} \mathbb{P}(A_n \mid B)\mathbb{P}(B),\end{aligned}$$

for any family of disjoint events $(A_n)_{n\geqslant 1}$ with $A_i \cap A_j = \emptyset$, $i \neq j$, and $\mathbb{P}(B) > 0$, $n \geqslant 1$. This also shows that conditional probability measures are probability measures, in the sense that whenever $\mathbb{P}(B) > 0$, we have

a) $\mathbb{P}(\Omega \mid B) = 1$, and

b) $\mathbb{P}\left(\bigcup_{n\geqslant 1} A_n \,\middle|\, B\right) = \sum_{n\geqslant 1} \mathbb{P}(A_n \mid B)$, whenever $A_k \cap A_l = \emptyset$, $k \neq l$.

In particular, if $\bigcup_{n\geqslant 1} A_n = \Omega$, $(A_n)_{n\geqslant 1}$ becomes a *partition* of Ω and we get the *law of total probability*

$$\mathbb{P}(B) = \sum_{n\geqslant 1} \mathbb{P}(B \cap A_n) = \sum_{n\geqslant 1} \mathbb{P}(A_n \mid B)\mathbb{P}(B) = \sum_{n\geqslant 1} \mathbb{P}(B \mid A_n)\mathbb{P}(A_n), \tag{1.3.1}$$

provided that $A_i \cap A_j = \emptyset$, $i \neq j$, and $\mathbb{P}(B) > 0$, $n \geqslant 1$.

Remark. In general we have

$$\mathbb{P}\left(A \,\middle|\, \bigcup_{n\geqslant 1} B_n\right) \neq \sum_{n\geqslant 1} \mathbb{P}(A \mid B_n),$$

even when $B_k \cap B_l = \emptyset$, $k \neq l$. Indeed, taking for example $A = \Omega = B_1 \cup B_2$ with $B_1 \cap B_2 = \emptyset$ and $\mathbb{P}(B_1) = \mathbb{P}(B_2) = 1/2$, we have

$$1 = \mathbb{P}(\Omega \mid B_1 \cup B_2) \neq \mathbb{P}(\Omega \mid B_1) + \mathbb{P}(\Omega \mid B_2) = 2.$$

Independent events

Definition 1.6. *Two events A and B such that $\mathbb{P}(A), \mathbb{P}(B) > 0$ are said to be* independent *if*

$$\mathbb{P}(A \mid B) = \mathbb{P}(A). \tag{1.3.2}$$

We note that the independence condition (1.3.2) is equivalent to

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

1.4 Random Variables

A real-valued random variable is a mapping*

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

from a probability sample space Ω into the state space $\mathbb{R}$. Given

$$X : \Omega \longrightarrow \mathbb{R}$$

a random variable and a (measurable)† subset A of $\mathbb{R}$, we denote by $\{X \in A\}$ the event

$$\{X \in A\} := \{\omega \in \Omega \ : \ X(\omega) \in A\}.$$

Examples

i) Let $\Omega := \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$, and consider the mapping

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ (k,l) &\longmapsto k + l. \end{aligned}$$

Then X is a random variable giving the sum of the two numbers appearing on each die.

ii) the time needed everyday to travel from home to work or school is a random variable, as the precise value of this time may change from day to day under unexpected circumstances.

iii) the price of a risky asset can be modeled using a random variable.

---

* See (MOE and UCLES 2022, page 14) and (MOE and UCLES 2020, page 19).

† Measurability of subsets of $\mathbb{R}$ refers to *Borel measurability*, a concept which will not be defined in this text.

In the sequel we will often use the notion of *indicator function* $\mathbb{1}_A$ of an event $A \subset \Omega$.

Definition 1.7. *For any $A \subset \Omega$, the indicator function $\mathbb{1}_A$ is the random variable*

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ \omega &\longmapsto \mathbb{1}_A(\omega) \end{aligned}$$

defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Indicator functions satisfy the property

$$\mathbb{1}_{A\cap B}(\omega) = \mathbb{1}_A(\omega)\mathbb{1}_B(\omega), \tag{1.4.1}$$

since

$$\begin{aligned} \mathbb{1}_{A\cap B}(\omega) = 1 &\Longleftrightarrow \omega \in A \cap B \\ &\Longleftrightarrow \omega \in A \text{ and } \omega \in B \\ &\Longleftrightarrow \mathbb{1}_A(\omega) = 1 \text{ and } \mathbb{1}_B(\omega) = 1 \\ &\Longleftrightarrow \mathbb{1}_A(\omega)\mathbb{1}_B(\omega) = 1. \end{aligned}$$

We also have

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A\mathbb{1}_B,$$

and

$$\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B, \tag{1.4.2}$$

if $A \cap B = \emptyset$.

For example, if $\Omega = \mathbb{N}$ and $A = \{k\}$, for all $l \geqslant 0$ we have

$$\mathbb{1}_{\{k\}}(l) = \begin{cases} 1 & \text{if } k = l, \\ \\ 0 & \text{if } k \neq l. \end{cases}$$

Given X a random variable, we also let

$$\mathbb{1}_{\{X=n\}} = \begin{cases} 1 & \text{if } X = n, \\ \\ 0 & \text{if } X \neq n, \end{cases}$$

and

$$\mathbb{1}_{\{X<n\}} = \begin{cases} 1 & \text{if } X < n, \\ \\ 0 & \text{if } X \geqslant n. \end{cases}$$

1.5 Probability Distributions

The *probability distribution* of a random variable $X : \Omega \longrightarrow \mathbb{R}$ is the collection

$$\{\mathbb{P}(X \in A) \ : \ A \text{ is a measurable subset of } \mathbb{R}\}.$$

As the collection of *measurable* subsets of $\mathbb{R}$ coincides with the σ-algebra generated by the intervals in $\mathbb{R}$, the distribution of X can be reduced to the knowledge of the probabilities

$$\{\mathbb{P}(a < X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a) \ : \ a < b \in \mathbb{R}\},$$

or of the cumulative distribution functions

$$\{\mathbb{P}(X \leqslant a) \ : \ a \in \mathbb{R}\}, \quad \text{or} \quad \{\mathbb{P}(X \geqslant a) \ : \ a \in \mathbb{R}\},$$

see *e.g.* Corollary 3.8 in Çınlar (2011).

Two random variables X and Y are said to be independent under the probability $\mathbb{P}$ if their probability distributions satisfy

$$\mathbb{P}(X \in A \ , \ Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all (measurable) subsets A and B of $\mathbb{R}$.

Distributions admitting a density

We say that the distribution of X admits a probability *density* distribution function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ if, for all $a \leqslant b$, the probability $\mathbb{P}(a \leqslant X \leqslant b)$ can be written as

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b \varphi_X(x) dx.$$

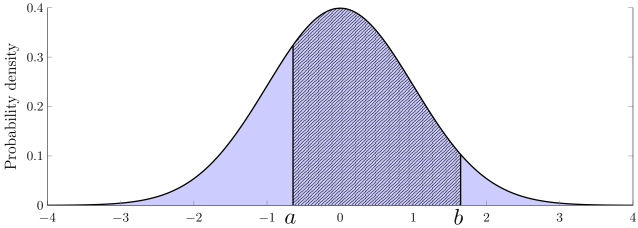


Fig. 1.1: Probability density function φ_X.

We also say that the distribution of X is absolutely continuous, or that X is an absolutely continuous random variable. This, however, does *not* imply that the density function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ is continuous.

In particular, we always have

$$\int_{-\infty}^{\infty} \varphi_X(x)dx = \mathbb{P}(-\infty \leqslant X \leqslant \infty) = 1$$

for any probability density functions $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$.

Remark 1.8. *Note that if the distribution of X admits a probability density function φ_X, then for all $a \in \mathbb{R}$ we have*

$$\mathbb{P}(X = a) = \int_a^a \varphi_X(x)dx = 0, \tag{1.5.1}$$

and this is not a contradiction.

In particular, Remark 1.8 shows that

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(X = a) + \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a < X < b),$$

for $a \leqslant b$. Property (1.5.1) appears for example in the framework of lottery games with a large number of participants, in which a given number "a" selected in advance has a very low (almost zero) probability to be chosen.

The probability density function φ_X can be recovered from the Cumulative Distribution Functions (CDFs)

$$x \longmapsto F_X(x) := \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} \varphi_X(s)ds,$$

and

$$x \longmapsto 1 - F_X(x) = \mathbb{P}(X \geqslant x) = \int_x^{\infty} \varphi_X(s)ds,$$

as

$$\varphi_X(x) = \frac{\partial F_X}{\partial x}(x) = \frac{\partial}{\partial x}\int_{-\infty}^{x} \varphi_X(s)ds = -\frac{\partial}{\partial x}\int_x^{\infty} \varphi_X(s)ds, \qquad x \in \mathbb{R}.$$

Examples

i) The *uniform* distribution on an interval.

The probability density function of the uniform distribution on the interval $[a, b]$, $a < b$, is given by

$$\varphi(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x), \qquad x \in \mathbb{R}.$$

ii) The *Gaussian* distribution.

The probability density function of the standard normal distribution is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

More generally, the probability density function of the Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ is given by

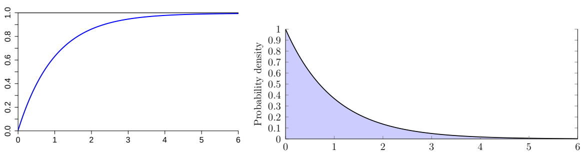
$$\varphi(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in \mathbb{R}.$$

In this case, we write $X \simeq \mathcal{N}(\mu, \sigma^2)$.

iii) The *exponential* distribution.

The probability density function of the exponential distribution with parameter $\lambda > 0$ is given by

$$\varphi(x) := \lambda \mathbb{1}_{[0,\infty)}(x) e^{-\lambda x} = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ \\ 0, & x < 0. \end{cases} \tag{1.5.2}$$



(a) Exponential CDF. (b) Exponential PDF.

Fig. 1.2: Exponential CDF and PDF.

We also have

$$\mathbb{P}(X > t) = e^{-\lambda t}, \qquad t \geqslant 0. \tag{1.5.3}$$

iv) The *gamma* distribution.

The probability density function of the gamma distribution is given by

$$\varphi(x) := \frac{a^\lambda}{\Gamma(\lambda)} \mathbb{1}_{[0,\infty)}(x) x^{\lambda-1} \mathrm{e}^{-ax} = \begin{cases} \dfrac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \mathrm{e}^{-ax}, & x \geqslant 0 \\ \\ 0, & x < 0, \end{cases}$$

where $a > 0$ and $\lambda > 0$ are scale and shape parameters, and

$$\Gamma(\lambda) := \int_0^\infty x^{\lambda-1} \mathrm{e}^{-x} dx, \qquad \lambda > 0,$$

is the gamma function.

v) The *Cauchy* distribution.

The probability density function of the Cauchy distribution is given by

$$\varphi(x) := \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}.$$

vi) The *lognormal* distribution.

The probability density function of the lognormal distribution is given by

$$\varphi(x) := \mathbb{1}_{[0,\infty)}(x) \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-(\mu - \log x)^2/(2\sigma^2)} = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-(\mu - \log x)^2/(2\sigma^2)}, & x \geqslant 0 \\ \\ 0, & x < 0. \end{cases}$$

Exercise: For each of the above probability density functions φ, check that the condition

$$\int_{-\infty}^\infty \varphi(x) dx = 1$$

is satisfied.

Joint densities

Given two absolutely continuous random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$, we can form the $\mathbb{R}^2$-valued random variable (X, Y) defined by

$$\begin{aligned} (X,Y) : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (X(\omega), Y(\omega)). \end{aligned}$$

We say that (X, Y) admits a joint probability density

$$\varphi_{(X,Y)} : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$$

when

$$\mathbb{P}((X,Y) \in A \times B) = \mathbb{P}(X \in A \text{ and } Y \in B) = \int_B \int_A \varphi_{(X,Y)}(x,y)dxdy$$

for all *measurable* subsets A, B of $\mathbb{R}$, see Figure 1.3.

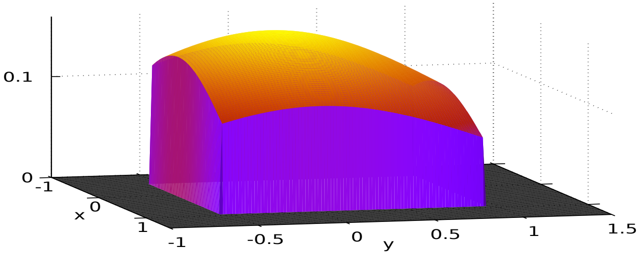


Fig. 1.3: Probability $\mathbb{P}((X,Y) \in [-0.5,1] \times [-0.5,1])$ computed as a volume integral.

The probability density function $\varphi_{(X,Y)}$ can be recovered from the joint cumulative distribution function

$$(x,y) \longmapsto F_{(X,Y)}(x,y) := \mathbb{P}(X \leqslant x \text{ and } Y \leqslant y) = \int_{-\infty}^x \int_{-\infty}^y \varphi_{(X,Y)}(s,t)dsdt,$$

and

$$(x,y) \longmapsto \mathbb{P}(X \geqslant x \text{ and } Y \geqslant y) = \int_x^\infty \int_y^\infty \varphi_{(X,Y)}(s,t)dsdt,$$

as

$$\begin{aligned} \varphi_{(X,Y)}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y) && (1.5.4) \\ &= \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^x \int_{-\infty}^y \varphi_{(X,Y)}(s,t)dsdt && (1.5.5) \\ &= \frac{\partial^2}{\partial x \partial y} \int_x^\infty \int_y^\infty \varphi_{(X,Y)}(s,t)dsdt, \end{aligned}$$

$x, y \in \mathbb{R}$.

The probability densities $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ and $\varphi_Y : \mathbb{R} \longrightarrow \mathbb{R}_+$ of $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ are called the *marginal densities* of (X,Y), and are given by

$$\varphi_X(x) = \int_{-\infty}^{\infty} \varphi_{(X,Y)}(x,y)dy, \qquad x \in \mathbb{R}, \tag{1.5.6}$$

and

$$\varphi_Y(y) = \int_{-\infty}^{\infty} \varphi_{(X,Y)}(x,y)dx, \qquad y \in \mathbb{R}.$$

The conditional probability density $\varphi_{X|Y=y} : \mathbb{R} \longrightarrow \mathbb{R}_+$ of X given $Y = y$ is defined by

$$\varphi_{X|Y=y}(x) := \frac{\varphi_{(X,Y)}(x,y)}{\varphi_Y(y)}, \qquad x, y \in \mathbb{R}, \tag{1.5.7}$$

provided that $\varphi_Y(y) > 0$. In particular, X and Y are independent if and only if

$$\varphi_{X|Y=y}(x) = \varphi_X(x), \quad i.e., \quad \varphi_{(X,Y)}(x,y) = \varphi_X(x)\varphi_Y(y), \qquad x, y \in \mathbb{R}.$$

Example

If $X_1, \ldots, X_n$ are independent exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_n$ we have

$$\begin{aligned} \mathbb{P}(\min(X_1, \ldots, X_n) > t) &= \mathbb{P}(X_1 > t, \ldots, X_n > t) \\ &= \mathbb{P}(X_1 > t) \cdots \mathbb{P}(X_n > t) \\ &= \mathrm{e}^{-(\lambda_1 + \cdots + \lambda_n)t}, \qquad t \geqslant 0, \end{aligned} \tag{1.5.8}$$

hence $\min(X_1, \ldots, X_n)$ is an exponentially distributed random variable with parameter $\lambda_1 + \cdots + \lambda_n$.

From the probability density function of (X_1, X_2) given by

$$\varphi_{(X_1,X_2)}(x,y) = \varphi_{X_1}(x)\varphi_{X_2}(y) = \lambda_1\lambda_2 \mathrm{e}^{-\lambda_1 x - \lambda_2 y}, \qquad x, y \geqslant 0,$$

we can write

$$\begin{aligned} \mathbb{P}(X_1 < X_2) &= \mathbb{P}(X_1 \leqslant X_2) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} \varphi_{(X_1,X_2)}(x,y)dxdy \\ &= \lambda_1\lambda_2 \int_0^{\infty} \int_0^{y} \mathrm{e}^{-\lambda_1 x - \lambda_2 y} dxdy \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}, \end{aligned} \tag{1.5.9}$$

and we note that

$$\mathbb{P}(X_1 = X_2) = \lambda_1\lambda_2 \int_{\{(x,y)\in\mathbb{R}_+^2 \ : \ x=y\}} \mathrm{e}^{-\lambda_1 x - \lambda_2 y} dxdy = 0.$$

Discrete distributions

We only consider integer-valued random variables, *i.e.* the distribution of X is given by the values of $\mathbb{P}(X = k)$, $k \geqslant 0$.

Examples

i) The *Bernoulli* distribution.

We have

$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p, \qquad (1.5.10)$$

where $p \in [0, 1]$ is a parameter.

Note that any Bernoulli random variable $X : \Omega \longrightarrow \{0, 1\}$ can be written as the indicator function

$$X = \mathbb{1}_A$$

on Ω with $A = \{X = 1\} = \{\omega \in \Omega \ : \ X(\omega) = 1\}$.

ii) The *binomial* distribution.

We have

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

where $n \geqslant 1$ and $p \in [0, 1]$ are parameters and $\binom{n}{k} = n!/(k!(n-k)!)$, $0 \leqslant k \leqslant n$.

iii) The *geometric* distribution.

In this case, we have

$$\mathbb{P}(X = k) = (1-p)p^k, \qquad k \geqslant 0, \qquad (1.5.11)$$

where $p \in (0, 1)$ is a parameter. For example, if $(X_k)_{k \in \mathbb{N}}$ is a sequence of independent Bernoulli random variables with distribution (1.5.10), then the random variable,*

$$T_0 := \inf\{k \geqslant 0 \ : \ X_k = 0\}$$

can denote the duration of a game until the time that the wealth X_k of a player reaches 0. The random variable T_0 has the geometric distribution (1.5.11) with parameter $p \in (0, 1)$.

* The notation "inf" stands for "infimum", meaning the smallest $n \geqslant 0$ such that $X_n = 0$, if such an n exists.

iv) The *negative binomial* (or *Pascal*) distribution.

We have

$$\mathbb{P}(X = k) = \binom{k + r - 1}{r - 1} (1 - p)^r p^k, \qquad k \geqslant 0, \tag{1.5.12}$$

where $p \in (0, 1)$ and $r \geqslant 1$ are parameters. Note that the sum of $r \geqslant 1$ independent geometric random variables with parameter p has a negative binomial distribution with parameter (r, p). In particular, the negative binomial distribution recovers the geometric distribution when $r = 1$.

v) The *Poisson* distribution.

We have

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k \geqslant 0,$$

where $\lambda > 0$ is a parameter.

The probability that a discrete nonnegative random variable $X : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ is finite is given by

$$\mathbb{P}(X < \infty) = \sum_{k \geqslant 0} \mathbb{P}(X = k), \tag{1.5.13}$$

and we have

$$1 = \mathbb{P}(X = \infty) + \mathbb{P}(X < \infty) = \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} \mathbb{P}(X = k).$$

Remark 1.9. *The distribution of a discrete random variable cannot admit a probability density. If this were the case, by Remark 1.8 we would have* $\mathbb{P}(X = k) = 0$ *for all* $k \geqslant 0$ *and*

$$1 = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(X \in \mathbb{N}) = \sum_{k \geqslant 0} \mathbb{P}(X = k) = 0,$$

which is a contradiction.

Given two discrete random variables X and Y, the conditional distribution of X given $Y = k$ is given by

$$\mathbb{P}(X = n \mid Y = k) = \frac{\mathbb{P}(X = n \text{ and } Y = k)}{\mathbb{P}(Y = k)}, \qquad n \geqslant 0,$$

provided that $\mathbb{P}(Y = k) > 0$, $k \geqslant 0$.

1.6 Expectation of Random Variables

The *expectation*, or *expected value*, of a random variable X is the mean, or average value, of X. In practice, expectations can be even more useful than probabilities. For example, knowing that a given equipment (such as a bridge) has a failure probability of 1.78493 out of a billion can be of less practical use than knowing the expected lifetime (*e.g.* 200000 years) of that equipment.

For example, the time $T(\omega)$ to travel from home to work/school can be a random variable with a new outcome and value every day, however we usually refer to its expectation $\mathbb{E}[T]$ rather than to its sample values that may change from day to day.

Expected value of a Bernoulli random variable

Any Bernoulli random variable $X : \Omega \longrightarrow \{0, 1\}$ can be written as the indicator function $X := \mathbb{1}_A$ where A is the event $A = \{X = 1\}$, and the parameter $p \in [0, 1]$ of X is given by

$$p = \mathbb{P}(X = 1) = \mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[X].$$

The expectation of a Bernoulli random variable with parameter p is defined as

$$\mathbb{E}[\mathbb{1}_A] := 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A). \quad (1.6.1)$$

Expected value of a discrete random variable

Next, let $X : \Omega \longrightarrow \mathbb{N}$ be a discrete random variable. The expectation $\mathbb{E}[X]$ of X is defined as the sum

$$\mathbb{E}[X] = \sum_{k \geqslant 0} k \mathbb{P}(X = k), \quad (1.6.2)$$

in which the possible values $k \geqslant 0$ of X are weighted by their probabilities. More generally we have

$$\mathbb{E}[\phi(X)] = \sum_{k \geqslant 0} \phi(k) \mathbb{P}(X = k),$$

for all sufficiently summable functions $\phi : \mathbb{N} \longrightarrow \mathbb{R}$.

The expectation of the indicator function $X = \mathbb{1}_A = \mathbb{1}_{\{X=1\}}$ can be recovered from (1.6.2) as

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_A] = 0 \times \mathbb{P}(\Omega \setminus A) + 1 \times \mathbb{P}(A) = 0 + \mathbb{P}(A) = \mathbb{P}(A).$$

Note that the expectation is a *linear* operation, *i.e.* we have

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \qquad a, b \in \mathbb{R}, \tag{1.6.3}$$

provided that

$$\mathbb{E}[|X|] + \mathbb{E}[|Y|] < \infty.$$

Examples

i) Expected value of a Poisson random variable with parameter $\lambda > 0$:

$$\mathbb{E}[X] = \sum_{k \geqslant 0} k\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \sum_{k \geqslant 1} k \frac{\lambda^k}{k!} = \lambda \mathrm{e}^{-\lambda} \sum_{k \geqslant 0} \frac{\lambda^k}{k!} = \lambda, \tag{1.6.4}$$

where we used the exponential series (A.1).

ii) Estimating the expected value of a Poisson random variable using R:

Taking $\lambda := 2$, we can use the following R code:

```
poisson_samples <- rpois(100000, lambda = 2)
poisson_samples
mean(poisson_samples)
```

Given $X : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ a discrete nonnegative random variable X, we have

$$\mathbb{P}(X < \infty) = \sum_{k \geqslant 0} \mathbb{P}(X = k),$$

and

$$1 = \mathbb{P}(X = \infty) + \mathbb{P}(X < \infty) = \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} \mathbb{P}(X = k),$$

and in general

$$\mathbb{E}[X] = +\infty \times \mathbb{P}(X = \infty) + \sum_{k \geqslant 0} k\mathbb{P}(X = k).$$

In particular, $\mathbb{P}(X = \infty) > 0$ implies $\mathbb{E}[X] = \infty$, and the finiteness condition $\mathbb{E}[X] < \infty$ implies $\mathbb{P}(X < \infty) = 1$, however the converse is *not true*. For example, assume that X has the geometric distribution

$$\mathbb{P}(X = k) := \frac{1}{2^{k+1}}, \qquad k \geqslant 0, \tag{1.6.5}$$

with parameter $p = 1/2$, and

$$\mathbb{E}[X] = \sum_{k\geqslant 0} \frac{k}{2^{k+1}} = \frac{1}{4}\sum_{k\geqslant 1} \frac{k}{2^{k-1}} = \frac{1}{4}\frac{1}{(1-1/2)^2} = 1 < \infty,$$

by (A.4). Letting $\phi(X) := 2^X$, we have

$$\mathbb{P}(\phi(X) < \infty) = \mathbb{P}(X < \infty) = \sum_{k\geqslant 0} \frac{1}{2^{k+1}} = 1,$$

and

$$\mathbb{E}[\phi(X)] = \sum_{k\geqslant 0} \phi(k)\mathbb{P}(X = k) = \sum_{k\geqslant 0} \frac{2^k}{2^{k+1}} = \sum_{k\geqslant 0} \frac{1}{2} = +\infty,$$

hence the expectation $\mathbb{E}[\phi(X)]$ is *infinite* although $\phi(X)$ is *finite* with probability one.*

Conditional expectation

The notion of expectation takes its full meaning under conditioning. For example, the expected return of a random asset usually depends on information such as economic data, location, etc. In this case, replacing the expectation by a conditional expectation will provide a better estimate of the expected value.

For instance, life expectancy is a natural example of a conditional expectation since it typically depends on location, gender, and other parameters.

The *conditional expectation* of a finite discrete random variable $X : \Omega \longrightarrow \mathbb{N}$ given an event A is defined by

$$\mathbb{E}[X \mid A] = \sum_{k\geqslant 0} k\mathbb{P}(X = k \mid A) = \sum_{k\geqslant 1} k\frac{\mathbb{P}(X = k \text{ and } A)}{\mathbb{P}(A)}.$$

Lemma 1.10. *Given an event A such that $\mathbb{P}(A) > 0$, we have*

$$\mathbb{E}[X \mid A] = \frac{1}{\mathbb{P}(A)}\mathbb{E}\big[X\mathbb{1}_A\big]. \tag{1.6.6}$$

Proof. The proof is done only for $X : \Omega \longrightarrow \mathbb{N}$ a discrete random variable, however (1.6.6) is valid for general real-valued random variables. By Relation (1.4.1) we have

$$\mathbb{E}[X \mid A] = \sum_{k\geqslant 0} k\mathbb{P}(X = k \mid A)$$

* This is the St. Petersburg paradox.

$$= \frac{1}{\mathbb{P}(A)} \sum_{k\geqslant 0} k\mathbb{P}(\{X=k\}\cap A) = \frac{1}{\mathbb{P}(A)} \sum_{k\geqslant 0} k\mathbb{E}\big[\mathbb{1}_{\{X=k\}\cap A}\big]$$
$$= \frac{1}{\mathbb{P}(A)} \sum_{k\geqslant 0} k\mathbb{E}\big[\mathbb{1}_{\{X=k\}}\mathbb{1}_A\big] = \frac{1}{\mathbb{P}(A)}\mathbb{E}\left[\mathbb{1}_A \sum_{k\geqslant 0} k\mathbb{1}_{\{X=k\}}\right]$$
$$= \frac{1}{\mathbb{P}(A)}\mathbb{E}\big[\mathbb{1}_A X\big],$$

where we used the relation

$$X = \sum_{k\geqslant 0} k\mathbb{1}_{\{X=k\}}$$

which holds since X takes only integer values. □

Example

i) For example, consider $\Omega = \{1, 3, -1, -2, 5, 7\}$ with the non-uniform probability measure given by

$$\mathbb{P}(\{-1\}) = \mathbb{P}(\{-2\}) = \mathbb{P}(\{1\}) = \mathbb{P}(\{3\}) = \frac{1}{7},\ \mathbb{P}(\{5\}) = \frac{2}{7},\ \mathbb{P}(\{7\}) = \frac{1}{7},$$

and the random variable

$$X : \Omega \longrightarrow \mathbb{Z}$$

given by

$$X(k) = k, \qquad k = 1, 3, -1, -2, 5, 7.$$

Here, $\mathbb{E}[X \mid X > 1]$ denotes the expected value of X given

$$A = \{X > 1\} = \{3, 5, 7\} \subset \Omega,$$

i.e. the mean value of X given that X is strictly positive. This conditional expectation can be computed as

$$\mathbb{E}[X \mid X > 1]$$
$$= 3 \times \mathbb{P}(X = 3 \mid X > 1) + 5 \times \mathbb{P}(X = 5 \mid X > 1) + 7 \times \mathbb{P}(X = 7 \mid X > 1)$$
$$= \frac{3 + 2 \times 5 + 7}{4}$$
$$= \frac{3 + 5 + 5 + 7}{7 \times 4/7}$$
$$= \frac{1}{\mathbb{P}(X > 1)}\mathbb{E}\big[X\mathbb{1}_{\{X>1\}}\big],$$

where $\mathbb{P}(X > 1) = 4/7$ and the truncated expectation $\mathbb{E}\big[X\mathbb{1}_{\{X>1\}}\big]$ is given by $\mathbb{E}\big[X\mathbb{1}_{\{X>1\}}\big] = (3 + 2 \times 5 + 7)/7$.

ii) Estimating a conditional expectation using R:

```
geo_samples <- rgeom(100000, prob = 1/4)
mean(geo_samples)
mean(geo_samples[geo_samples<10])
```

Taking $p := 3/4$, by (A.4) we have

$$\mathbb{E}[X] = (1-p)\sum_{k\geqslant 1} kp^k = \frac{p}{1-p} = 3,$$

and

$$\begin{aligned}\mathbb{E}[X \mid X < 10] &= \frac{1}{\mathbb{P}(X<10)}\mathbb{E}\big[X\mathbb{1}_{\{X<10\}}\big] \\ &= \frac{1}{\mathbb{P}(X<10)}\sum_{k=0}^{9} k\mathbb{P}(X=k) \\ &= \frac{1}{\sum_{k=0}^{9} p^k}\sum_{k=1}^{9} kp^k \\ &= \frac{p(1-p)}{1-p^{10}}\frac{\partial}{\partial p}\sum_{k=0}^{9} p^k \\ &= \frac{p(1-p)}{1-p^{10}}\frac{\partial}{\partial p}\left(\frac{1-p^{10}}{1-p}\right) \\ &= \frac{p(1-p^{10}-10(1-p)p^9)}{(1-p)(1-p^{10})} \\ &\simeq 2.4032603455.\end{aligned}$$

If the random variable $X : \Omega \longrightarrow \mathbb{N}$ is independent* of the event A, we have

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[X]\mathbb{P}(A),$$

and we naturally find

$$\mathbb{E}[X \mid A] = \mathbb{E}[X]. \tag{1.6.7}$$

Taking $X = \mathbb{1}_A$ with

$$\begin{aligned}\mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ \omega &\longmapsto \mathbb{1}_A := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}\end{aligned}$$

shows that, in particular,

* *i.e.*, $\mathbb{P}(\{X=k\}\cap A) = \mathbb{P}(\{X=k\})\mathbb{P}(A)$ for all $k \geqslant 0$.

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A \mid A] &= 0 \times \mathbb{P}(X = 0 \mid A) + 1 \times \mathbb{P}(X = 1 \mid A)\\ &= \mathbb{P}(X = 1 \mid A)\\ &= \mathbb{P}(A \mid A)\\ &= 1.\end{aligned}$$

One can also define the conditional expectation of X given $A = \{Y = k\}$, as

$$\mathbb{E}[X \mid Y = k] = \sum_{n \geqslant 0} n \mathbb{P}(X = n \mid Y = k),$$

where $Y : \Omega \longrightarrow \mathbb{N}$ is a discrete random variable.

Proposition 1.11. *Given X a discrete random variable such that $\mathbb{E}[|X|] < \infty$, we have the relation*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]], \tag{1.6.8}$$

which is sometimes referred to as the tower property.

Proof. We have

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X \mid Y]] &= \sum_{k \geqslant 0} \mathbb{E}[X \mid Y = k] \mathbb{P}(Y = k)\\ &= \sum_{k \geqslant 0} \sum_{n \geqslant 0} n \mathbb{P}(X = n \mid Y = k) \mathbb{P}(Y = k)\\ &= \sum_{n \geqslant 0} n \sum_{k \geqslant 0} \mathbb{P}(X = n \text{ and } Y = k)\\ &= \sum_{n \geqslant 0} n \mathbb{P}(X = n) = \mathbb{E}[X],\end{aligned}$$

where we used the marginal distribution

$$\mathbb{P}(X = n) = \sum_{k \geqslant 0} \mathbb{P}(X = n \text{ and } Y = k), \qquad n \geqslant 0,$$

that follows from the *law of total probability* (1.3.1) with $A_k = \{Y = k\}$, $k \geqslant 0$. □

Taking

$$Y = \sum_{k \geqslant 0} k \mathbb{1}_{A_k},$$

with $A_k := \{Y = k\}$, $k \geqslant 0$, from (1.6.8) we also get the *law of total expectation*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] \tag{1.6.9}$$

$$= \sum_{k \geqslant 0} \mathbb{E}[X \mid Y = k] \mathbb{P}(Y = k)$$
$$= \sum_{k \geqslant 0} \mathbb{E}[X \mid A_k] \mathbb{P}(A_k).$$

Example

Life expectancy in Singapore is $\mathbb{E}[T] = 80$ years overall, where T denotes the lifetime of a given individual chosen at random. Let $G \in \{m, w\}$ denote the gender of that individual. The statistics show that

$$\mathbb{E}[T \mid G = m] = 78 \quad \text{and} \quad \mathbb{E}[T \mid G = w] = 81.9,$$

and we have

$$\begin{aligned} 80 &= \mathbb{E}[T] \\ &= \mathbb{E}[\mathbb{E}[T|G]] \\ &= \mathbb{P}(G = w)\mathbb{E}[T \mid G = w] + \mathbb{P}(G = m)\mathbb{E}[T \mid G = m] \\ &= 81.9 \times \mathbb{P}(G = w) + 78 \times \mathbb{P}(G = m) \\ &= 81.9 \times (1 - \mathbb{P}(G = m)) + 78 \times \mathbb{P}(G = m), \end{aligned}$$

showing that

$$80 = 81.9 \times (1 - \mathbb{P}(G = m)) + 78 \times \mathbb{P}(G = m),$$

i.e.

$$\mathbb{P}(G = m) = \frac{81.9 - 80}{81.9 - 78} = \frac{1.9}{3.9} = 0.487.$$

Variance

The *variance* of a random variable X is defined by

$$\mathrm{Var}[X] := \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2,$$

provided that $\mathbb{E}\left[|X|^2\right] < \infty$. If $(X_k)_{k=1,\ldots,n}$ is a sequence of independent random variables, we have

$$\mathrm{Var}\left[\sum_{k=1}^n X_k\right] = \mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] - \left(\mathbb{E}\left[\sum_{k=1}^n X_k\right]\right)^2$$
$$= \mathbb{E}\left[\sum_{k=1}^n X_k \sum_{l=1}^n X_l\right] - \mathbb{E}\left[\sum_{k=1}^n X_k\right] \mathbb{E}\left[\sum_{l=1}^n X_l\right]$$

$$= \mathbb{E}\left[\sum_{k=1}^n \sum_{l=1}^n X_k X_l\right] - \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}[X_k]\mathbb{E}[X_l]$$
$$= \sum_{k=1}^n \mathbb{E}\big[X_k^2\big] + \sum_{1\leqslant k\neq l\leqslant n} \mathbb{E}[X_k X_l] - \sum_{k=1}^n (\mathbb{E}[X_k])^2 - \sum_{1\leqslant k\neq l\leqslant n} \mathbb{E}[X_k]\mathbb{E}[X_l]$$
$$= \sum_{k=1}^n \left(\mathbb{E}\big[X_k^2\big] - (\mathbb{E}[X_k])^2\right)$$
$$= \sum_{k=1}^n \mathrm{Var}[X_k]. \tag{1.6.10}$$

Random sums

In the sequel we consider $Y : \Omega \longrightarrow \mathbb{N}$ an *a.s.* finite, integer-valued random variable, *i.e.* we have $\mathbb{P}(Y < \infty) = 1$ and $\mathbb{P}(Y = \infty) = 0$. Based on the tower property of conditional expectations (1.6.8) or ordinary conditioning, the expectation of a random sum $\sum_{k=1}^Y X_k$, where $(X_k)_{k\in\mathbb{N}}$ is a sequence of random variables, can be computed from the *tower property* (1.6.8) or from the *law of total expectation* (1.6.9) as

$$\mathbb{E}\left[\sum_{k=1}^Y X_k\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^Y X_k \,\Big|\, Y\right]\right]$$
$$= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^Y X_k \,\Big|\, Y = n\right] \mathbb{P}(Y = n)$$
$$= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^n X_k \,\Big|\, Y = n\right] \mathbb{P}(Y = n),$$

and if the sequence $(X_k)_{k\in\mathbb{N}}$ is (mutually) independent of Y, this yields

$$\mathbb{E}\left[\sum_{k=1}^Y X_k\right] = \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^n X_k\right] \mathbb{P}(Y = n)$$
$$= \sum_{n\geqslant 0} \mathbb{P}(Y = n) \sum_{k=1}^n \mathbb{E}[X_k].$$

Random products

Similarly, for a random product we will have, using the independence of Y with $(X_k)_{k\in\mathbb{N}}$,

$$\mathbb{E}\left[\prod_{k=1}^{Y} X_k\right] = \sum_{n\geqslant 0}\mathbb{E}\left[\prod_{k=1}^{n} X_k\right]\mathbb{P}(Y=n) \qquad (1.6.11)$$
$$= \sum_{n\geqslant 0}\mathbb{P}(Y=n)\prod_{k=1}^{n}\mathbb{E}[X_k],$$

where the last equality requires the (mutual) independence of the random variables in the sequence $(X_k)_{k\geqslant 1}$.

Distributions admitting a density

Given a random variable X whose distribution admits a probability density $\varphi_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\varphi_X(x)dx,$$

and more generally,

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty}\phi(x)\varphi_X(x)dx, \qquad (1.6.12)$$

for all sufficiently integrable function ϕ on $\mathbb{R}$. For example, if X has a standard normal distribution we have

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty}\phi(x)\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

Examples

a) In case X has a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, we have

$$\mathbb{E}[\phi(X)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\phi(x)\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}dx. \qquad (1.6.13)$$

b) The uniform random variable U on $[0,1]$ satisfies $\mathbb{E}[U] = 1/2 < \infty$ and

$$\mathbb{P}(1/U < \infty) = \mathbb{P}(U > 0) = \mathbb{P}(U \in (0,1]) = 1,$$

however we have

$$\mathbb{E}[1/U] = \int_0^1 \frac{dx}{x} = +\infty,$$

and $\mathbb{P}(1/U = +\infty) = \mathbb{P}(U = 0) = 0$.

c) If the random variable X has an exponential distribution with parameter $\mu > 0$ we have

$$\mathbb{E}\left[e^{\lambda X}\right] = \mu \int_0^\infty e^{\lambda x} e^{-\mu x} dx = \begin{cases} \dfrac{\mu}{\mu - \lambda} < \infty & \text{if } \mu > \lambda, \\ \\ +\infty, & \text{if } \mu \leqslant \lambda. \end{cases}$$

Exercise: In case $X \simeq \mathcal{N}(\mu, \sigma^2)$ has a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, check that

$$\mu = \mathbb{E}[X] \qquad \text{and} \qquad \sigma^2 = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

When $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ is a $\mathbb{R}^2$-valued couple of random variables whose distribution admits a probability density $\varphi_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ we have

$$\mathbb{E}[\phi(X, Y)] = \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(x, y) \varphi_{X,Y}(x, y) dx dy,$$

for all sufficiently integrable function ϕ on $\mathbb{R}^2$.

The expectation of an absolutely continuous random variable satisfies the same linearity property (1.6.3) as in the discrete case.

The conditional expectation of an absolutely continuous random variable can be defined as

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^\infty x \varphi_{X|Y=y}(x) dx$$

where the conditional probability density $\varphi_{X|Y=y}(x)$ is defined in (1.5.7), with the relation

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] \tag{1.6.14}$$

which is called the *tower property* and holds as in the discrete case, since

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{-\infty}^\infty \mathbb{E}[X \mid Y = y] \varphi_Y(y) dy \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty x \varphi_{X|Y=y}(x) \varphi_Y(y) dx dy \\ &= \int_{-\infty}^\infty x \int_{-\infty}^\infty \varphi_{(X,Y)}(x, y) dy dx \\ &= \int_{-\infty}^\infty x \varphi_X(x) dx = \mathbb{E}[X], \end{aligned}$$

where we used Relation (1.5.6) between the probability density of (X, Y) and its marginal X.

For example, an exponentially distributed random variable X with probability density function (1.5.2) has the expected value

$$\mathbb{E}[X] = \lambda \int_0^\infty x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

1.7 Moment and Probability Generating Functions

Characteristic functions

The *characteristic function* of a random variable X is the function

$$\Psi_X : \mathbb{R} \longrightarrow \mathbb{C}$$

defined by

$$\Psi_X(t) = \mathbb{E}\big[e^{itX}\big], \qquad t \in \mathbb{R}.$$

The characteristic function Ψ_X of a random variable X with probability density function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ satisfies

$$\Psi_X(t) = \int_{-\infty}^\infty e^{ixt} \varphi(x) dx, \qquad t \in \mathbb{R}.$$

On the other hand, if $X : \Omega \longrightarrow \mathbb{N}$ is a discrete random variable we have

$$\Psi_X(t) = \sum_{n \geqslant 0} e^{itn} \mathbb{P}(X = n), \qquad t \in \mathbb{R}.$$

One of the main applications of characteristic functions is to provide a characterization of probability distributions, as in the following theorem.

Theorem 1.12. *Two random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ have same distribution if and only if*

$$\Psi_X(t) = \Psi_Y(t), \qquad t \in \mathbb{R}.$$

Theorem 1.12 is used to identify or to determine the probability distribution of a random variable X, by comparison with the characteristic function Ψ_Y of a random variable Y whose distribution is known.

The characteristic function of a random vector (X, Y) is the function $\Psi_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{C}$ defined by

$$\Psi_{X,Y}(s, t) = \mathbb{E}\big[e^{isX + itY}\big], \qquad s, t \in \mathbb{R}.$$

Theorem 1.13. *The random variables $X : \Omega \longrightarrow \mathbb{R}$ and $Y : \Omega \longrightarrow \mathbb{R}$ are* independent *if and only if*

$$\Psi_{X,Y}(s,t) = \Psi_X(s)\Psi_Y(t), \qquad s,t \in \mathbb{R}.$$

A random variable X has a Gaussian distribution with mean μ and variance σ^2 if and only if its characteristic function satisfies

$$\mathbb{E}\big[e^{i\alpha X}\big] = e^{i\alpha\mu - \alpha^2\sigma^2/2}, \qquad \alpha \in \mathbb{R}. \tag{1.7.1}$$

From Theorems 1.12 and 1.13, we deduce the following proposition.

Proposition 1.14. *Let $X \simeq \mathcal{N}(\mu, \sigma_X^2)$ and $Y \simeq \mathcal{N}(\nu, \sigma_Y^2)$ be independent Gaussian random variables. Then $X + Y$ also has a Gaussian distribution*

$$X + Y \simeq \mathcal{N}(\mu + \nu, \sigma_X^2 + \sigma_Y^2).$$

Proof. Since X and Y are independent, by Theorem 1.13 the characteristic function Ψ_{X+Y} of $X + Y$ is given by

$$\begin{aligned}\Phi_{X+Y}(t) &= \Phi_X(t)\Phi_Y(t)\\ &= e^{it\mu - t^2\sigma_X^2/2}e^{it\nu - t^2\sigma_Y^2/2}\\ &= e^{it(\mu+\nu) - t^2(\sigma_X^2+\sigma_Y^2)/2}, \qquad t \in \mathbb{R},\end{aligned}$$

where we used (1.7.1). Consequently, the characteristic function of $X + Y$ is that of a Gaussian random variable with mean $\mu + \nu$ and variance $\sigma_X^2 + \sigma_Y^2$ and we conclude by Theorem 1.12. □

Moment generating functions

The *moment generating function* of a random variable X is the function $\Phi_X : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$\Phi_X(t) := \mathbb{E}\big[e^{tX}\big],$$

for t in a neighborhood of 0. In particular, we have

$$\mathbb{E}[X^n] = \frac{\partial^n}{\partial t^n}\Phi_X(0), \qquad n \geqslant 1,$$

provided that $\mathbb{E}[|X|^n] < \infty$, and

$$\Phi_X(t) = \mathbb{E}\big[e^{tX}\big] = \sum_{n\geqslant 0}\frac{t^n}{n!}\mathbb{E}[X^n],$$

provided that $\mathbb{E}\big[e^{t|X|}\big] < \infty$, $t \in \mathbb{R}$, and for this reason the moment generating function G_X characterizes the *moments* $\mathbb{E}[X^n]$ of $X : \Omega \longrightarrow \mathbb{N}$, $n \geqslant 0$.

The moment generating function Φ_X of a random variable X with probability density function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ satisfies

$$\Phi_X(t) = \int_{-\infty}^{\infty} e^{xt} \varphi(x) dx, \qquad t \in \mathbb{R}.$$

For example, the moment generating functions (MGF) of a Gaussian random variable X with mean μ and variance σ^2 is given by

$$\mathbb{E}\left[e^{\alpha X}\right] = e^{\alpha\mu + \alpha^2\sigma^2/2}, \qquad \alpha \in \mathbb{R}. \tag{1.7.2}$$

Note that in probability, the moment generating function is written as a *bilateral* transform defined using an integral from $-\infty$ to $+\infty$.

Probability generating functions

Consider

$$X : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$$

a *discrete* random variable possibly taking infinite values. The *probability generating function* of X is the *function*

$$\begin{aligned} G_X : [-1, 1] &\longrightarrow \mathbb{R} \\ s &\longmapsto G_X(s) \end{aligned}$$

defined by

$$G_X(s) := \mathbb{E}\left[s^X \mathbb{1}_{\{X<\infty\}}\right] = \sum_{n \geqslant 0} s^n \mathbb{P}(X = n), \qquad -1 \leqslant s \leqslant 1. \tag{1.7.3}$$

Note that the series summation in (1.7.3) is over the *finite* integers, which explains the presence of the truncating indicator function $\mathbb{1}_{\{X<\infty\}}$ inside the expectation in (1.7.3). If the random variable $X : \Omega \longrightarrow \mathbb{N}$ is almost surely finite, *i.e.* $\mathbb{P}(X < \infty) = 1$, we simply have

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{n \geqslant 0} s^n \mathbb{P}(X = n), \qquad -1 \leqslant s \leqslant 1,$$

and for this reason the probability generating function G_X characterizes the *probability distribution* $\mathbb{P}(X = n)$, $n \geqslant 0$, of $X : \Omega \longrightarrow \mathbb{N}$.

Examples

i) Poisson distribution. Consider a random variable X with probability generating function

$$G_X(s) = e^{\lambda(s-1)}, \qquad -1 \leqslant s \leqslant 1,$$

for some $\lambda > 0$. What is the distribution of X?

Using the exponential series (A.1) we have

$$G_X(s) = \mathrm{e}^{\lambda(s-1)} = \mathrm{e}^{-\lambda} \sum_{n \geqslant 0} s^n \frac{\lambda^n}{n!}, \qquad -1 \leqslant s \leqslant 1, \tag{1.7.4}$$

hence by identification with (1.7.3) we find

$$\mathbb{P}(X = n) = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}, \qquad n \geqslant 0,$$

i.e. X has the Poisson distribution with parameter λ.

ii) Geometric distribution. Given X a random variable with geometric distribution $\mathbb{P}(X = n) = (1-p)p^n$, $n \geqslant 0$, we have

$$G_X(s) = \sum_{n \geqslant 0} s^n \mathbb{P}(X = n) = (1-p) \sum_{n \geqslant 0} s^n p^n = \frac{1-p}{1-ps}, \qquad -1 < s < 1,$$

where we applied the geometric series (A.3).

We note that from (1.7.3) we can write

$$G_X(s) = \mathbb{E}[s^X], \qquad -1 < s < 1,$$

since $s^X = s^X \mathbb{1}_{\{X < \infty\}}$ when $-1 < s < 1$.

Properties of probability generating functions

i) Taking $s = 1$, we have

$$G_X(1) = \sum_{n \geqslant 0} \mathbb{P}(X = n) = \mathbb{P}(X < \infty) = \mathbb{E}\big[\mathbb{1}_{\{X < \infty\}}\big],$$

hence

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

ii) Taking $s = 0$, we have

$$G_X(0) = \mathbb{E}\big[0^X\big] = \mathbb{E}\big[\mathbb{1}_{\{X=0\}}\big] = \mathbb{P}(X = 0),$$

since $0^0 = 1$ and $0^X = \mathbb{1}_{\{X=0\}}$, hence

$$G_X(0) = \mathbb{P}(X = 0). \tag{1.7.5}$$

iii) The *derivative** $G'_X(s)$ of $G_X(s)$ with respect to s satisfies

$$G'_X(s) = \sum_{n \geqslant 1} n s^{n-1} \mathbb{P}(X = n), \qquad -1 < s < 1,$$

hence if $\mathbb{P}(X < \infty) = 1$ we have

$$G'_X(1^-) = \mathbb{E}[X] = \sum_{k \geqslant 0} k \mathbb{P}(X = k), \qquad (1.7.6)$$

provided that $\mathbb{E}[X] < \infty$.

iv) By computing the second derivative

$$\begin{aligned} G''_X(s) &= \sum_{k \geqslant 2} k(k-1) s^{k-2} \mathbb{P}(X = k) \\ &= \sum_{k \geqslant 0} k(k-1) s^{k-2} \mathbb{P}(X = k) \\ &= \sum_{k \geqslant 0} k^2 s^{k-2} \mathbb{P}(X = k) - \sum_{k \geqslant 0} k s^{k-2} \mathbb{P}(X = k), \qquad -1 < s < 1, \end{aligned}$$

we similarly find

$$\begin{aligned} G''_X(1^-) &= \sum_{k \geqslant 0} k(k-1) \mathbb{P}(X = k) \\ &= \sum_{k \geqslant 0} k^2 \mathbb{P}(X = k) - \sum_{k \geqslant 0} k \mathbb{P}(X = k) \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X] \\ &= \mathbb{E}[X(X-1)], \end{aligned}$$

provided that $\mathbb{E}\left[X^2\right] < \infty$.

More generally, using the *k-th* derivative of G_X we can compute the *factorial moment*

$$G_X^{(k)}(1^-) = \mathbb{E}[X(X-1) \cdots (X-k+1)], \qquad k \geqslant 1, \qquad (1.7.7)$$

provided that $\mathbb{E}[|X^k|] < \infty$. In particular, we have

$$\mathrm{Var}[X] = G''_X(1^-) + G'_X(1^-)(1 - G'_X(1^-)), \quad (1.7.8)$$

* Here $G'_X(1^-)$ denotes the derivative on the left at the point $s = 1$.

provided that $\mathbb{E}[X^2] < \infty$.

v) When $X : \Omega \longrightarrow \mathbb{N}$ and $Y : \Omega \longrightarrow \mathbb{N}$ are two finite independent random variables we have

$$G_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y] = G_X(s)G_Y(s), \tag{1.7.9}$$

$-1 \leqslant s \leqslant 1$.

vi) The probability generating function can also be used from (1.7.3) to recover the distribution of the discrete random variable X as

$$\mathbb{P}(X = n) = \frac{1}{n!}\frac{\partial^n}{\partial s^n} G_X(s)_{|s=0}, \qquad n \geqslant 0, \tag{1.7.10}$$

extending (1.7.5) to all $n \geqslant 0$.

Exercise: Show that the probability generating function of a Poisson random variable X with parameter $\lambda > 0$ is given by

$$G_X(s) = \mathrm{e}^{\lambda(s-1)}, \qquad -1 \leqslant s \leqslant 1.$$

From the generating function we also recover the mean

$$\mathbb{E}[X] = G'_X(1^-) = \lambda \mathrm{e}^{\lambda(s-1)}_{|s=1} = \lambda,$$

of the Poisson random variable X with parameter λ, and its variance

$$\begin{aligned}
\mathrm{Var}[X] &= G''_X(1^-) + G'_X(1^-) - (G'_X(1^-))^2 \\
&= \lambda^2 \mathrm{e}^{\lambda(s-1)}_{|s=1} + \lambda s \mathrm{e}^{\lambda(s-1)}_{|s=1} - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda,
\end{aligned}$$

by (1.7.8).

Moments and cumulants

The *cumulants* of a random variable X are the sequence $(\kappa_n^X)_{n\in\mathbb{N}}$ of numbers defined by the logarithmic generating function

$$\log\left(\mathbb{E}\left[e^{tX}\right]\right)=\sum_{n\geqslant 1}\kappa_n^X\frac{t^n}{n!},\qquad t\in\mathbb{R}.$$

The *n-th* moment of a random variable X can be written in terms of its cumulants as

$$\mathbb{E}[X^n]=\sum_{k=0}^{n}\frac{1}{k!}\sum_{\substack{d_1+\cdots+d_k=n\\ d_1,\ldots,d_k\geqslant 1}}\frac{n!}{d_1!\cdots d_k!}\kappa_{d_1}^X\cdots\kappa_{d_k}^X \tag{1.7.11}$$

$$=\sum_{k=1}^{n}\sum_{B_1,\ldots,B_k}\kappa_{|B_1|}^X\cdots\kappa_{|B_k|}^X,$$

where the above sum runs over the partitions $B_1,\ldots,B_k$ of $\{1,\ldots,n\}$ with cardinality $|B_i|$ (Faà di Bruno formula). This also shows that

$$\mathbb{E}[X^n]=A_n(\kappa_1^X,\kappa_2^X,\ldots,\kappa_n^X),$$

where

$$A_n(x_1,\ldots,x_n)=n!\sum_{\substack{r_1+2r_2+\cdots+nr_n=n\\ r_1,\ldots,r_n\geqslant 0}}\prod_{l=1}^{n}\left(\frac{1}{r_l!}\left(\frac{x_l}{l!}\right)^{r_l}\right)$$

is the Bell polynomial of degree n. These relations follow from the Faà di Bruno formula, cf. *e.g.* § 2.4 and Relation (2.4.4) page 27 of Lukacs (1970). Indeed we have

$$\begin{aligned}
&\sum_{n\geqslant 0}\frac{t^n}{n!}\mathbb{E}[X^n]=\mathbb{E}\left[e^{tX}\right]\\
&=\exp\left(\log\left(\mathbb{E}\left[e^{tX}\right]\right)\right)=\sum_{k\geqslant 0}\frac{1}{k!}\left(\log\left(\mathbb{E}\left[e^{tX}\right]\right)\right)^k\\
&=\sum_{k\geqslant 0}\frac{1}{k!}\left(\sum_{n\geqslant 1}\kappa_n^X\frac{t^n}{n!}\right)^k\\
&=\sum_{k\geqslant 0}\frac{1}{k!}\sum_{n\geqslant 1}t^n\sum_{\substack{d_1+\cdots+d_k=n\\ d_1\geqslant 1,\ldots,d_k\geqslant 1}}\frac{\kappa_{d_1}^X}{d_1!}\cdots\frac{\kappa_{d_1}^X}{d_k!}\\
&=\sum_{n\geqslant 0}\frac{t^n}{n!}\sum_{k=0}^{n}\frac{1}{k!}\sum_{\substack{d_1+\cdots+d_k=n\\ d_1\geqslant 1,\ldots,d_k\geqslant 1}}\frac{n!}{d_1!\cdots d_k!}\frac{\kappa_{d_1}^X}{d_1!}\cdots\frac{\kappa_{d_1}^X}{d_k!},
\end{aligned}$$

which shows (1.7.11).

1.8 Conditional Expectation

The construction of conditional expectations of the form $\mathbb{E}[X \mid Y]$ given above for discrete and absolutely continuous random variables can be generalized to σ-algebras.

Definition 1.15. *Given $\mathcal{F}$ a σ-algebra on Ω, a random variable $X : \Omega \longrightarrow \mathbb{R}$ is said to be $\mathcal{F}$-measurable if*

$$\{X \leqslant x\} := \{\omega \in \Omega \ : \ X(\omega) \leqslant x\} \in \mathcal{F},$$

for all $x \in \mathbb{R}$.

Intuitively, when X is $\mathcal{F}$-measurable, the knowledge of the values of X depends only on the information contained in $\mathcal{F}$. For example, when $\mathcal{F} = \sigma(A_1, \ldots, A_n)$ where $(A_n)_{n\geqslant 1}$ is a *partition* of Ω with $\bigcup_{n\geqslant 1} A_n = \Omega$, any $\mathcal{F}$-measurable random variable X can be written as

$$X(\omega) = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}(\omega), \qquad \omega \in \Omega,$$

for some $c_1, \ldots, c_n \in \mathbb{R}$.

Definition 1.16. *Given $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space we let $L^2(\Omega, \mathcal{F})$ denote the space of $\mathcal{F}$-measurable and square-integrable random variables, i.e.*

$$L^2(\Omega, \mathcal{F}) := \big\{X : \Omega \longrightarrow \mathbb{R} \ : \ \mathbb{E}\big[|X|^2\big] < \infty\big\}.$$

More generally, for $p \geqslant 1$ one can define the space $L^p(\Omega, \mathcal{F})$ of $\mathcal{F}$-measurable and p-integrable random variables as

$$L^p(\Omega, \mathcal{F}) := \big\{X : \Omega \longrightarrow \mathbb{R} \ : \ \mathbb{E}[|X|^p] < \infty\big\}.$$

We define an *inner product* $\langle \cdot, \cdot \rangle_{L^2(\Omega,\mathcal{F})}$ between elements of $L^2(\Omega, \mathcal{F})$, as

$$\langle X, Y \rangle_{L^2(\Omega,\mathcal{F})} := \mathbb{E}[XY], \quad X, Y \in L^2(\Omega, \mathcal{F}). \tag{1.8.1}$$

This inner product is associated to the norm $\| \cdot \|_{L^2(\Omega)}$ by the relation

$$\|X\|_{L^2(\Omega)} = \sqrt{\mathbb{E}\big[X^2\big]} = \sqrt{\langle X, X \rangle_{L^2(\Omega,\mathcal{F})}}, \qquad X \in L^2(\Omega, \mathcal{F}).$$

The norm $\| \cdot \|_{L^2(\Omega)}$ also defines the *mean-square* distance

$$\|X - Y\|_{L^2(\Omega)} = \sqrt{\mathbb{E}\big[(X - Y)^2\big]}$$

between random variables $X, Y \in L^2(\Omega, \mathcal{F})$, and it induces a notion of *orthogonality*, namely X is *orthogonal* to Y in $L^2(\Omega, \mathcal{F})$ if and only if

$$\langle X, Y \rangle_{L^2(\Omega,\mathcal{F})} = 0.$$

Proposition 1.17. *The ordinary expectation* $\mathbb{E}[X]$ *achieves the minimum distance*

$$\left\| X - \mathbb{E}[X] \right\|^2_{L^2(\Omega)} = \min_{c\in\mathbb{R}} \| X - c \|^2_{L^2(\Omega)}. \tag{1.8.2}$$

Proof. It suffices to differentiate

$$\frac{\partial}{\partial c}\mathbb{E}\left[(X-c)^2\right] = -2\mathbb{E}[X-c] = 0,$$

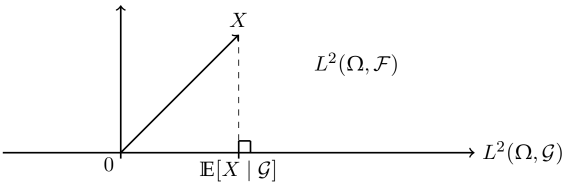
showing that the minimum in (1.8.2) is reached when $\mathbb{E}[X - c] = 0$, *i.e.* $c = \mathbb{E}[X]$. □

Similarly to Proposition 1.17, the conditional expectation will be defined by a distance minimizing procedure.

Definition 1.18. *Given* $\mathcal{G} \subset \mathcal{F}$ *a sub* σ*-algebra of* $\mathcal{F}$ *and* $X \in L^2(\Omega, \mathcal{F})$, *the conditional expectation of* X *given* $\mathcal{G}$, *and denoted*

$$\mathbb{E}[X \mid \mathcal{G}],$$

is defined as the orthogonal projection *of* X *onto* $L^2(\Omega, \mathcal{G})$.



As a consequence of the uniqueness of the orthogonal projection onto the subspace $L^2(\Omega, \mathcal{G})$ of $L^2(\Omega, \mathcal{F})$, the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is characterized by the relation

$$\langle Y, X - \mathbb{E}[X \mid \mathcal{G}] \rangle_{L^2(\Omega,\mathcal{F})} = 0,$$

which rewrites as

$$\mathbb{E}[Y(X - \mathbb{E}[X \mid \mathcal{G}])] = 0,$$

i.e.

$$\mathbb{E}[YX] = \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]],$$

for all bounded and $\mathcal{G}$-measurable random variables Y, where $\langle \cdot , \cdot \rangle_{L^2(\Omega,\mathcal{F})}$ denotes the inner product (1.8.1) in $L^2(\Omega, \mathcal{F})$. The next proposition extends Proposition 1.17 as a consequence of Definition 1.18. See Theorem 5.1.4 page 197 of Stroock (2011) for an extension of the construction of conditional expectation to the space $L^1(\Omega, \mathcal{F})$ of integrable random variable.

Proposition 1.19. *The conditional expectation* $\mathbb{E}[X \mid \mathcal{G}]$ *realizes the minimum in mean-square distance between* $X \in L^2(\Omega, \mathcal{F})$ *and* $L^2(\Omega, \mathcal{G})$*, i.e. we have*

$$\|X - \mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\Omega)} = \min_{Y \in L^2(\Omega,\mathcal{G})} \|X - Y\|_{L^2(\Omega)}. \tag{1.8.3}$$

Proof. This is a consequence of the Pythagorean theorem written as

$$\|X - Y\|_{L^2(\Omega)} = \|X - \mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\Omega)} + \|\mathbb{E}[X \mid \mathcal{G}] - Y\|_{L^2(\Omega)},$$

for any $Y \in L^2(\Omega, \mathcal{G})$. □

The following proposition will often be used as a characterization of $\mathbb{E}[X \mid \mathcal{G}]$.

Proposition 1.20. *Given* $X \in L^2(\Omega, \mathcal{F})$*,* $Z := \mathbb{E}[X \mid \mathcal{G}]$ *is the unique random variable* Z *in* $L^2(\Omega, \mathcal{G})$ *that satisfies the relation*

$$\mathbb{E}[YX] = \mathbb{E}[YZ] \tag{1.8.4}$$

for all bounded and $\mathcal{G}$*-measurable random variables* Y*.*

We note that taking $Y = \mathbf{1}$ in (1.8.4) yields

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]. \tag{1.8.5}$$

In particular, when $\mathcal{G} = \{\emptyset, \Omega\}$ we have $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid \{\emptyset, \Omega\}]$, and

$$\mathbb{E}[X \mid \{\emptyset, \Omega\}] = \mathbb{E}[\mathbb{E}[X \mid \{\emptyset, \Omega\}]] = \mathbb{E}[X], \tag{1.8.6}$$

because $\mathbb{E}[X \mid \{\emptyset, \Omega\}]$ is in $L^2(\Omega, \{\emptyset, \Omega\})$ and is *a.s.* constant. In addition, the conditional expectation operator has the following properties.

i) $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$ if Y depends only on the information contained in $\mathcal{G}$.

   Proof. By the characterization (1.8.4) it suffices to show that

$$\mathbb{E}[H(XY)] = \mathbb{E}[H(Y\mathbb{E}[X|\mathcal{G}])], \tag{1.8.7}$$

   for all bounded and $\mathcal{G}$-measurable random variables H, which implies $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$.

   Relation (1.8.7) holds from (1.8.4) because the product HY is $\mathcal{G}$-measurable hence Y in (1.8.4) can be replaced with HY.

ii) $\mathbb{E}[Y|\mathcal{G}] = Y$ when Y depends only on the information contained in $\mathcal{G}$.

Proof. This is a consequence of point (i) above by taking $X := \mathbf{1}$.

iii) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ if $\mathcal{H} \subset \mathcal{G}$, called the *tower property*.

Proof. First, we note that by (1.8.5), (iii) holds when $\mathcal{H} = \{\emptyset, \Omega\}$. Next, by the characterization (1.8.4) it suffices to show that

$$\mathbb{E}[H\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[H\mathbb{E}[X|\mathcal{H}]], \tag{1.8.8}$$

for all bounded and $\mathcal{H}$-measurable random variables H, which will imply (iii) from (1.8.4).

In order to prove (1.8.8) we check that by point (i) above and (1.8.5) we have

$$\begin{aligned}\mathbb{E}[H\mathbb{E}[X|\mathcal{G}]] &= \mathbb{E}[\mathbb{E}[HX|\mathcal{G}]] = \mathbb{E}[HX] \\ &= \mathbb{E}[\mathbb{E}[HX|\mathcal{H}]] = \mathbb{E}[H\mathbb{E}[X|\mathcal{H}]],\end{aligned}$$

and we conclude by the characterization (1.8.4).

iv) $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ when X "does not depend" on the information contained in $\mathcal{G}$ or, more precisely stated, when the random variable X is *independent* of the σ-algebra $\mathcal{G}$.

Proof. It suffices to note that for all bounded $\mathcal{G}$-measurable Y we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[Y\mathbb{E}[X]],$$

and we conclude again by (1.8.4).

v) If Y depends only on $\mathcal{G}$ and X is independent of $\mathcal{G}$, then

$$\mathbb{E}[h(X,Y)|\mathcal{G}] = \mathbb{E}[h(X,x)]_{x=Y}. \tag{1.8.9}$$

Proof. This relation can be proved using the tower property, by noting that for any bounded $Z \in L^2(\Omega, \mathcal{G})$ we have

$$\begin{aligned}\mathbb{E}[Z\mathbb{E}[h(x,X)]_{x=Y}] &= \mathbb{E}[Z\mathbb{E}[h(x,X) \mid \mathcal{G}]_{x=Y}] \\ &= \mathbb{E}[Z\mathbb{E}[h(Y,X) \mid \mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[Zh(Y,X) \mid \mathcal{G}]] \\ &= \mathbb{E}[Zh(Y,X)],\end{aligned}$$

which yields (1.8.9) by the characterization (1.8.4).

The notion of conditional expectation can be extended from square-integrable random variables in $L^2(\Omega, \mathcal{F})$ to integrable random variables in $L^1(\Omega, \mathcal{F})$, see *e.g.* Theorem 5.1 in Kallenberg (2002).

Proposition 1.21. *When the σ-algebra $\mathcal{G} := \sigma(A_1, A_2, \ldots, A_n)$ is generated by n disjoint events $A_1, A_2, \ldots, A_n \in \mathcal{F}$, we have*

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{k=1}^n \mathbb{1}_{A_k} \mathbb{E}[X \mid A_k] = \sum_{k=1}^n \mathbb{1}_{A_k} \frac{\mathbb{E}[X \mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}.$$

Proof. It suffices to note that the $\mathcal{G}$-measurable random variables can be generated by indicator functions of the form $\mathbb{1}_{A_l}$, and that

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{A_l} \sum_{k=1}^n \mathbb{1}_{A_k} \frac{\mathbb{E}[X \mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}\right] &= \mathbb{E}\left[\mathbb{1}_{A_l} \frac{\mathbb{E}[X \mathbb{1}_{A_l}]}{\mathbb{P}(A_l)}\right] \\ &= \frac{\mathbb{E}[X \mathbb{1}_{A_l}]}{\mathbb{P}(A_l)} \mathbb{E}\big[\mathbb{1}_{A_l}\big] \\ &= \mathbb{E}\big[X \mathbb{1}_{A_l}\big], \qquad l = 1, 2, \ldots, n,\end{aligned}$$

showing (1.8.4). The relation

$$\mathbb{E}[X \mid A_k] = \frac{\mathbb{E}[X \mathbb{1}_{A_k}]}{\mathbb{P}(A_k)}, \qquad k = 1, 2, \ldots, n,$$

follows from Lemma 1.10. □

For example, in case $\Omega = \{a, b, c, d\}$ and $\mathcal{G} = \big\{\emptyset, \Omega, \{a, b\}, \{c\}, \{d\}\big\}$, we have

$$\begin{aligned}\mathbb{E}[X \mid \mathcal{G}] &= \mathbb{1}_{\{a,b\}} \mathbb{E}[X \mid \{a, b\}] + \mathbb{1}_{\{c\}} \mathbb{E}[X \mid \{c\}] + \mathbb{1}_{\{d\}} \mathbb{E}[X \mid \{d\}] \\ &= \mathbb{1}_{\{a,b\}} \frac{\mathbb{E}\big[X \mathbb{1}_{\{a,b\}}\big]}{\mathbb{P}(\{a, b\})} + \mathbb{1}_{\{c\}} \frac{\mathbb{E}\big[X \mathbb{1}_{\{c\}}\big]}{\mathbb{P}(\{c\})} + \mathbb{1}_{\{d\}} \frac{\mathbb{E}\big[X \mathbb{1}_{\{d\}}\big]}{\mathbb{P}(\{d\})}.\end{aligned}$$

Regarding conditional probabilities we have similarly, for $A \subset \Omega = \{a, b, c, d\}$,

$$\begin{aligned}\mathbb{P}(A \mid \mathcal{G}) &= \mathbb{1}_{\{a,b\}} \frac{\mathbb{P}\big(A \cap \{a, b\}\big)}{\mathbb{P}(\{a, b\})} + \mathbb{1}_{\{c\}} \frac{\mathbb{P}\big(A \cap \{c\}\big)}{\mathbb{P}(\{c\})} + \mathbb{1}_{\{d\}} \frac{\mathbb{P}\big(A \cap \{d\}\big)}{\mathbb{P}(\{d\})} \\ &= \mathbb{1}_{\{a,b\}} \mathbb{P}(A \mid \{a, b\}) + \mathbb{1}_{\{c\}} \mathbb{P}(A \mid \{c\}) + \mathbb{1}_{\{d\}} \mathbb{P}(A \mid \{d\}).\end{aligned}$$

In particular, if $A = \{a\} \subset \Omega = \{a, b, c, d\}$ we find

$$\begin{aligned}\mathbb{P}(\{a\} \mid \mathcal{G}) &= \mathbb{1}_{\{a,b\}} \mathbb{P}(\{a\} \mid \{a, b\}) \\ &= \mathbb{1}_{\{a,b\}} \frac{\mathbb{P}\big(\{a\} \cap \{a, b\}\big)}{\mathbb{P}(\{a, b\})} \\ &= \mathbb{1}_{\{a,b\}} \frac{\mathbb{P}(\{a\})}{\mathbb{P}(\{a, b\})}.\end{aligned}$$

In other words, the probability of getting the outcome a is $\mathbb{P}(\{a\})/\mathbb{P}(\{a,b\})$ knowing that the outcome is either a or b, otherwise it is zero.

Exercises

Exercise 1.1 Prove the equivalence between the two versions (1.2.2)-(1.2.3) of the inclusion-exclusion principle.

Exercise 1.2 Consider the random variable $X : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$ with distribution

$$\mathbb{P}(X = k) = qp^k, \qquad k \in \mathbb{N} = \{0, 1, 2, \ldots\},$$

where $q \in [0, 1 - p]$ and $0 \leqslant p < 1$.

a) Compute $\mathbb{P}(X < \infty)$ and $\mathbb{P}(X = \infty)$ by considering two cases, and give the value of $\mathbb{E}[X]$ when $0 \leqslant q < 1 - p$.
b) Assume that $q = 1 - p$ and consider the random variable $Y := r^X$ for some $r > 0$. Explain why $\mathbb{P}(Y < \infty) = 1$ and compute $\mathbb{E}[Y]$ by considering two cases depending on the value of $r > 0$.

Exercise 1.3 Let $N \in \{1, 2, 3, 4, 5, 6\}$ denote the integer random variable obtained by tossing a six faced die and by noting the number on the upper side of the die. Given the value of N, an *independent*, unbiased coin is thrown N times. We denote by Z the total number of heads that appear in the process of throwing the coin N times.

a) Using conditioning on the value of $N \in \{1, 2, 3, 4, 5, 6\}$, compute the mean and the variance of the random variable Z.
b) Determine the probability distribution of Z.
c) Recover the result of Question (a) from the data of the probability distribution computed in Question (b).

Exercise 1.4 Thinning of Poisson random variables. Given a random sample N of a Poisson random variable with parameter λ, we perform a number N of *independent* $\{0, 1\}$-valued Bernoulli experiments *independent* of N, each of them with parameter $p \in (0, 1)$. We let Z denote the total number of $+1$ outcomes occurring in the N Bernoulli trials.

a) Express Z as a random sum, and use this expression to compute the mean and variance of Z.
b) Compute the probability distribution of Z.

c) Recover the result of Question (a) from the data of the probability distribution computed in Question (b).

Exercise 1.5 Given X and Y two independent exponentially distributed random variables with parameters λ and μ, show the relation

$$\mathbb{E}[\min(X,Y) \mid X<Y] = \frac{1}{\lambda+\mu} = \mathbb{E}[\min(X,Y)]. \tag{1.8.10}$$

Exercise 1.6 Given a random sample L of a gamma random variable with probability density function

$$\varphi_L(x) = \mathbb{1}_{[0,\infty)} x \mathrm{e}^{-x},$$

consider U a uniform random variable taking values in the interval $[0,L]$ and let $V = L-U$.

Compute the joint probability density function of the couple (U,V) of random variables.

Exercise 1.7 Let X and Y denote two independent Poisson random variables with parameters λ and μ.

a) Show that the random variable $X+Y$ has the Poisson distribution with parameter $\lambda+\mu$.
b) Compute the conditional distribution $\mathbb{P}(X=k \mid X+Y=n)$ given that $X+Y=n$, for all $k,n \in \mathbb{N}$.
c) Assume that respective parameters of the distributions of X and Y are random, independent, and chosen according to an exponential distribution with parameter $\theta>0$.

   Give the probability distributions of X and Y, and compute the conditional distribution $\mathbb{P}(X=k \mid X+Y=n)$ given that $X+Y=n$, for all $k,n \in \mathbb{N}$.
d) Assume now that X and Y have same random parameter represented by a single exponentially distributed random variable Λ with parameter $\theta>0$, independent of X and Y.

   Compute the conditional distribution $\mathbb{P}(X=k \mid X+Y=n)$ given that $X+Y=n$, for all $k,n \in \mathbb{N}$.

Exercise 1.8 A red pen and a green pen are put in a hat. A pen is chosen at random in the hat, and replaced inside after it color has been noted.

- In case the pen is of red color, then a supplementary red pen is placed in the hat.

- On the other hand if the pen color is green, then another green pen is added.

After this first part of the experiment is completed, a second pen is chosen at random.

Determine the probability that the first drawn pen was red, given that the color of the second pen chosen was red.

Exercise 1.9 A machine relies on the functioning of three parts, each of which having a probability $1-p$ of being under failure, and a probability p of functioning correctly. All three parts are functioning independently of the others, and the machine is working if and only if two at least of the parts are operating.

a) Compute the probability that the machine is functioning.
b) Suppose that the machine itself is set in a random environment in which the value of the probability p becomes random. Precisely we assume that p is a uniform random variable taking real values between 0 and 1, independently of the state of the system.

   Compute the probability that the machine operates in this random environment.

Chapter 2
Gambling Problems

This chapter consists in a detailed study of a fundamental example of random walk that can only evolve by going up of down by one unit within the finite state space $\mathbb{S} = \{0, 1, \ldots, S\}$. This allows us in particular to have a first look at the first step analysis technique that will be repeatedly used in the general framework of Markov chains, particularly in Chapter 5.

2.1 Constrained Random Walk

To begin, let us repeat that this chapter on "gambling problems" is not primarily designed to help a reader dealing with problem gambling, although some comments on this topic are made of the end of Section 2.3.

We consider an amount $\$S$ of S dollars which is to be shared between two players A and B. At each round, Player A may earn \$1 with probability $p \in (0, 1)$, and in this case Player B loses \$1. Conversely, Player A may lose \$1 with probability $q := 1 - p$, in which case Player B gains \$1, and the successive rounds are independent.

We let X_n represent the wealth of Player A at time $n \in \mathbb{N}$, while $S - X_n$ represents the wealth of Player B at time $n \in \mathbb{N}$.

The initial wealth X_0 of Player A could be negative,* but for simplicity we will assume that it is comprised between 0 and S. Assuming that the value of X_n, $n \geqslant 0$, belongs to $\{1, 2, \ldots, S-1\}$ at the time step n, at the next step $n+1$ we will have

$$X_{n+1} = \begin{cases} X_n + 1 & \text{if Player A wins round } n+1, \\ \\ X_n - 1 & \text{if Player B wins round } n+1. \end{cases}$$

Moreover, as soon as X_n hits one of the boundary points $\{0, S\}$, the process remains frozen at that state over time, *i.e.*

$$(X_n = 0) \Longrightarrow (X_{n+1} = 0) \quad \text{and} \quad (X_n = S) \Longrightarrow (X_{n+1} = S),$$

or

$$\mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = 1 \quad \text{and} \quad \mathbb{P}(X_{n+1} = S \mid X_n = S) = 1, \quad n \geqslant 0.$$

In other words, the game ends whenever the wealth of any of the two players reaches \$0, in which case the other player's account contains \$$S$, see Figure 2.1.

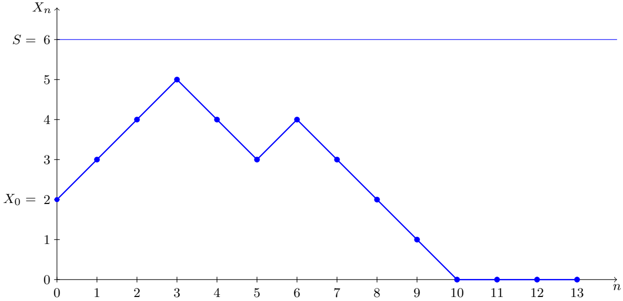


Fig. 2.1: Sample path of a gambling process $(X_n)_{n\geqslant 0}$.

Among the main issues of interest are:

- the probability that Player A (or B) gets eventually ruined,

- the mean duration of the game.

* Daily entry levy: Singaporeans and permanent residents may have to start with $X_0 = -\$150$.

We will also be interested in the probability distribution of the random game duration T, *i.e.* in the knowledge of $\mathbb{P}(T = n)$, $n \geqslant 0$.

According to the above problem description, for all $n \in \mathbb{N}$ we have

$$\mathbb{P}(X_{n+1} = k+1 \mid X_n = k) = p \quad \text{and} \quad \mathbb{P}(X_{n+1} = k-1 \mid X_n = k) = q,$$

$k = 1, 2, \ldots, S-1$, and in this case the chain is said to be *time homogeneous* since the transition probabilities do not depend on the time index n.

Since we do not focus on the behavior of the chain after it hits states 0 or S, the probability distribution of X_{n+1} given $\{X_n = 0\}$ or $\{X_n = S\}$ can be left unspecified.

The probability sample space Ω corresponding to this experiment could be taken as the (uncountable) set

$$\Omega := \{-1, +1\}^{\mathbb{N}} = \big\{\omega = (\omega_0, \omega_1, \ldots) \ : \ \omega_i = \pm 1, \ n \in \mathbb{N}\big\},$$

with any element $\omega \in \Omega$ represented by a countable sequence of $+1$ or -1, depending whether the process goes up or down at each time step. However, in what follows we will not focus on this particular expression of Ω.

2.2 Ruin Probabilities

We are interested in the event

$$R_A = \text{“Player } A \text{ loses all his capital at some time”} = \bigcup_{n \in \mathbb{N}} \{X_n = 0\}, \tag{2.2.1}$$

and in computing the conditional probability

$$f_S(k) := \mathbb{P}(R_A \mid X_0 = k), \qquad k = 0, 1, \ldots, S. \tag{2.2.2}$$

Pathwise analysis

First, let us note that the problem is easy to solve in the cases $S = 1$, $S = 2$ and $S = 3$.

i) $\underline{S = 1}$.

In this case the boundary $\{0, 1\}$ is reached from time 0 and we find

$$\begin{cases} f_1(0) = \mathbb{P}(R_A \mid X_0 = 0) = 1, \\ f_1(1) = \mathbb{P}(R_A \mid X_0 = 1) = 0. \end{cases} \tag{2.2.3}$$

ii) $\underline{S = 2}$.

In this case we find

$$\begin{cases} f_2(0) = \mathbb{P}(R_A \mid X_0 = 0) = 1, \\ f_2(1) = \mathbb{P}(R_A \mid X_0 = 1) = q, \\ f_2(2) = \mathbb{P}(R_A \mid X_0 = 2) = 0. \end{cases} \tag{2.2.4}$$

iii) $\underline{S = 3}$.

The value of $f_3(1) = \mathbb{P}(R_A \mid X_0 = 1)$ is computed by noting that starting from state ⓵, one can reach state ⓪ only by an odd number $2n+1$ of steps, $n \in \mathbb{N}$, and that every such path decomposes into $n+1$ independent downwards steps, each of them having probability q, and n upwards steps, each of them with probability p. By summation over n using the geometric series identity (A.3), this yields

$$\begin{cases} f_3(0) = \mathbb{P}(R_A \mid X_0 = 0) = 1, \\ f_3(1) = \mathbb{P}(R_A \mid X_0 = 1) = q \sum_{n \geqslant 0} (pq)^n = \dfrac{q}{1-pq}, \\ f_3(2) = \mathbb{P}(R_A \mid X_0 = 2) = q^2 \sum_{n \geqslant 0} (pq)^n = \dfrac{q^2}{1-pq}, \\ f_3(3) = \mathbb{P}(R_A \mid X_0 = 3) = 0. \end{cases} \tag{2.2.5}$$

The value of $f_3(2)$ is computed similarly by considering $n+2$ independent downwards steps, each of them with probability q, and n upwards steps, each of them with probability p. Clearly, things become quite complicated for $S \geqslant 4$, and increasingly difficult as S gets larger.

First step analysis

The general case will be solved by the method of *first step analysis*, which will be repeatedly applied to other Markov processes in Chapters 3 and 5 and elsewhere.

Lemma 2.1. *For all $k = 1, 2, \ldots, S-1$ we have*

$$\mathbb{P}(R_A \mid X_0 = k) = p\mathbb{P}(R_A \mid X_0 = k+1) + q\mathbb{P}(R_A \mid X_0 = k-1).$$

Proof. The idea is to apply conditioning given the first transition from X_0 to X_1. For all $k = 1, 2, \ldots, S-1$, by (1.2.1) we have

$$
\begin{aligned}
&\mathbb{P}(R_A \mid X_0 = k)\\
&= \mathbb{P}(R_A \text{ and } X_1 = k+1 \mid X_0 = k) + \mathbb{P}(R_A \text{ and } X_1 = k-1 \mid X_0 = k)\\
&= \frac{\mathbb{P}(R_A \text{ and } X_1 = k+1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)} + \frac{\mathbb{P}(R_A \text{ and } X_1 = k-1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)}\\
&= \frac{\mathbb{P}(R_A \text{ and } X_1 = k+1 \text{ and } X_0 = k)}{\mathbb{P}(X_1 = k+1 \text{ and } X_0 = k)} \times \frac{\mathbb{P}(X_1 = k+1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)}\\
&\quad + \frac{\mathbb{P}(R_A \text{ and } X_1 = k-1 \text{ and } X_0 = k)}{\mathbb{P}(X_1 = k-1 \text{ and } X_0 = k)} \times \frac{\mathbb{P}(X_1 = k-1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)}\\
&= \mathbb{P}(R_A \mid X_1 = k+1 \text{ and } \ X_0 = k)\mathbb{P}(X_1 = k+1 \mid X_0 = k)\\
&\quad + \mathbb{P}(R_A \mid X_1 = k-1 \text{ and } \ X_0 = k)\mathbb{P}(X_1 = k-1 \mid X_0 = k)\\
&= p\mathbb{P}(R_A \mid X_1 = k+1 \text{ and } X_0 = k) + q\mathbb{P}(R_A \mid X_1 = k-1 \text{ and } X_0 = k)\\
&= p\mathbb{P}(R_A \mid X_0 = k+1) + q\mathbb{P}(R_A \mid X_0 = k-1),
\end{aligned}
$$

where we used Lemma 2.2 below on the last step. □

In the case $S = 3$, Lemma 2.1 shows that

$$
\begin{cases}
f_3(0) = \mathbb{P}(R_A \mid X_0 = 0) = 1,\\
\\
f_3(1) = pf_3(2) + qf_3(0) = pf_3(2) + q = pqf_3(1) + q,\\
\\
f_3(2) = pf_3(3) + qf_3(1) = qf_3(1) = pqf_3(2) + q^2,\\
\\
f_3(3) = \mathbb{P}(R_A \mid X_0 = 3) = 0,
\end{cases}
$$

whose solution can be checked to be given by (2.2.5).

More generally, Lemma 2.1 shows that the function

$$
f_S : \{0, 1, \ldots, S\} \longrightarrow [0, 1]
$$

defined by (2.2.2) satisfies the linear equation*

* Due to the relation $(f + g)(k) = f(k) + g(k)$ we can check that if f and g are two solutions of (2.2.6) then $f + g$ is also a solution of (2.2.6), hence the equation is *linear*.

$$f_S(k) = pf_S(k+1) + qf_S(k-1), \qquad k = 1, 2, \ldots, S-1, \quad (2.2.6)$$

subject to the *boundary conditions*

$$f_S(0) = \mathbb{P}(R_A \mid X_0 = 0) = 1, \quad (2.2.7)$$

and

$$f_S(S) = \mathbb{P}(R_A \mid X_0 = S) = 0, \quad (2.2.8)$$

for $k \in \{0, S\}$. It can be easily checked that the expressions (2.2.3), (2.2.4) and (2.2.5) do satisfy the above Equation (2.2.6) and the boundary conditions (2.2.7) and (2.2.8).

Note that Lemma 2.1 is frequently stated without proof. The last step of the proof stated above relies on the following lemma, which shows that the data of X_1 entirely determines the probability of the ruin event R_A. In other words, the probability of ruin depends only on the initial amount k owned by the gambler when he enters the casino. Whether he enters the casino at time 1 with $X_1 = k \pm 1$ or at time 0 with $X_0 = k \pm 1$ makes no difference on the ruin probability.

Lemma 2.2. *For all $k = 1, 2, \ldots, S-1$ we have*

$$\mathbb{P}(R_A \mid X_1 = k \pm 1 \text{ and } X_0 = k) = \mathbb{P}(R_A \mid X_1 = k \pm 1) = \mathbb{P}(R_A \mid X_0 = k \pm 1).$$

In other words, the ruin probability depends on the data of the starting point and not on the starting time.

Proof. This relation can be shown in various ways:

1. Descriptive proof (*preferred*): we note that given $X_1 = k + 1$, the transition from X_0 to X_1 has no influence on the future of the process after time 1, and the probability of ruin starting at time 1 is the same as if the process is started at time 0.

2. Algebraic proof. First, for $1 < k \leqslant S - 1$ and $k \pm 1 \geqslant 1$, letting $\widetilde{X}_0 := X_1 - Z$ where $Z \simeq X_1 - X_0$ has same distribution as $X_1 - X_0$ and is independent of X_1, by (2.2.1) we have

$$\begin{aligned}
&\mathbb{P}(R_A \mid X_1 = k \pm 1 \text{ and } X_0 = k) \\
&= \mathbb{P}\left(\bigcup_{n \geqslant 0} \{X_n = 0\} \;\middle|\; X_1 = k \pm 1,\ X_0 = k \right) \\
&= \frac{\mathbb{P}\left(\left(\bigcup_{n \geqslant 0} \{X_n = 0\} \right) \cap \{X_1 = k \pm 1\} \cap \{X_0 = k\} \right)}{\mathbb{P}\left(\{X_1 = k \pm 1\} \cap \{X_0 = k\} \right)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{P}\left(\bigcup_{n\geqslant 0}\left(\{X_n=0\}\cap\{X_1=k\pm 1\}\cap\{X_0=k\}\right)\right)}{\mathbb{P}\left(\{X_1=k\pm 1\}\cap\{X_0=k\}\right)}\\
&= \frac{\mathbb{P}\left(\bigcup_{n\geqslant 2}\left(\{X_n=0\}\cap\{X_1=k\pm 1\}\cap\{X_0=k\}\right)\right)}{\mathbb{P}\left(\{X_1=k\pm 1\}\cap\{X_0=k\}\right)}\\
&= \frac{\mathbb{P}\left(\left(\bigcup_{n\geqslant 2}\{X_n=0\}\right)\cap\{X_1=k\pm 1\}\cap\{X_0=k\}\right)}{\mathbb{P}\left(\{X_1=k\pm 1\}\cap\{X_0=k\}\right)} \qquad (2.2.9)\\
&= \frac{\mathbb{P}\left(\left(\bigcup_{n\geqslant 2}\{X_n=0\}\right)\cap\{X_1=k\pm 1\}\cap\left\{X_1-\widetilde{X}_0=\pm 1\right\}\right)}{\mathbb{P}\big(\{X_1=k\pm 1\}\cap\{X_1-\widetilde{X}_0=\pm 1\}\big)} \qquad (2.2.10)\\
&= \frac{\mathbb{P}\left(\left(\bigcup_{n\geqslant 2}\{X_n=0\}\right)\cap\{X_1=k\pm 1\}\right)\mathbb{P}\big(\left\{X_1-\widetilde{X}_0=\pm 1\right\}\big)}{\mathbb{P}(\{X_1=k\pm 1\})\mathbb{P}(\{X_1-\widetilde{X}_0=\pm 1\})} \qquad (2.2.11)\\
&= \mathbb{P}\left(\bigcup_{n\geqslant 2}\{X_n=0\}\ \middle|\ \{X_1=k\pm 1\}\right)\\
&= \mathbb{P}\left(\bigcup_{n\geqslant 1}\{X_n=0\}\ \middle|\ \{X_0=k\pm 1\}\right)\\
&= \mathbb{P}(R_A \mid X_0=k\pm 1).
\end{aligned}
$$

Note that when switching from (2.2.9) to (2.2.10) and then to (2.2.11) using the relation $\widetilde{X}_0 := X_1 - Z$ we regard the process increment starting from X_1 as run backward in time. In case $k=1$ we easily find that

$$
\mathbb{P}(R_A \mid X_1 = 0 \text{ and } X_0 = 1) = 1 = \mathbb{P}(R_A \mid X_0 = 0),
$$

since $\{X_1=0\} \subset R_A = \bigcup_{n\in\mathbb{N}}\{X_n=0\}$.

□

In the remaining of this section we will prove that in the non-symmetric case $p \neq q$ the solution of (2.2.6) is given* by

$$f_S(k) = \mathbb{P}(R_A \mid X_0 = k) = \frac{(q/p)^k - (q/p)^S}{1-(q/p)^S} = \frac{1-(p/q)^{S-k}}{1-(p/q)^S}, \quad (2.2.12)$$

$k = 0, 1, \ldots, S$, and that in the symmetric case $p = q = 1/2$ the solution of (2.2.6) is given by

$$f_S(k) = \mathbb{P}(R_A \mid X_0 = k) = \frac{S-k}{S} = 1 - \frac{k}{S}, \quad (2.2.13)$$

$k = 0, 1, \ldots, S$, cf. also Exercise 2.3 for a different derivation.

Remark that (2.2.12) and (2.2.13) do satisfy both boundary conditions (2.2.7) and (2.2.8). When the number S of states becomes large we find that, for all $k \geqslant 0$,

$$f_\infty(k) := \lim_{S\to\infty} \mathbb{P}(R_A \mid X_0 = k) = \begin{cases} 1 & \text{if } q \geqslant p, \\ \left(\frac{q}{p}\right)^k & \text{if } p > q, \end{cases} \quad (2.2.14)$$

which represents the probability of hitting the origin starting from state $\textcircled{k}$, cf. (2.3.14) and (3.4.16) below, and also Exercise 3.2-(c) for a different derivation of this statement in the framework of random walks.

Exercise: Check that (2.2.12) agrees with (2.2.4) and (2.2.5) when $S = 2$ and $S = 3$.

In the graph of Figure 2.2 the ruin probability (2.2.12) is plotted as a function of k for $p = 0.45$ and $q = 0.55$.

* The techniques used to solve (2.2.6) can be found in MH1301 Discrete Mathematics.

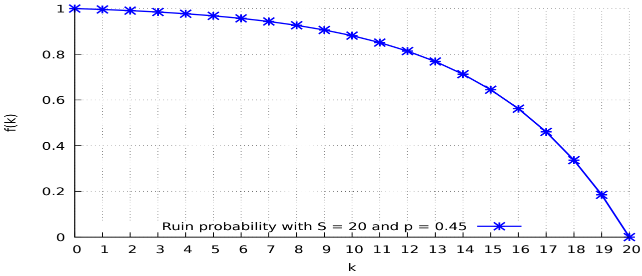


Fig. 2.2: Ruin probability $f_{20}(k)$ function of $X_0 = k \in [0, 20]$ for $S = 20$ and $p = 0.45$.

We now turn to the solution of (2.2.6), for which we develop two different approaches (called here the "standard solution" and the "direct solution") that both recover (2.2.12) and (2.2.13).

Standard solution method

We decide to look for a solution of (2.2.6) of the form *

$$k \longmapsto f_S(k) = Ca^k, \tag{2.2.15}$$

where C and a are constants which will be determined from the boundary conditions and from the equation (2.2.6), respectively.

Substituting (2.2.15) into (2.2.6) when C is non-zero yields the *characteristic equation*

$$pa^2 - a + q = p(a-1)(a - q/p) = 0, \tag{2.2.16}$$

of degree 2 in the unknown a, and this equation admits in general two solutions a_1 and a_2 given by

$$\{a_1, a_2\} = \left(\frac{1 + \sqrt{1-4pq}}{2p}, \frac{1 - \sqrt{1-4pq}}{2p} \right) = \left(1, \frac{q}{p}\right) = (1, r),$$

for all $p \in (0, 1]$, with

$$a_1 = 1 \quad \text{and} \quad a_2 = r = \frac{q}{p}.$$

Note that we have $a_1 = a_2 = 1$ in case $p = q = 1/2$.

* Where did we get this idea? From intuition, experience, or empirically by multiple trials and errors.

Non-symmetric case: $p \neq q$ - Proof of (2.2.12)

In this case we have $p \neq q$, *i.e.*[*] $r \neq 1$, and

$$f_S(k) = C_1 a_1^k = C_1 \qquad \text{and} \qquad f_S(k) = C_2 r^k$$

are both solutions of (2.2.6). Since (2.2.6) is linear, the sum of two solutions remains a solution, hence the general solution of (2.2.6) is given by

$$f_S(k) = C_1 a_1^k + C_2 a_2^k = C_1 + C_2 r^k, \qquad k = 0, 1, \ldots, S, \tag{2.2.17}$$

where $r = q/p$ and C_1, C_2 are two constants to be determined from the boundary conditions, see also the command

RSolve[f[k]=pf[k+1]+(1-p)f[k-1],f[k],k].

From (2.2.7), (2.2.8) and (2.2.17) we have

$$\begin{cases} f_S(0) = 1 = C_1 + C_2, \\ \\ f_S(S) = 0 = C_1 + C_2 r^S, \end{cases} \tag{2.2.18}$$

and solving the system (2.2.18) of two equations we find

$$C_1 = -\frac{r^S}{1 - r^S} \quad \text{and} \quad C_2 = \frac{1}{1 - r^S},$$

which yields (2.2.12) as by (2.2.17) we have

$$f_S(k) = C_1 + C_2 r^k = \frac{r^k - r^S}{1 - r^S} = \frac{(q/p)^k - (q/p)^S}{1 - (q/p)^S}, \quad k = 0, 1, \ldots, S.$$

Symmetric case: $p = q = 1/2$ - Proof of (2.2.13)

In this case, Equation (2.2.6) rewrites as

$$f(k) = \frac{1}{2} f(k+1) + \frac{1}{2} f(k-1), \qquad k = 1, 2, \ldots, S-1, \tag{2.2.19}$$

and we have $r = 1$ (fair game) and (2.2.16) reads

$$a^2 - 2a + 1 = (a-1)^2 = 0,$$

which has the unique solution $a = 1$, since the constant function $f(k) = C$ is solution of (2.2.6).

[*] From the Latin "id est" meaning "that is".

However this is not enough and we need to combine $f(k) = C_1$ with a second solution. Noting that $g(k) = C_2 k$ is also solution of (2.2.6), the general solution is found to have the form

$$f_S(k) = f(k) + g(k) = C_1 + C_2 k, \tag{2.2.20}$$

see also the command

RSolve[f[k]=(1/2)f[k+1]+(1/2)f[k-1],f[k],k].

From (2.2.7), (2.2.8) and (2.2.20) we have

$$\begin{cases} f_S(0) = 1 = C_1, \\ \\ f_S(S) = 0 = C_1 + C_2 S, \end{cases} \tag{2.2.21}$$

and solving the system (2.2.21) of two equations we find

$$C_1 = 1 \quad \text{and} \quad C_2 = -1/S,$$

which yields the quite intuitive solution

$$f_S(k) = \mathbb{P}(R_A \mid X_0 = k) = \frac{S-k}{S} = 1 - \frac{k}{S}, \quad k = 0, 1, \ldots, S, \tag{2.2.22}$$

see also the command

RSolve[f[k] ==(1/2)*f[k+1]+(1/2)*f[k-1], f[0]==1,f[10]==0,f[k], k].

Direct solution method

Noting that $p + q = 1$ and due to its special form, we can rewrite (2.2.6) as

$$(p+q)f_S(k) = pf_S(k+1) + qf_S(k-1),$$

$k = 1, 2, \ldots, S-1$, *i.e.* as the *difference equation*

$$p(f_S(k+1) - f_S(k)) - q(f_S(k) - f_S(k-1)) = 0, \tag{2.2.23}$$

$k = 1, 2, \ldots, S-1$, which rewrites as

$$f_S(k+1) - f_S(k) = \frac{q}{p}(f_S(k) - f_S(k-1)), \qquad k = 1, 2, \ldots, S-1,$$

hence for $k = 1$ we have

$$f_S(2) - f_S(1) = \frac{q}{p}(f_S(1) - f_S(0)),$$

and for $k = 2$ we find

$$f_S(3) - f_S(2) = \frac{q}{p}(f_S(2) - f_S(1)) = \left(\frac{q}{p}\right)^2 (f_S(1) - f_S(0)).$$

Following by induction on $k \geqslant 2$, we can show that

$$f_S(k+1) - f_S(k) = \left(\frac{q}{p}\right)^k (f_S(1) - f_S(0)), \tag{2.2.24}$$

$k = 0, 1, \ldots, S-1$. Next, by the telescoping sum

$$f_S(n) = f_S(0) + \sum_{k=0}^{n-1} (f_S(k+1) - f_S(k)),$$

Relation (2.2.24) implies

$$f_S(n) = f_S(0) + (f_S(1) - f_S(0)) \sum_{k=0}^{n-1} \left(\frac{q}{p}\right)^k, \tag{2.2.25}$$

$n = 1, 2, \ldots, S-1$. The remaining question is how to find $f_S(1) - f_S(0)$, knowing that $f_S(0) = 1$ by (2.2.7).

Non-symmetric case: $p \neq q$

In this case we have $r = q/p \neq 1$ and we get

$$\begin{aligned} f_S(n) &= f_S(0) + (f_S(1) - f_S(0)) \sum_{k=0}^{n-1} r^k \\ &= f_S(0) + \frac{1-r^n}{1-r}(f_S(1) - f_S(0)), \end{aligned} \tag{2.2.26}$$

$n = 1, 2, \ldots, S-1$, where we used (A.2).

Conditions (2.2.7) and (2.2.8) show that

$$0 = f_S(S) = 1 + \frac{1-r^S}{1-r}(f_S(1) - f_S(0)),$$

hence

$$f_S(1) - f_S(0) = -\frac{1-r}{1-r^S},$$

and combining this relation with (2.2.26) yields

$$f_S(n) = f_S(0) - \frac{1-r^n}{1-r^S} = 1 - \frac{1-r^n}{1-r^S} = \frac{r^n - r^S}{1-r^S},$$

$n = 0, 1, \ldots, S$, which recovers (2.2.12).

Symmetric case: $p = q = 1/2$

In this case we have $r = 1$ and in order to solve (2.2.19) we note that (2.2.23) simply becomes

$$f_S(k+1) - f_S(k) = f_S(1) - f_S(0), \qquad k = 0, 1, \ldots, S-1,$$

and (2.2.25) reads

$$f_S(n) = f_S(0) + n(f_S(1) - f_S(0)), \qquad n = 1, 2, \ldots, S-1,$$

which has the form (2.2.20). Then the conditions $f_S(0) = 1$ and $f_S(S) = 0$, cf. (2.2.7) and (2.2.8), yield

$$0 = f_S(S) = 1 + S(f_S(1) - f_S(0)), \quad \text{hence} \quad f_S(1) - f_S(0) = -\frac{1}{S},$$

and

$$f_S(n) = f_S(0) - \frac{n}{S} = 1 - \frac{n}{S} = \frac{S-n}{S},$$

$n = 0, 1, \ldots, S$, which coincides with (2.2.22).

Remark 2.3. *Note that when $p = q = 1/2$, (2.2.23) can be read as a discretization of a continuous Laplace equation as*

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(x) &\simeq \frac{\partial f}{\partial x}(x+1/2) - \frac{\partial f}{\partial x}(x-1/2)\\ &\simeq (f(x+1) - f(x) - (f(x) - f(x-1)))\\ &= 0, \qquad x \in \mathbb{R},\end{aligned} \tag{2.2.27}$$

which admits a solution of the form

$$f(x) = f(0) + xf'(0) = f(0) + x(f(1) - f(0)), \qquad x \in \mathbb{R},$$

showing the intuition behind the linear form of (2.2.20).

In order to compute the probability of ruin of Player B given that $X_0 = k$ we only need to swap k to $S-k$ and to exchange p and q in (2.2.12). In other words, when $X_0 = k$ then Player B starts with an initial amount $S-k$ and a probability q of winning each round, which by (2.2.12) yields

$$\mathbb{P}(R_B \mid X_0 = k) = \frac{(p/q)^{S-k} - (p/q)^S}{1 - (p/q)^S} = \frac{1 - (q/p)^k}{1 - (q/p)^S} \quad \text{if } p \neq q, \tag{2.2.28}$$

where

$$R_B := \text{"Player } B \text{ loses all his capital at some time"} = \bigcup_{n\in\mathbb{N}} \{X_n = S\}.$$

In the symmetric case $p = q = 1/2$ we similarly find

$$\mathbb{P}(R_B \mid X_0 = k) = \frac{k}{S}, \quad k = 0, 1, \ldots, S, \quad \text{if } p = q = \frac{1}{2}, \qquad (2.2.29)$$

see also Exercise 2.3 below.

Note that (2.2.28) and (2.2.29) satisfy the expected boundary conditions

$$\mathbb{P}(R_B \mid X_0 = 0) = 0 \quad \text{and} \quad \mathbb{P}(R_B \mid X_0 = S) = 1,$$

since X_0 represents the wealth of Player A at time 0.

By (2.2.12) and (2.2.28) we can check that*

$$\begin{aligned}\mathbb{P}(R_A \cup R_B \mid X_0 = k) &= \mathbb{P}(R_A \mid X_0 = k) + \mathbb{P}(R_B \mid X_0 = k) \\ &= \frac{(q/p)^k - (q/p)^S}{1-(q/p)^S} + \frac{1-(q/p)^k}{1-(q/p)^S} \\ &= 1, \qquad k = 0, 1, \ldots, S, \qquad (2.2.30)\end{aligned}$$

which means that eventually one of the two players has to lose the game, see also Exercise 10.1. This means in particular that, with probability one, the game cannot continue endlessly.

In other words, we have

$$\mathbb{P}\big(R_A^c \cap R_B^c \mid X_0 = k\big) = 0, \qquad k = 0, 1, \ldots, S.$$

In the particular case $S = 3$ we can indeed check that, taking

$$A_n := \bigcap_{k=1}^{n} \big\{X_{2k-1} = 1 \text{ and } X_{2k} = 2\big\}, \qquad n \geqslant 1,$$

the sequence $(A_n)_{n\geqslant 1}$ is non-increasing, hence by (1.2.6) we have

$$\begin{aligned}&\mathbb{P}\left(\bigcap_{n\geqslant 1} \big\{X_{2n-1} = 1 \text{ and } X_{2n} = 2\big\} \;\middle|\; X_0 = 2\right) \\ &\qquad = \mathbb{P}\left(\bigcap_{n\geqslant 1} \bigcap_{k=1}^{n} \big\{X_{2k-1} = 1 \text{ and } X_{2k} = 2\big\} \;\middle|\; X_0 = 2\right)\end{aligned}$$

* Exercise: check by hand computation that the equality to 1 holds as stated.

$$
\begin{aligned}
&= \mathbb{P}\left(\bigcap_{n \geqslant 1} A_n \mid X_0 = 2 \right) \\
&= \lim_{n \to \infty} \mathbb{P}(A_n \mid X_0 = 2) \\
&= \lim_{n \to \infty} \mathbb{P}\left(\bigcap_{k=1}^{n} \{X_{2k-1} = 1 \text{ and } X_{2k} = 2\} \;\Big|\; X_0 = 2 \right) \\
&= q \lim_{n \to \infty} (pq)^n = 0,
\end{aligned}
$$

since we always have $0 \leqslant pq < 1$. However, this is *a priori* not completely obvious when $S \geqslant 4$.

The expected terminal wealth of Player A starting from state $\textcircled{k}$ is given by

$$
0 \times \mathbb{P}(R_A \mid X_0 = k) + S \times \mathbb{P}(R_B \mid X_0 = k) = \begin{cases} S \dfrac{1 - (q/p)^k}{1 - (q/p)^S} & \text{if } p \neq q, \\ \\ k & \text{if } p = q = \dfrac{1}{2}. \end{cases}
$$

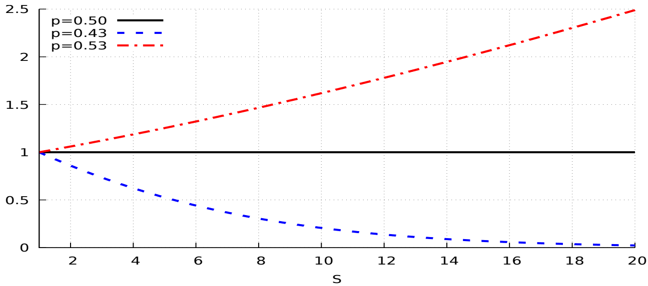


Fig. 2.3: Expected terminal wealth of Player A starting from state $k = 1$.

When the number S of states becomes large, (2.2.28) also shows that for all $k \geqslant 0$ we have

$$
\lim_{S \to \infty} \mathbb{P}(R_B \mid X_0 = k) = \begin{cases} 0 & \text{if } p \leqslant q, \\ \\ 1 - \left(\dfrac{q}{p} \right)^k & \text{if } p > q, \end{cases}
$$

which represents the complement of the probability (2.2.14) of hitting the origin starting from state (k), and is the probability that the process $(X_n)_{n\geqslant 0}$ "escapes to infinity".

In Figure 2.4 below the ruin probability (2.2.12) is plotted as a function of p for $S = 20$ and $k = 10$.

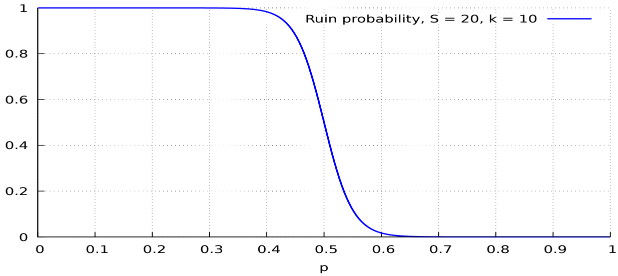


Fig. 2.4: Ruin probability as a function of $p \in [0, 1]$ for $S = 20$ and $k = 10$.

Gambling machines in casinos are computer controlled and most countries permit by law a certain degree of "unfairness" (see the notions of "payout percentage" or "return to player") by taking $p < 1/2$ in order to allow the house to make an income.* Interestingly, we can note that taking *e.g.* $p = 0.45 < 1/2$ gives a ruin probability

$$\mathbb{P}(R_A \mid X_0 = 10) = 0.8815,$$

almost equal to 90%, which means that the slightly unfair probability $p = 0.45$ at the level of each round translates into a probability of only $0.1185 \simeq \%12$ of finally winning the game, *i.e.* a division by 4 from 0.45, although the average proportion of winning rounds is still 45%.

Consequently, a "slightly unfair" game at each round may become devastatingly unfair in the long run.† Most (but not all) gamblers are aware that gambling machines can be slightly unfair, however most people would intuitively believe that a small degree of unfairness on each round should only translate into a reasonably low degree of unfairness in the long run.

* In this game, the payout is \$2 and the payout percentage is $2p$.

† "You can't lose with house odds and time. You can't win with gambler odds and time." Peter Kaufman via Dennis Hong.

2.3 Mean Game Duration

Let now

$$T_{0,S} = \inf\{n \geqslant 0 \ : \ X_n = 0 \text{ or } X_n = S\}$$

denote the time* until any of the states ⓪ or Ⓢ is reached by $(X_n)_{n\geqslant 0}$, with $T_{0,S} = +\infty$ in case neither states are ever reached, *i.e.* when there exists no integer $n \geqslant 0$ such that $X_n = 0$ or $X_n = S$, see Figure 2.5.

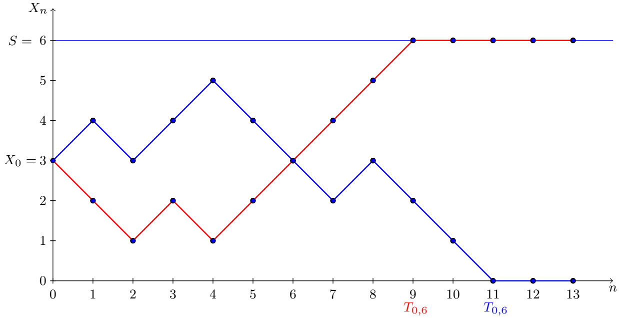


Fig. 2.5: Sample paths of a gambling process $(X_n)_{n\geqslant 0}$.

Note that by (2.2.30) we have

$$\mathbb{P}(T_{0,S} < \infty \mid X_0 = k) = \mathbb{P}(R_A \cup R_B \mid X_0 = k) = 1, \quad k = 0, 1, \ldots, S.$$

and therefore

$$\mathbb{P}(T_{0,S} = \infty \mid X_0 = k) = 0, \qquad k = 0, 1, \ldots, S.$$

We are now interested in computing the expected duration

$$h_S(k) := \mathbb{E}[T_{0,S} \mid X_0 = k]$$

of the game given that Player A starts with a wealth equal to $X_0 = k \in \{0, 1, \ldots, S\}$. Clearly, we have the boundary conditions

$$\begin{cases} h_S(0) = \mathbb{E}[T_{0,S} \mid X_0 = 0] = 0, & (2.3.1a) \\ h_S(S) = \mathbb{E}[T_{0,S} \mid X_0 = S] = 0. & (2.3.1b) \end{cases}$$

* The notation "inf" stands for "infimum", meaning the smallest $n \geqslant 0$ such that $X_n = 0$ or $X_n = S$, if such an n exists.

We start by considering the particular cases $S = 2$ and $S = 3$.

i) $\underline{S = 2}$.

We have

$$T_{0,2} = \begin{cases} 0 \ \text{ if } X_0 = 0, \\ \\ 1 \ \text{ if } X_0 = 1, \\ \\ 0 \ \text{ if } X_0 = 2, \end{cases}$$

thus $T_{0,2}$ is *deterministic* given the value of X_0 and we simply have $h_2(1) = T_{0,2} = 1$ when $X_0 = 1$.

ii) $\underline{S = 3}$.

In this case the probability distribution of $T_{0,3}$ given $X_0 \in \{0, 1, 2, 3\}$ can be determined explicitly and we find, when $X_0 = 1$,

$$\begin{cases} \mathbb{P}(T_{0,3} = 2k \mid X_0 = 1) = p^2(pq)^{k-1}, \qquad k \geqslant 1, \\ \\ \mathbb{P}(T_{0,3} = 2k+1 \mid X_0 = 1) = q(pq)^k, \qquad k \geqslant 0, \end{cases}$$

since in an even number $2k$ of steps we can only exit through state ③ after starting from ①, while in an odd number $2k + 1$ of steps we can only exit through state ⓪. By exchanging p with q in the above formulas we get, when $X_0 = 2$,

$$\begin{cases} \mathbb{P}(T_{0,3} = 2k \mid X_0 = 2) = q^2(pq)^{k-1}, \qquad k \geqslant 1, \\ \\ \mathbb{P}(T_{0,3} = 2k+1 \mid X_0 = 2) = p(pq)^k, \qquad k \geqslant 0, \end{cases}$$

whereas $T_{0,3} = 0$ whenever $X_0 = 0$ or $X_0 = 3$.

As a consequence, we can directly compute

$$\begin{aligned} h_3(2) = \mathbb{E}[T_{0,3} \mid X_0 = 2] &= 2\sum_{k\geqslant 1} k\mathbb{P}(T_{0,3} = 2k \mid X_0 = 2) \\ &\quad + \sum_{k\geqslant 0}(2k+1)\mathbb{P}(T_{0,3} = 2k+1 \mid X_0 = 2) \\ &= 2q^2\sum_{k\geqslant 1} k(pq)^{k-1} + p\sum_{k\geqslant 0}(2k+1)(pq)^k \\ &= \frac{2q^2}{(1-pq)^2} + \frac{2p^2q}{(1-pq)^2} + \frac{p}{1-pq} \\ &= \frac{2q^2 + p + qp^2}{(1-pq)^2} \\ &= \frac{2(1-p)q + 1 - q + q(1-q)p}{(1-pq)^2} \end{aligned} \tag{2.3.2}$$

$$= \frac{1+q-pq-q^2p}{(1-pq)^2}$$
$$= \frac{1+q}{1-pq}, \tag{2.3.3}$$

where we applied (A.4), and by exchanging p and q we get

$$h_3(1) = \mathbb{E}[T_{0,3} \mid X_0 = 1] = \frac{2p^2+q+pq^2}{(1-pq)^2} = \frac{1+p}{1-pq}. \tag{2.3.4}$$

Again, things can become quite complicated for $S \geqslant 4$, and increasingly difficult when S becomes larger.

In the general case $S \geqslant 4$ we will only compute the conditional expectation of $T_{0,S}$ and not its probability distribution. For this we rely again on *first step analysis*, as stated in the following lemma.

Lemma 2.4. *For all* $k = 1, 2, \ldots, S-1$ *we have*

$$\mathbb{E}[T_{0,S} \mid X_0 = k] = 1 + p\mathbb{E}[T_{0,S} \mid X_0 = k+1] + q\mathbb{E}[T_{0,S} \mid X_0 = k-1].$$

Proof. We condition on the first transition from X_0 to X_1. Using the equality $\mathbb{1}_A = \mathbb{1}_{A\cap B} + \mathbb{1}_{A\cap B^c}$ under the form

$$\mathbb{1}_{\{X_0=k\}} = \mathbb{1}_{\{X_1=k+1, X_0=k\}} + \mathbb{1}_{\{X_1=k-1, X_0=k\}},$$

cf. (1.4.2), and conditional expectations we show, by first step analysis, that for all $k = 1, 2, \ldots, S-1$, applying Lemma 1.10 and (1.6.6) successively to $A = \{X_0 = k\}$, $A = \{X_0 = k-1\}$ and $A = \{X_0 = k+1\}$, we have

$$\begin{aligned}
\mathbb{E}[T_{0,S} \mid X_0 = k] &= \frac{1}{\mathbb{P}(X_0 = k)} \mathbb{E}\left[T_{0,S}\mathbb{1}_{\{X_0=k\}}\right] \\
&= \frac{1}{\mathbb{P}(X_0 = k)} \left(\mathbb{E}\left[T_{0,S}\mathbb{1}_{\{X_1=k+1,X_0=k\}}\right] + \mathbb{E}\left[T_{0,S}\mathbb{1}_{\{X_1=k-1,X_0=k\}}\right]\right) \\
&= \frac{\mathbb{P}(X_1 = k+1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)} \mathbb{E}[T_{0,S} \mid X_1 = k+1, X_0 = k] \\
&\quad + \frac{\mathbb{P}(X_1 = k-1 \text{ and } X_0 = k)}{\mathbb{P}(X_0 = k)} \mathbb{E}[T_{0,S} \mid X_1 = k-1, X_0 = k] \\
&= \mathbb{P}(X_1 = k+1 \mid X_0 = k)\mathbb{E}[T_{0,S} \mid X_1 = k+1, X_0 = k] \\
&\quad + \mathbb{P}(X_1 = k-1 \mid X_0 = k)\mathbb{E}[T_{0,S} \mid X_1 = k-1, X_0 = k] \\
&= p\mathbb{E}[T_{0,S} \mid X_1 = k+1, X_0 = k] + q\mathbb{E}[T_{0,S} \mid X_1 = k-1, X_0 = k] \qquad (2.3.5) \\
&= p\mathbb{E}[T_{0,S} + 1 \mid X_0 = k+1, X_{-1} = k] + q\mathbb{E}[T_{0,S} + 1 \mid X_0 = k-1, X_{-1} = k] \qquad (2.3.6) \\
&= p\mathbb{E}[T_{0,S} + 1 \mid X_0 = k+1] + q\mathbb{E}[T_{0,S} + 1 \mid X_0 = k-1]
\end{aligned}$$

$$= p(1 + \mathbb{E}[T_{0,S} \mid X_0 = k+1]) + q(1 + \mathbb{E}[T_{0,S} \mid X_0 = k-1])$$
$$= p + q + p\mathbb{E}[T_{0,S} \mid X_0 = k+1] + q\mathbb{E}[T_{0,S} \mid X_0 = k-1]$$
$$= 1 + p\mathbb{E}[T_{0,S} \mid X_0 = k+1] + q\mathbb{E}[T_{0,S} \mid X_0 = k-1].$$

From (2.3.5) to (2.3.6) we relabelled X_1 as X_0, which amounts to changing $T_{0,S} - 1$ into $T_{0,S}$, or equivalently changing $T_{0,S}$ into $T_{0,S} + 1$. □

In the case $S = 3$, Lemma 2.4 shows that

$$\begin{cases} h_3(0) = \mathbb{E}[T_{0,S} \mid X_0 = 0] = 0, \\ h_3(1) = 1 + ph_3(2) + qh_3(0) = 1 + ph_3(2) = 1 + p(1 + qh_3(1)) = 1 + p + pqh_3(1), \\ h_3(2) = 1 + ph_3(3) + qh_3(1) = 1 + qh_3(1) = 1 + q(1 + ph_3(2)) = 1 + q + pqh_3(2), \\ h_3(3) = \mathbb{E}[T_{0,S} \mid X_0 = 3] = 0, \end{cases}$$

whose solution can be checked to be given by (2.3.3)-(2.3.4).

More generally, defining the function $h_S : \{0, 1, \ldots, S\} \longrightarrow \mathbb{R}_+$ by

$$h_S(k) := \mathbb{E}[T_{0,S} \mid X_0 = k], \qquad k = 0, 1, \ldots, S,$$

Lemma 2.4 shows that

$$h_S(k) = p(1 + h_S(k+1)) + q(1 + h_S(k-1)) = 1 + ph_S(k+1) + qh_S(k-1),$$

$k = 1, 2, \ldots, S-1$, *i.e.* we have to solve the equation

$$\begin{cases} h_S(k) = 1 + ph_S(k+1) + qh_S(k-1), \quad 1 \leqslant k < S, & \text{(2.3.7a)} \\ h_S(0) = h_S(S) = 0, & \text{(2.3.7b)} \end{cases}$$

for the function $h_S(k)$. Using the fact that $p + q = 1$, we can rewrite (2.3.7a) as

$$(p+q)h_S(k) = 1 + ph_S(k+1) + qh_S(k-1), \qquad k = 1, 2, \ldots, S-1,$$

or as the *difference equation*

$$p(h_S(k+1) - h_S(k)) - q(h_S(k) - h_S(k-1)) = -1, \qquad k = 1, 2, \ldots, S-1, \tag{2.3.8}$$

under the *boundary conditions** (2.3.7b).

* The techniques used to solve (2.3.7a) can be found in MH1301 Discrete Mathematics.

The equation

$$p(f(k+1)-f(k))-q(f(k)-f(k-1))=0, \qquad k=1,2,\ldots,S-1, \tag{2.3.9}$$

cf. (2.2.23), is called the *homogeneous equation* associated to (2.3.8).

We will use the following fact:

The general solution to (2.3.8) can be written as the sum of a *homogeneous solution* of (2.3.9) and a *particular solution* of (2.3.8).

Non-symmetric case: $p \neq q$

By (2.2.17) we know that the *homogeneous solution* of (2.3.9) is of the form $C_1 + C_2 r^k$. Next, searching for a *particular solution* of (2.3.8) of the form $k \longmapsto Ck$ shows that C has to be equal to $C = 1/(q-p)$. Therefore, when $r = q/p \neq 1$, the general solution of (2.3.8) has the form

$$h_S(k) = C_1 + C_2 r^k + \frac{1}{q-p}k, \tag{2.3.10}$$

see also the command

RSolve[f[k]=1+pf[k+1]+(1-p)f[k-1],f[k],k].

From the boundary conditions (2.3.1a) and (2.3.1b) and from (2.3.10) we have

$$\begin{cases} h_S(0) = 0 = C_1 + C_2, \\ \\ h_S(S) = 0 = C_1 + C_2 r^S + \dfrac{1}{q-p}S, \end{cases} \tag{2.3.11}$$

and solving the system (2.3.11) of two equations we find

$$C_1 = -\frac{S}{(q-p)(1-r^S)} \quad \text{and} \quad C_2 = \frac{S}{(q-p)(1-r^S)},$$

hence from (2.3.10) we get

$$\begin{aligned} h_S(k) = \mathbb{E}[T_{0,S} \mid X_0 = k] &= \frac{1}{q-p}\left(k - S\frac{1-(q/p)^k}{1-(q/p)^S}\right) \qquad (2.3.12) \\ &= \frac{1}{q-p}\left(k - S\mathbb{P}(R_B \mid X_0 = k)\right), \quad k = 0,1,2,\ldots,S, \end{aligned}$$

which does satisfy the boundary conditions (2.3.1a) and (2.3.1b). Note that changing k to $S-k$ and p to q does not modify (2.3.12), as it also represents the mean game duration for Player B. In particular, we note that

$$\mathbb{E}[T_{0,S} \mid X_0 = k] < +\infty, \qquad k = 0, 1, 2, \ldots, S.$$

When $p = 1$, *i.e.* $r = 0$, we can check easily that *e.g.* $h_S(k) = S - k$, $k = 0, 1, 2, \ldots, S$. On the other hand, when the number S of states becomes large, we find that for all $k \geqslant 1$,

$$h_\infty(k) := \lim_{S\to\infty} h_S(k) = \lim_{S\to\infty} \mathbb{E}[T_{0,S} \mid X_0 = k] = \begin{cases} \infty & \text{if } q \leqslant p, \\ \\ \dfrac{k}{q-p} & \text{if } q > p, \end{cases} \tag{2.3.13}$$

with $h_\infty(0) = 0$, cf. also the symmetric case treated in (2.3.19) below when $p = q = 1/2$. In particular, for $k \geqslant 1$ we have, see (2.2.14),

$$\begin{cases} \mathbb{P}(T_0 < \infty \mid X_0 = k) = \left(\dfrac{q}{p}\right)^k < 1 \ \text{ and } \ \mathbb{E}[T_0 \mid X_0 = k] = \infty, \ \text{ for } \ p > q, \\ \\ \mathbb{P}(T_0 < \infty \mid X_0 = k) = 1 \ \text{ and } \ \mathbb{E}[T_0 \mid X_0 = k] = \infty, \ \text{ for } \ p = q = \dfrac{1}{2}, \\ \\ \mathbb{P}(T_0 < \infty \mid X_0 = k) = 1 \ \text{ and } \ \mathbb{E}[T_0 \mid X_0 = k] < \infty, \ \text{ for } \ p < q. \end{cases} \tag{2.3.14}$$

We note in particular that the mean game duration $\mathbb{E}[T_0 \mid X_0 = k]$ is infinite in the fair game case $p = q = 1/2$. When $r = q/p = 1$, this yields an example of a random variable T_0 which is (almost surely*) finite, while its expectation is infinite.† This situation is similar to that of the St. Petersburg paradox as in (1.6.5). Similarly, one can find sequences $(X_n)_{n\geqslant 0}$ of random variables such that $X_n \to 0$ with probability one, while $\mathbb{E}[X_n] \to \infty$ as n tends to infinity.

When $S = 2$ it is easy to show that (2.3.12) yields $h_2(1) = 1$. When $S = 3$, (2.3.12) shows that, using the relation $p + q = 1$,‡

$$\mathbb{E}[T_{0,3} \mid X_0 = 1] = \frac{1}{q-p}\left(1 - 3\frac{1 - q/p}{1 - (q/p)^3}\right) = \frac{1+p}{1-pq}, \tag{2.3.15}$$

and

$$\mathbb{E}[T_{0,3} \mid X_0 = 2] = \frac{1}{q-p}\left(2 - 3\frac{1 - (q/p)^2}{1 - (q/p)^3}\right) = \frac{1+q}{1-pq}, \tag{2.3.16}$$

* "Almost surely" means "with probability 1".

† Recall that an infinite set of finite data values may have an infinite average. Similarly, there exist finite surfaces that enclose an infinite volume.

‡ This point is left as exercise.

however it takes more time to show that (2.3.15) and (2.3.16) agree respectively with (2.3.2) and (2.3.4), see for example *here*. In Figure 2.6 the mean game duration (2.3.12) is plotted as a function of k for $p = 0.45$.

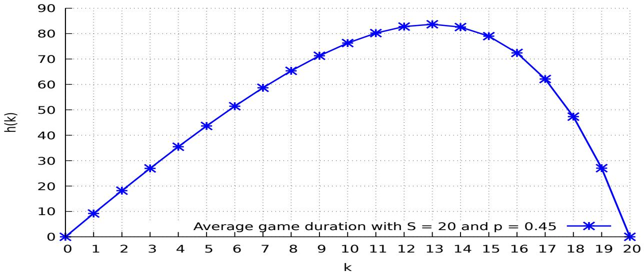


Fig. 2.6: Mean game duration $h_{20}(k)$ as a function of $X_0 = k \in [0, 20]$ for $p = 0.45$.

Symmetric case: $p = q = 1/2$

In this case (fair game) the *homogeneous solution* of (2.3.9) is $C_1 + C_2 k$, given by (2.2.20).

Since $r = 1$ we see that $k \longmapsto Ck$ can no longer be a particular solution of (2.3.8). However we can search for a particular solution of the form $k \longmapsto Ck^2$, in which case we find that C has to be equal to $C = -1$.

Therefore when $r = q/p = 1$ the general solution of (2.3.8) has the form

$$h_S(k) = C_1 + C_2 k - k^2, \qquad k = 0, 1, 2, \ldots, S, \tag{2.3.17}$$

see also the command

RSolve[f[k]=1+(1/2)f[k+1]+(1/2)f[k-1],f[k],k].

From the boundary conditions (2.3.1a) and (2.3.1b) and from (2.3.17) we have

$$\begin{cases} h_S(0) = 0 = C_1, \\ \\ h_S(S) = 0 = C_1 + C_2 S - S^2, \end{cases} \tag{2.3.18}$$

and solving the above system (2.3.18) of two equations yields

$$C_1 = 0 \quad \text{and} \quad C_2 = S,$$

hence from (2.3.17) we get

$$h_S(k) = \mathbb{E}[T_{0,S} \mid X_0 = k] = k(S-k), \qquad k = 0, 1, 2, \ldots, S, \tag{2.3.19}$$

which does satisfy the boundary conditions (2.3.1a) and (2.3.1b) and coincides with (2.3.13) when S goes to infinity, see also the command

RSolve[f[k] ==1+(1/2)*f[k+1]+(1/2)*f[k-1], f[0]==1,f[10]==0,f[k], k].

In particular, we have

$$\mathbb{E}[T_{0,S} \mid X_0 = k] < +\infty, \qquad k = 0, 1, 2, \ldots, S.$$

We also note that for all values of $p \in [0,1]$ the expectation $\mathbb{E}[T_{0,S} \mid X_0 = k]$ has a *finite* value, which recovers the fact that the game duration $T_{0,S}$ is finite with probability one for all $k = 0, 1, \ldots, S$, *i.e.* $\mathbb{P}(T_{0,S} = \infty \mid X_0 = k) = 0$ for all $k = 0, 1, \ldots, S$, see Exercise 10.1.

Remark 2.5. *When $r = 1$, by the same argument as in (2.2.27) we find that (2.3.8) is a discretization of the continuous Poisson equation*

$$\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x) = -1, \qquad x \in \mathbb{R},$$

which has for solution

$$f(x) = f(0) + x f'(0) - x^2, \qquad x \in \mathbb{R}.$$

Relation (2.3.19) can also be recovered from (2.3.12) by letting p go to $1/2$. In the next Figure 2.7 the expected game duration (2.3.12) is plotted as a function of p for $S = 20$ and $k = 10$.

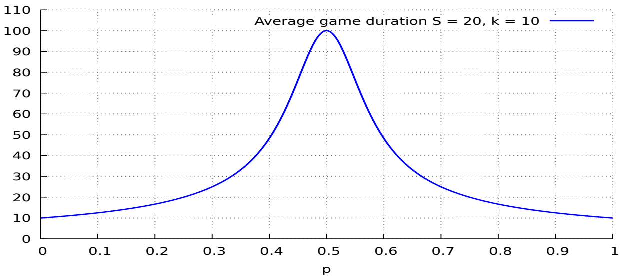


Fig. 2.7: Mean game duration as a function of $p \in [0,1]$ for $S = 20$ and $k = 10$.

As expected, the duration will be maximal in a fair game for $p = q = 1/2$. On the other hand, it always takes exactly $10 = S - k = k$ steps to end the game in case $p = 0$ or $p = 1$, in which case there is no randomness. When $p = 0.45$ the expected duration of the game becomes 76.3, which represents only a drop of 24% from the "fair" value 100, as opposed to the 73% drop noticed above in terms of winning probabilities. Thus, a game with $p = 0.45$ is only slightly shorter than a fair game, whereas the probability of winning the game drops down to 0.12.

Remark 2.6. *In this Chapter 2 we have noticed an interesting connection between analysis and probability. That is, a probabilistic quantity such as* $k \longmapsto \mathbb{P}(R_A \mid X_0 = k)$ *or* $k \longmapsto \mathbb{E}[T_{0,S} \mid X_0 = k]$ *can be shown to satisfy a difference equation which is solved by analytic methods. This fact actually extends beyond the present simple framework, and in continuous time it yields other connections between probability and partial differential equations.*

In the next chapter we will consider a family of simple random walks which can be seen as "unrestricted" gambling processes.

Exercises

Exercise 2.1 We consider a gambling problem with the possibility of a draw,* *i.e.* at time n the gain X_n of Player A can increase by one unit with probability $r \in (0, 1/2]$, decrease by one unit with probability r, or remain stable with probability $1 - 2r$. We let

$$f(k) := \mathbb{P}(R_A \mid X_0 = k)$$

denote the probability of ruin of Player A, and let

$$h(k) := \mathbb{E}[T_{0,S} \mid X_0 = k]$$

denote the expectation of the game duration $T_{0,S}$ starting from $X_0 = k$, $k = 0, 1, \ldots, S$.

a) Using first step analysis, write down the difference equation satisfied by $f(k)$ and its boundary conditions, $k = 0, 1, \ldots, S$. We refer to this equation as the *homogeneous equation*.
b) Solve the *homogeneous equation* of Question (a) by your preferred method. Is this solution compatible with your intuition of the problem? Why?
c) Using first step analysis, write down the difference equation satisfied by $h(k)$ and its boundary conditions, $k = 0, 1, \ldots, S$.
d) Find a particular solution of the equation of Question (c).

* Also called "lazy random walk"

e) Solve the equation of Question (c).

   Hint: recall that the general solution of the equation is the sum of a particular solution and a solution of the *homogeneous equation*.

f) How does the mean duration $h(k)$ behave as r goes to zero? Is this solution compatible with your intuition of the problem? Why?

Exercise 2.2 Recall that for any standard gambling process $(Z_k)_{k\geqslant 0}$ on a state space $\{a, a+1, \ldots, b-1, b\}$ with absorption at states $\textcircled{a}$ and $\textcircled{b}$ and probabilities $p \neq q$ of moving by ± 1, the probability of hitting state $\textcircled{a}$ before hitting state $\textcircled{b}$ after starting from state $Z_0 = k \in \{a, a+1, \ldots, b-1, b\}$ is given by

$$\frac{1-(p/q)^{b-k}}{1-(p/q)^{b-a}}. \tag{2.3.20}$$

In questions (a)-(b)-(c) below we consider a gambling process $(X_k)_{k\geqslant 0}$ on the state space $\{0, 1, \ldots, S\}$ with absorption at $\textcircled{0}$ and $\textcircled{S}$ and probabilities $p \neq q$ of moving by ± 1.

a) Using Relation (2.3.20), give the probability of coming back in finite time to a given state $m \in \{1, 2, \ldots, S-1\}$ after starting from $X_0 = k \in \{m+1, \ldots, S\}$.
b) Using Relation (2.3.20), give the probability of coming back in finite time to the given state $m \in \{1, 2, \ldots, S-1\}$ after starting from $X_0 = k \in \{0, 1, \ldots, m-1\}$.
c) Using first step analysis, give the probability of coming back to state $\textcircled{m}$ in finite time after starting from $X_0 = m$.
d) Using first step analysis, compute the mean time to either come back to m of reach any of the two boundaries $\{0, S\}$, whichever comes first?
e) Repeat the above questions (c)-(d) with equal probabilities $p = q = 1/2$, in which case the probability of hitting state $\textcircled{a}$ before hitting state $\textcircled{b}$ after starting from state $Z_0 = k$ is given by

$$\frac{b-k}{b-a}, \qquad k = a, a+1, \ldots, b-1, b. \tag{2.3.21}$$

Exercise 2.3 Consider the gambling process $(X_n)_{n\geqslant 0}$ on the state space $\mathbb{S} = \{0, 1, \ldots, S\}$, with probability p, resp. q, of moving up, resp. down, at each time step. For $k = 0, 1, \ldots, S$, let τ_k denote the first hitting time

$$\tau_k := \inf\{n \geqslant 0 \ : X_n = k\}.$$

of state $\textcircled{k}$ by the process $(X_n)_{n\geqslant 0}$, and let

$$p_k := \mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k), \qquad k = 0, 1, \ldots, S-1,$$

denote the probability of hitting state $\boxed{k+1}$ before hitting state ⓪.

a) Show that $p_k = \mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k)$ satisfies the recurrence equation

$$p_k = p + qp_{k-1}p_k, \qquad k = 1, 2, \ldots, S-1, \tag{2.3.22}$$

i.e.

$$p_k = \frac{p}{1 - qp_{k-1}}, \qquad k = 1, 2, \ldots, S-1.$$

b) Check by substitution that the solution of (2.3.22) is given by

$$p_k = \frac{1-(q/p)^k}{1-(q/p)^{k+1}}, \qquad k = 0, 1, \ldots, S-1. \tag{2.3.23}$$

c) Compute $\mathbb{P}(\tau_S < \tau_0 \mid X_0 = k)$ by a product formula and recover (2.2.12) and (2.2.28) based on the result of part (2.3.23).

d) Show that (2.2.13) and (2.2.29) can be recovered in a similar way in the symmetric case $p = q = 1/2$ by trying the solution $p_k = k/(k+1)$, $k = 0, 1, \ldots, S-1$.

Exercise 2.4 Consider a gambling process $(X_n)_{n\geqslant 0}$ on the state space $\{0, 1, 2\}$, with transition probabilities given by

$$\begin{bmatrix} \mathbb{P}(X_1 = 0 \mid X_0 = 0) & \mathbb{P}(X_1 = 1 \mid X_0 = 0) & \mathbb{P}(X_1 = 2 \mid X_0 = 0) \\ \mathbb{P}(X_1 = 0 \mid X_0 = 1) & \mathbb{P}(X_1 = 1 \mid X_0 = 1) & \mathbb{P}(X_1 = 2 \mid X_0 = 1) \\ \mathbb{P}(X_1 = 0 \mid X_0 = 2) & \mathbb{P}(X_1 = 1 \mid X_0 = 2) & \mathbb{P}(X_1 = 2 \mid X_0 = 2) \end{bmatrix}$$

$$= \begin{matrix} \\ 0 \\ 1 \\ 2 \end{matrix} \begin{matrix} \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{bmatrix} q & p & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{bmatrix} \end{matrix},$$

where $0 < p < 1$ and $q = 1 - p$. In this game, Player A is allowed to "rebound" from state ⓪ to state ① with probability p, and state ② is absorbing.

In order to be ruined, Player A has to visit state ⓪ **<u>twice</u>**. Let

$$f(k) := \mathbb{P}(R_A \mid X_0 = k), \qquad k = 0, 1, 2,$$

denote the probability of ruin of Player A starting from $k = 0, 1, 2$. Starting from ⓪ counts as one visit to ⓪.

a) Compute the boundary condition $f(0)$ using pathwise analysis.

b) Give the value of the boundary condition $f(2)$, and compute $f(1)$ by first step analysis.

Exercise 2.5

a) Recover (2.3.19) from (2.3.12) by letting p go to $1/2$, *i.e.* when $r = q/p$ goes to 1.
b) Recover (2.2.22) from (2.2.12) by letting p go to $1/2$, *i.e.* when $r = q/p$ goes to 1.

Exercise 2.6 Extend the setting of Exercise 2.1 to a non-symmetric gambling process with draw and respective probabilities $\alpha > 0$, $\beta > 0$, and $1 - \alpha - \beta > 0$ of increase, decrease, and draw. Compute the ruin probability $f(k)$ and the mean game duration $h(k)$ in this extended framework. Check that when $\alpha = \beta \in (0, 1/2)$ we recover the result of Exercise 2.1.

Problem 2.7 We consider a discrete-time process $(X_n)_{n\geqslant 0}$ that models the wealth of a gambler within $\{0, 1, \ldots, S\}$, with the transition probabilities

$$\begin{cases} \mathbb{P}(X_{n+1} = k+1 \mid X_n = k) = p, & k = 0, 1, \ldots, S-1, \\ \mathbb{P}(X_{n+1} = k-1 \mid X_n = k) = q, & k = 1, 2, \ldots, S-1, \end{cases}$$

with

$$\mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = q \quad \text{and} \quad \mathbb{P}(X_{n+1} = S \mid X_n = S) = 1,$$

for all $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, where $q = 1 - p$ and $p \in (0, 1]$. In this model the gambler is given a second chance, and may be allowed to "rebound" to state ⓵ after reaching ⓪. Let

$$W = \bigcup_{n\in\mathbb{N}} \{X_n = S\}$$

denote the event that the player eventually wins the game.

a) Let
$$g(k) := \mathbb{P}(W \mid X_0 = k)$$
denote the probability that the player eventually wins after starting from state $k \in \{0, 1, \ldots, S\}$. Using first step analysis, write down the difference equations satisfied by $g(k)$, $k = 0, 1, \ldots, S-1$, and their boundary condition(s), which may not be given in explicit form. This question is standard, however one has to pay attention to the special behavior of the process at state ⓪.
b) Obtain $\mathbb{P}(W \mid X_0 = k)$ for all $k = 0, 1, \ldots, S$ as the unique solution to the system of equations stated in Question (a).

   The answer to this question is very simple and can be obtained through intuition. However, a (mathematical) proof is required.

c) Let

$$T_S = \inf\{n \geqslant 0 \ : \ X_n = S\}$$

denote the first hitting time of S by the process $(X_n)_{n\geqslant 0}$. Let

$$h(k) := \mathbb{E}[T_S \mid X_0 = k]$$

denote the expected time until the gambler wins after starting from state $k \in \{0, 1, \ldots, S\}$. Using first step analysis, write down the difference equations satisfied by $h(k)$ for $k = 0, 1, \ldots, S-1$, and state the corresponding boundary condition(s).

Again, one has to pay attention to the special behavior of the process at state ⓪, as the equation obtained by first step analysis for $h(0)$ will take a particular form and can be viewed as a second boundary condition.

d) Compute $\mathbb{E}[T_S \mid X_0 = k]$ for all $k = 0, 1, \ldots, S$ by solving the equations of Question (c).

This question is more difficult than Question (b), and it could be skipped at first reading since its result is not used in the sequel. One can solve the associated *homogeneous equation* for $k = 1, 2, \ldots, S-1$ using the results of Section 2.3, and a particular solution can be found by observing that here we consider the time until Player A (not B) wins. As usual, the cases $p \neq q$ and $p = q = 1/2$ have to be considered separately at some point. The formula obtained for $p = 1$ should be quite intuitive and may help you check your result.

e) Let now

$$T_0 = \inf\{n \geqslant 0 \ : \ X_n = 0\}$$

denote the first hitting time of 0 by the process $(X_n)_{n\geqslant 0}$. Using the results of Section 2.2 for the ruin of Player B, write down the value of

$$p_k := \mathbb{P}(T_S < T_0 \mid X_0 = k)$$

as a function of p, S, and $k = 0, 1, \ldots, S$.

Note that according to the notation of Chapter 2, $\{T_S < T_0\}$ denotes the event "Player A wins the game".

f) Explain why the equality

$$\begin{aligned}\mathbb{P}(T_S < T_0 \mid X_1 = k+1 \text{ and } X_0 = k) &= \mathbb{P}(T_S < T_0 \mid X_1 = k+1) \\ &= \mathbb{P}(T_S < T_0 \mid X_0 = k+1). \qquad (2.3.24)\end{aligned}$$

holds for $k \in \{0, 1, \ldots, S-1\}$ (an explanation in words will be sufficient here).

g) Using Relation (2.3.24), show that the probability

$$\mathbb{P}(X_1 = k+1 \mid X_0 = k \text{ and } T_S < T_0)$$

of an upward step given that state Ⓢ is reached first, is equal to

$$\mathbb{P}(X_1 = k+1 \mid X_0 = k \text{ and } T_S < T_0) = p\frac{\mathbb{P}(T_S < T_0 \mid X_0 = k+1)}{\mathbb{P}(T_S < T_0 \mid X_0 = k)} = p\frac{p_{k+1}}{p_k}, \quad (2.3.25)$$

$k = 1, 2, \ldots, S-1$, to be computed explicitly from the result of Question (e). How does this probability compare to the value of p?

No particular difficulty here, the proof should be a straightforward application of the definition of conditional probabilities.

h) Compute the probability

$$\mathbb{P}(X_1 = k-1 \mid X_0 = k \text{ and } T_0 < T_S), \qquad k = 1, 2, \ldots, S,$$

of a downward step given that state ⓪ is reached first, using similar arguments to Question (g).

i) Let

$$h(k) = \mathbb{E}[T_S \mid X_0 = k, \; T_S < T_0], \qquad k = 1, 2, \ldots, S,$$

denote the expected time until the player wins, given that state ⓪ is never reached. Using the transition probabilities (2.3.25), state the finite difference equations satisfied by $h(k)$, $k = 1, 2, \ldots, S-1$, and their boundary condition(s).

The derivation of the equation is standard, but you have to make a careful use of conditional transition probabilities given $\{T_S < T_0\}$. There is an issue on whether and how $h(0)$ should appear in the system of equations, but this point can actually be solved.

j) Solve the equation of Question (i) when $\underline{p = 1/2}$ and compute $h(k)$ for $k = 1, 2, \ldots, S$. What can be said of $h(0)$?

There is actually a way to transform this equation using an homogeneous equation already solved in Section 2.3.

Problem 2.8 Let $S \geqslant 1$. We consider a discrete-time process $(X_n)_{n \geqslant 0}$ that models the wealth of a gambler within $\{0, 1, \ldots, S\}$, with the transition probabilities

$$\mathbb{P}(X_{n+1} = k+2 \mid X_n = k) = p, \quad \mathbb{P}(X_{n+1} = k-1 \mid X_n = k) = 2p,$$

and

$$\mathbb{P}(X_{n+1} = k \mid X_n = k) = r, \quad k \in \mathbb{Z},$$

for all $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, where $p > 0$, $r \geqslant 0$, and $3p + r = 1$. We let

$$\tau := \inf\{n \geqslant 0 \; : \; X_n \leqslant 0 \text{ or } X_n \geqslant S\}.$$

a) Consider the probability

$$g(k) := \mathbb{P}(X_\tau \geqslant S \mid X_0 = k)$$

that the game ends with Player A winning the game, starting from $X_0 = k$. Give the values of $g(0)$, $g(S)$ and $g(S+1)$.

b) Using first step analysis, write down the difference equation satisfied by $g(k)$, $k = 1, 2, \ldots, S-1$, and its boundary conditions, by taking *overshoot* into account. We refer to this equation as the *homogeneous equation*.

c) Solve the equation of Question (b) from its characteristic equation as in (2.2.16).

d) Does the answer to Question (c) depend on p? Why?

e) Consider the expected time

$$h(k) := \mathbb{E}[\tau \mid X_0 = k], \qquad k = 0, 1, \ldots, S+1,$$

spent until the end of the game. Give the values of $h(0)$, $h(S)$ and $h(S+1)$.

f) Using first step analysis, write down the difference equation satisfied by $h(k)$, $k = 1, 2, \ldots, S-1$, and its boundary conditions.

g) Find a particular solution of the equation of Question (e).

h) Solve the equation of Question (c).

   Hint: the general solution of the equation is the sum of a particular solution and a solution of the associated *homogeneous equation*.

i) How does the mean duration $h(k)$ behave as p goes to zero? Is this compatible with your intuition of the problem? Why?

j) How do the values of $g(k)$ and $h(k)$ behave for fixed $k \in \{1, 2, \ldots, S-1\}$ as S tends to infinity?

Problem 2.9 (Anděl and Hudecová (2012)). Consider a gambling process $(X_n)_{n\geqslant 0}$ on the state space $\mathbb{S} = \{0, 1, \ldots, S\}$, with transition probabilities

$$\mathbb{P}(X_{n+1} = k+1 \mid X_n = k) = p, \qquad \mathbb{P}(X_{n+1} = k-1 \mid X_n = k) = q,$$

$k = 1, 2, \ldots, S-1$, with $p + q = 1$. Let

$$\tau := \inf\{n \geqslant 0 \ : \ X_n = 0 \text{ or } X_n = S\}$$

denote the time until the process hits either state $\textcircled{0}$ or state $\textcircled{S}$, and consider the second moment

$$h(k) := \mathbb{E}\big[\tau^2 \mid X_0 = k\big],$$

of τ after starting from $k = 0, 1, 2, \ldots, S$.

a) Give the values of $h(0)$ and $h(S)$.

b) Using first step analysis, find an equation satisfied by $h(k)$ and involving $\mathbb{E}[\tau \mid X_0 = k+1]$ and $\mathbb{E}[\tau \mid X_0 = k-1]$, $k = 1, 2, \ldots, S-1$.

c) From now on we take $\underline{p = q = 1/2}$. Recall that in this case we have

$$\mathbb{E}[\tau \mid X_0 = k] = (S-k)k, \qquad k = 0, 1, \ldots, S.$$

Show that the function $h(k)$ satisfies the finite difference equation

$$h(k) = -1 + 2(S-k)k + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1), \qquad k = 1, 2, \ldots, S-1. \tag{2.3.26}$$

d) Knowing that

$$k \longmapsto \frac{2}{3}k^2 - \frac{2S}{3}k^3 + \frac{k^4}{3}$$

is a particular solution of the equation (2.3.26) of Question (c), and that the solution of the *homogeneous equation*

$$f(k) = \frac{1}{2}f(k+1) + \frac{1}{2}f(k-1), \qquad k = 1, 2, \ldots, S-1,$$

takes the form

$$f(k) = C_1 + C_2 k,$$

compute the value of the expectation $h(k)$ solution of (2.3.26) for all $k = 0, 1, \ldots, S$.

e) Compute the variance

$$v(k) = \mathbb{E}[\tau^2 \mid X_0 = k] - (\mathbb{E}[\tau \mid X_0 = k])^2$$

of the game duration starting from $k = 0, 1, \ldots, S$.

f) Compute $v(1)$ when $\underline{S = 2}$ and explain why the result makes pathwise sense.

Chapter 3
Random Walks

In this chapter we consider our second important example of discrete-time stochastic process, which is a random walk allowed to evolve over the set $\mathbb{Z}$ of signed integers without any boundary restriction. Of particular importance are the probabilities of return to a given state in finite time, as well as the corresponding mean return time.

3.1 Unrestricted Random Walk

The simple unrestricted random walk $(S_n)_{n\geqslant 0}$, also called the *Bernoulli* random walk, is defined by $S_0 = 0$ and

$$S_n = \sum_{k=1}^{n} X_k = X_1 + \cdots + X_n, \qquad n \geqslant 1,$$

where the random walk *increments* $(X_k)_{k\geqslant 1}$ form a family of independent, $\{-1, +1\}$-valued random variables.

We will assume in addition that the family $(X_k)_{k\geqslant 1}$ is *i.i.d.*, *i.e.* it is made of *independent and identically distributed* Bernoulli random variables, with distribution

$$\begin{cases} \mathbb{P}(X_k = +1) = p, \\ \\ \mathbb{P}(X_k = -1) = q, \qquad k \geqslant 1, \end{cases}$$

with $p + q = 1$.

3.2 Mean and Variance

In this case the mean and variance of X_n are given by

$$\mathbb{E}[X_n] = -1 \times q + 1 \times p = 2p - 1 = p - q,$$

and

$$\begin{aligned}\mathrm{Var}[X_n] &= \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 \\ &= 1 \times q + 1 \times p - (2p-1)^2 \\ &= 4p(1-p) = 4pq.\end{aligned}$$

As a consequence, we find that

$$\mathbb{E}[S_n \mid S_0 = 0] = \mathbb{E}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathbb{E}\left[X_k\right] = n(2p-1) = n(p-q),$$

and the variance can be computed by (1.6.10) as

$$\mathrm{Var}[S_n \mid S_0 = 0] = \mathrm{Var}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathrm{Var}[X_k] = 4npq,$$

where we used (1.6.10).

3.3 Distribution

First we note that in an even number of time steps, $(S_n)_{n\geqslant 0}$ can only reach an even state in $\mathbb{Z}$ starting from ⓪. Similarly, in an odd number of time steps, $(S_n)_{n\geqslant 0}$ can only reach an odd state in $\mathbb{Z}$ starting from ⓪. Indeed, starting from $S_n = k$ the value of S_{n+2} after two time steps can only belong to $\{k-2, k, k+2\}$. Consequently, we have

$$\begin{cases} \mathbb{P}(S_{2n} = 2k+1 \mid S_0 = 0) = 0, & k \in \mathbb{Z},\ n \geqslant 0, \\ \\ \mathbb{P}(S_{2n+1} = 2k \mid S_0 = 0) = 0, & k \in \mathbb{Z},\ n \geqslant 0, \end{cases} \tag{3.3.1}$$

and

$$\mathbb{P}(S_n = k \mid S_0 = 0) = 0, \qquad \text{for } k < -n \text{ or } k > n, \tag{3.3.2}$$

since $S_0 = 0$. Next, let l denote the number of upwards steps between time 0 and time $2n$, whereas $2n - l$ will denote the number of downwards steps. If $S_{2n} = 2k$ we have

$$2k = l - (2n - l) = 2l - 2n,$$

hence there are $l = n + k$ upwards steps and $2n - l = n - k$ downwards steps, $-n \leqslant k \leqslant n$. The probability of a given paths having $l = n + k$ upwards steps and $2n - l = n - k$ downwards steps is

$$p^{n+k} q^{n-k}$$

and in order to find $\mathbb{P}(S_{2n} = 2k \mid S_0 = 0)$ we need to multiply this probability by the total number of paths leading from ⓪ to $\boxed{2k}$ in $2n$ steps. We find that this number of paths is

$$\binom{2n}{n+k} = \binom{2n}{n-k}$$

which represents the number of ways to arrange $n + k$ upwards steps (or $n - k$ downwards steps) within $2n$ time steps.

Hence we have

$$\mathbb{P}(S_{2n} = 2k \mid S_0 = 0) = \binom{2n}{n+k} p^{n+k} q^{n-k}, \quad -n \leqslant k \leqslant n, \quad (3.3.3)$$

in addition to (3.3.1) and (3.3.2). In Figure 3.1 we enumerate the $120 = \binom{10}{7} = \binom{10}{3}$ possible paths corresponding to $n = 5$ and $k = 2$.

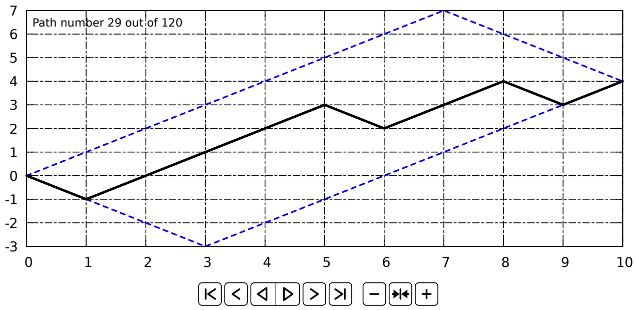


Fig. 3.1: Graph of $120 = \binom{10}{7} = \binom{10}{3}$ paths linking $(0, 0)$ to $(10, 4)$.*

Exercises:

i) Show by a similar analysis that

$$\mathbb{P}(S_{2n+1} = 2k+1 \mid S_0 = 0) = \binom{2n+1}{n+k+1} p^{n+k+1} q^{n-k}, \quad -n \leqslant k \leqslant n, \tag{3.3.4}$$

i.e. $(2n+1+S_{2n+1})/2$ is a binomial random variable with parameter $(2n+1,p)$, and

$$\mathbb{P}\left(\frac{2n+1+S_{2n+1}}{2} = k \;\middle|\; S_0 = 0\right) = \mathbb{P}\left(S_{2n+1} = 2k-2n-1 \mid S_0 = 0\right) = \binom{2n+1}{k} p^k q^{2n+1-k},$$

$k = 0, 1, \ldots, 2n+1$.

ii) Show that $n + S_{2n}/2$ is a binomial* random variable with parameter $(2n,p)$, *i.e.*, show that

$$\mathbb{P}\left(n + \frac{S_{2n}}{2} = k \;\middle|\; S_0 = 0\right) = \mathbb{P}\left(S_{2n} = 2k-2n \mid S_0 = 0\right) = \binom{2n}{k} p^k q^{2n-k}, \qquad k = 0, 1, \ldots, 2n.$$

3.4 First Return to Zero

Let

$$T_0^r := \inf\{n \geqslant 1 \ : \ S_n = 0\}$$

denote the time of first return to ⓪ of the random walk started at ⓪, with the convention $\inf \emptyset = +\infty$.† We are interested in particular in computing the mean time $\mathbb{E}[T_0^r \mid S_0 = 0]$ it takes to return to state ⓪ after starting from state ⓪, see Figure 3.2.

* Animated figure (works in Acrobat Reader).

* Note that S_{2n} is always an even number after we start from $S_0 = 0$.

† Recall that the notation "inf" stands for "infimum", meaning the smallest $n \geqslant 0$ such that $S_n = 0$, with $T_0^r = +\infty$ if no such $n \geqslant 0$ exists.

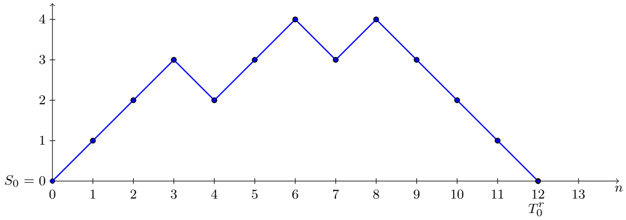


Fig. 3.2: Sample path of the random walk $(S_n)_{n \geqslant 0}$.

We are interested in computing the distribution

$$g(n) = \mathbb{P}(T_0^r = n \mid S_0 = 0), \qquad n \geqslant 1,$$

of the first return time T_0^r to ⓪. It is easy to show by pathwise analysis that T_0^r can only be even-valued starting from ⓪, hence $g(2k+1) = 0$ for all $k \in \mathbb{N}$, and in particular we have

$$\mathbb{P}(T_0^r = 1 \mid S_0 = 0) = 0, \quad \mathbb{P}(T_0^r = 2 \mid S_0 = 0) = 2pq, \tag{3.4.1}$$

and

$$\mathbb{P}(T_0^r = 4 \mid S_0 = 0) = 2p^2q^2, \tag{3.4.2}$$

by considering the two paths leading from ⓪ to ⓪ in two steps and the only two paths leading from ⓪ to ⓪ in four steps without hitting ⓪. However the computation of $\mathbb{P}(T_0^r = 2n \mid S_0 = 0)$ by this method is difficult to extend to all $n \geqslant 3$.

In order to completely solve this problem we will rely on the computation of the probability generating function $G_{T_0^r}$ of T_0^r, cf. (3.4.9) below.

This computation will use the following tools:

- convolution equation, see Relation (3.4.3) below,

- Taylor expansions, see Relation (3.4.22) below,

- probability generating functions.

First, we will need the following Lemma 3.1 which will be used in the proof of Lemma 3.4 below.

Lemma 3.1. *(Convolution equation). The function*

$$g : \{1, 2, 3, \ldots\} \longrightarrow [0, 1]$$

$$n \longmapsto g(n)$$

defined by

$$g(n) := \mathbb{P}(T_0^r = n \mid S_0 = 0), \qquad n \geqslant 1,$$

satisfies the convolution equation

$$h(n) = \sum_{k=0}^{n-2} g(n-k)h(k), \qquad n \geqslant 1, \tag{3.4.3}$$

with the initial condition $g(1) = 0$*, where* $h(n) := \mathbb{P}(S_n = 0 \mid S_0 = 0)$ *is given from* (3.3.3) *by*

$$h(2n) = \binom{2n}{n} p^n q^n, \quad \text{and} \quad h(2n+1) = 0, \qquad n \in \mathbb{N}. \tag{3.4.4}$$

Proof. We first partition the event $\{S_n = 0\}$ into

$$\{S_n = 0\} = \bigcup_{k=0}^{n-2} \{S_k = 0, \ S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0, \ S_n = 0\}, \qquad n \geqslant 1,$$

according to all possible times $k = 0, 1, \ldots, n-2$ of *last* return to state ⓪ *before* time n, with $\{S_1 = 0\} = \emptyset$ since we are starting from $S_0 = 0$, see Figure 3.3.

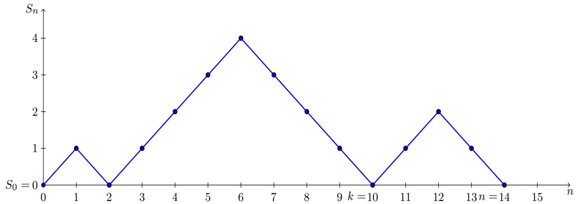


Fig. 3.3: Last return to state 0 at time $k = 10$.

Then we have

$$h(n) := \mathbb{P}(S_n = 0 \mid S_0 = 0)$$
$$= \sum_{k=0}^{n-2} \mathbb{P}\big(S_k = 0, \ S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0, \ S_n = 0 \mid S_0 = 0\big)$$

$$
\begin{aligned}
&= \sum_{k=0}^{n-2} \mathbb{P}\big(S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0,\ S_n = 0 \mid S_k = 0,\ S_0 = 0\big)\mathbb{P}(S_k = 0 \mid S_0 = 0) \\
&= \sum_{k=0}^{n-2} \mathbb{P}\big(S_{k+1} \neq 0, \ldots, S_{n-1} \neq 0,\ S_n = 0 \mid S_k = 0\big)\mathbb{P}(S_k = 0 \mid S_0 = 0) \quad (3.4.5) \\
&= \sum_{k=0}^{n-2} \mathbb{P}\big(S_1 \neq 0, \ldots, S_{n-k-1} \neq 0,\ S_{n-k} = 0 \mid S_0 = 0\big)\mathbb{P}(S_k = 0 \mid S_0 = 0) \quad (3.4.6) \\
&= \sum_{k=0}^{n-2} \mathbb{P}(T_0^r = n-k \mid S_0 = 0)\mathbb{P}(S_k = 0 \mid S_0 = 0) \\
&= \sum_{k=0}^{n-2} h(k)g(n-k), \qquad n \geqslant 1, \quad (3.4.7)
\end{aligned}
$$

where from (3.4.5) to (3.4.6) we applied a backward shift of k steps in time. □

We now need to solve the convolution equation (3.4.3) for $g(n) = \mathbb{P}(T_0^r = n \mid S_0 = 0)$, $n \geqslant 1$, knowing that $g(1) = 0$. For this we will derive a simple equation for the probability generating function

$$
\begin{aligned}
G_{T_0^r} : [-1, 1] &\longrightarrow \mathbb{R} \\
s &\longmapsto G_{T_0^r}(s)
\end{aligned}
$$

of the random variable T_0^r, defined by

$$
G_{T_0^r}(s) := \mathbb{E}\big[s^{T_0^r}\mathbb{1}_{\{T_0^r<\infty\}} \,\big|\, S_0 = 0\big] = \sum_{n\geqslant 0} s^n \mathbb{P}(T_0^r = n \mid S_0 = 0) = \sum_{n\geqslant 0} s^n g(n),
$$

$-1 \leqslant s \leqslant 1$, cf. (1.7.3).

Recall that the knowledge of $G_{T_0^r}(s)$ provides certain information on the distribution of T_0^r, such as the probability

$$
\mathbb{P}(T_0^r < \infty \mid S_0 = 0) = \mathbb{E}\big[\mathbb{1}_{\{T_0^r<\infty\}} \,\big|\, S_0 = 0\big] = G_{T_0^r}(1)
$$

and the expected value

$$
\mathbb{E}\big[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \,\big|\, S_0 = 0\big] = \sum_{n\geqslant 1} n\mathbb{P}(T_0^r = n \mid S_0 = 0) = G'_{T_0^r}(1^-).
$$

In Lemma 3.4 below we will compute $G_{T_0^r}(s)$ for all $s \in [-1, 1]$. First, let the function

$$H : \mathbb{R} \longrightarrow \mathbb{R}$$
$$s \longmapsto H(s)$$

be defined by

$$H(s) := \sum_{k\geqslant 0} h(k)s^k = \sum_{k\geqslant 0} s^k \mathbb{P}(S_k = 0 \mid S_0 = 0), \qquad -1 \leqslant s \leqslant 1.$$

In the following lemma we show that the function $H(s)$ can be computed in closed form.

Proposition 3.2. *We have*

$$H(s) = (1 - 4pqs^2)^{-1/2}, \qquad |s| < \frac{1}{2\sqrt{pq}}.$$

Proof. By (3.4.4) and the fact that $\mathbb{P}(S_{2k+1} = 0 \mid S_0 = 0) = 0$, $k \in \mathbb{N}$, we have

$$\begin{aligned}
H(s) &= \sum_{k\geqslant 0} s^k \mathbb{P}(S_k = 0 \mid S_0 = 0) \qquad (3.4.8)\\
&= \sum_{k\geqslant 0} s^{2k} \mathbb{P}(S_{2k} = 0 \mid S_0 = 0) = \sum_{k\geqslant 0} s^{2k} \binom{2k}{k} p^k q^k \\
&= \sum_{k\geqslant 0} (pq)^k s^{2k} \frac{(2k)(2k-1)(2k-2)(2k-3) \times \cdots \times 4 \times 3 \times 2 \times 1}{(k(k-1) \times \cdots \times 2 \times 1)^2} \\
&= \sum_{k\geqslant 0} (4pq)^k s^{2k} \frac{k(k-1/2)(k-2/2)(k-3/2) \times \cdots \times (4/2) \times (3/2) \times (2/2) \times (1/2)}{(k(k-1) \times \cdots \times 2 \times 1)^2} \\
&= \sum_{k\geqslant 0} (4pq)^k s^{2k} \frac{(k-1/2)(k-3/2) \times \cdots \times (3/2) \times (1/2)}{k(k-1) \times \cdots \times 2 \times 1} \\
&= \sum_{k\geqslant 0} (-1)^k (4pq)^k s^{2k} \frac{(-1/2-(k-1))(3/2-k) \times \cdots \times (-3/2) \times (-1/2)}{k(k-1) \times \cdots \times 2 \times 1} \\
&= \sum_{k\geqslant 0} (-4pqs^2)^k \frac{(-1/2) \times (-3/2) \times \cdots \times (3/2-k)(-1/2-(k-1))}{k!} \\
&= (1 - 4pqs^2)^{-1/2},
\end{aligned}$$

for $|4pqs^2| < 1$,* see for example *here*. □

* We used the Taylor expansion $(1+x)^\alpha = \sum_{k\geqslant 0} \frac{x^k}{k!}\alpha(\alpha-1) \times \cdots \times (\alpha-(k-1))$, cf. Relation (A.8).

Remark 3.3. *We note that, taking $s = 1$, by (1.6.1) we have*

$$\begin{aligned} H(1) &= \sum_{k\geqslant 0} \mathbb{P}(S_k = 0 \mid S_0 = 0) \\ &= \sum_{k\geqslant 0} \mathbb{E}[\mathbb{1}_{\{S_k=0\}} \mid S_0 = 0] \\ &= \mathbb{E}\left[\sum_{k\geqslant 0} \mathbb{1}_{\{S_k=0\}} \;\middle|\; S_0 = 0\right], \end{aligned}$$

hence $H(1) = 1/\sqrt{1-4pq}$ represents the mean number of visits of the random walk $(S_n)_{n\geqslant 0}$ to state $\textcircled{0}$.

Next, based on the convolution equation (3.4.3) of Lemma 3.1 we compute $G_{T_0^r}(s)$ in the next Lemma 3.4 by deriving and solving an Equation (3.4.13) for $G_{T_0^r}(s)$. This method has some similarities with the z-transform method used in electrical engineering, see also the command Sum[Bin[2*n,n]*(p*q*s^2)^n/(2*n-1),{n,1,Infinity}].

Lemma 3.4. *The probability generating function $G_{T_0^r}$ of the first return time T_0^r to $\textcircled{0}$ is given by*

$$G_{T_0^r}(s) = 1 - \frac{1}{H(s)} = 1 - \sqrt{1 - 4pqs^2}, \qquad 4pqs^2 < 1. \tag{3.4.9}$$

Proof. We have, taking into account the relation $g(1) = \mathbb{P}(T_0^r = 1 \mid S_0 = 0) = 0$,

$$G_{T_0^r}(s)H(s) = \left(\sum_{n\geqslant 1} s^n g(n)\right)\left(\sum_{k\geqslant 0} s^k h(k)\right)$$

$$= \sum_{n\geqslant 2}\sum_{k\geqslant 0} s^{n+k} g(n)h(k) = \sum_{k\geqslant 0}\sum_{n\geqslant 2} s^{n+k} g(n)h(k) \tag{3.4.10}$$

$$= \sum_{l\geqslant 2} s^l \sum_{k=0}^{l-2} g(l-k)h(k) \tag{3.4.11}$$

$$= \sum_{l\geqslant 1} s^l h(l) \tag{3.4.12}$$

$$= \sum_{l\geqslant 1} s^l \mathbb{P}(S_l = 0 \mid S_0 = 0)$$

$$= -1 + \sum_{l\geqslant 0} s^l \mathbb{P}(S_l = 0 \mid S_0 = 0) = H(s) - 1,$$

where from line (3.4.10) to line (3.4.11) we have applied the change of variable $(k, n) \longmapsto (k, l)$ with $l = n + k$, and from line (3.4.11) to line (3.4.12) we have used the convolution equation (3.4.3) of Lemma 3.1. This shows that $G_{T_0^r}(s)$ satisfies the equation

$$G_{T_0^r}(s) H(s) = H(s) - 1, \qquad 4pqs^2 < 1. \tag{3.4.13}$$

Solving (3.4.13) yields the value of $G_{T_0^r}(s)$ for all s such that $4pqs^2 < 1$. □

See Exercise 3.4-(e) for another derivation of (3.4.9) based on first step analysis.*

We will apply our knowledge of $G_{T_0^r}(s)$ to the computation of the first return time distribution of T_0^r, the probability of return to ⓪ in finite time, and the mean return time to ⓪.

Probability of return to zero in finite time

The probability that the first return to ⓪ occurs within a finite time is

$$\begin{aligned} \mathbb{P}(T_0^r < \infty \mid S_0 = 0) &= \mathbb{E}\left[\mathbb{1}_{\{T_0^r < \infty\}} \,\middle|\, S_0 = 0\right] \\ &= \mathbb{E}\left[1^{T_0^r} \mathbb{1}_{\{T_0^r < \infty\}} \,\middle|\, S_0 = 0\right] \\ &= G_{T_0^r}(1) = 1 - \sqrt{1 - 4pq} \\ &= 1 - |2p - 1| = 1 - |p - q| = \begin{cases} 2q, & p \geqslant 1/2, \\ \\ 2p, & p \leqslant 1/2, \end{cases} \\ &= 2\min(p, q), \end{aligned} \tag{3.4.14}$$

hence

$$\mathbb{P}(T_0^r = \infty \mid S_0 = 0) = |2p - 1| = |p - q|. \tag{3.4.15}$$

* "Any good theorem should have several proofs, the more the better. For two reasons: usually, different proofs have different strengths and weaknesses, and they generalise in different directions - they are not just repetitions of each other. Some of them are good for this application, some are good for that application. They all shed light on the area. If you cannot look at a problem from different directions, it is probably not very interesting; the more perspectives, the better !" - Sir Michael Atiyah.

see also the command Sum[Bin[2*n,n]*(p*q)^n/(2*n-1),{n,1,Infinity}], or using the following Python code:

```
from sympy import *
import sympy as sp
k = sp.Symbol("k");p = sp.Symbol("p"); q = sp.Symbol("q")
prob=summation(p**k*q**k*factorial(2*k)/factorial(k)**2/(2*k-1), (k, 1, oo))
simplify(prob.args[0][0])
```

Note that in (2.2.14) above we have shown that the probability of hitting state ⓪ in finite time starting from any state ⓚ with $k \geqslant 1$ is given by

$$\mathbb{P}\big(T_0^r < \infty \mid S_0 = k\big) = \min\left(1, \left(\frac{q}{p}\right)^k\right), \qquad k \geqslant 1, \tag{3.4.16}$$

i.e.

$$\mathbb{P}(T_0^r = \infty \mid S_0 = k) = \mathrm{Max}\left(0, 1 - \left(\frac{q}{p}\right)^k\right), \qquad k \geqslant 1,$$

see Exercise 3.2-(c), Exercise 3.4-(c) and Problem 3.15-(e), and also Exercises 3.4-(b) and 5.10-(a) for other derivations of (3.4.16).

In the non-symmetric case $p \neq q$, Relation (3.4.14) shows that

$$\mathbb{P}(T_0^r < \infty \mid S_0 = 0) < 1 \qquad \text{and} \qquad \mathbb{P}(T_0^r = \infty \mid S_0 = 0) > 0,$$

whereas in the symmetric case (or fair game) $p = q = 1/2$ we find that

$$\mathbb{P}(T_0^r < \infty \mid S_0 = 0) = 1 \qquad \text{and} \qquad \mathbb{P}(T_0^r = \infty \mid S_0 = 0) = 0,$$

i.e. the random walk returns to ⓪ with probability one.

Mean return time to zero

i) In the non-symmetric case $p \neq q$, by (3.4.15), the time T_0^r needed to return to state ⓪ is infinite with probability

$$\mathbb{P}(T_0^r = \infty \mid S_0 = 0) = |p - q| > 0,$$

hence the expected value*

$$\begin{aligned}\mathbb{E}[T_0^r \mid S_0 = 0] &= \infty \times \mathbb{P}(T_0^r = \infty \mid S_0 = 0) + \sum_{k \geqslant 1} k \mathbb{P}(T_0^r = k \mid S_0 = 0) \\ &= \infty \end{aligned} \tag{3.4.17}$$

is infinite in that case.

* Note that the summation $\sum_{k \geqslant 1} = \sum_{1 \leqslant k < \infty}$ actually *excludes* the value $k = \infty$.

Note that starting from $S_0 = k \geqslant 1$, by (2.3.13) we have found that the mean hitting time of state (0) equals

$$\mathbb{E}[T_0^r \mid S_0 = k] = \begin{cases} \infty & \text{if } q \leqslant p, \\ \dfrac{k}{q-p} & \text{if } q > p, \end{cases} \tag{3.4.18}$$

see also Problem 3.15-(f).

In particular, we have $\mathbb{P}(T_0^r < \infty \mid S_0 = k) = 1$ when $q > p$ and $k \geqslant 1$, which is consistent with (3.4.16). See Exercise 5.10-(b) for other derivations of (2.3.13)-(3.4.18) using the probability generating function $s \mapsto G_{T_0^r}(s)$.

Remark 3.5. *By (3.4.9), the truncated expectation* $\mathbb{E}\big[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \mid S_0 = 0\big]$ *satisfies*

$$\begin{aligned}
\mathbb{E}\big[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \mid S_0 = 0\big] &= \sum_{n\geqslant 1} n \mathbb{P}(T_0^r = n \mid S_0 = 0) \\
&= G'_{T_0^r}(1^-) \\
&= \frac{4pqs}{\sqrt{1-4pqs^2}}\Big|_{s=1} \\
&= \frac{4pq}{\sqrt{1-4pq}} \\
&= \frac{4pq}{|p-q|},
\end{aligned} \tag{3.4.19}$$

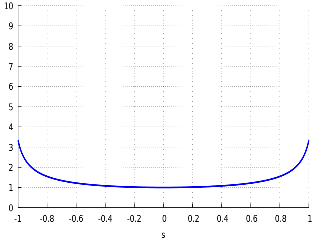
when $p \neq q$, see for example Figure 3.4-a), this code and the following Python code:

```
from sympy import *
import sympy as sp
n = sp.Symbol("n");p = sp.Symbol("p"); q = sp.Symbol("q")
expectation=summation(2*n*p**n*q**n*factorial(2*n)/factorial(n)**2/(2*n-1), (n, 1, oo))
expectation.args[1].args[0][0]
```

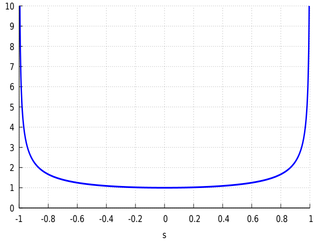
which shows in particular from (3.4.14) and Lemma 1.10 that

$$\begin{aligned}
\mathbb{E}[T_0^r \mid T_0^r < \infty,\ S_0 = 0] &= \frac{1}{\mathbb{P}(T_0^r < \infty \mid S_0 = 0)} \mathbb{E}[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \mid S_0 = 0] \\
&= \frac{1}{\min(p,q)} \times \frac{2pq}{|p-q|} = 2\frac{\operatorname{Max}(p,q)}{|p-q|},
\end{aligned}$$

see (3.4.14).



(a) Graph of $G_{T_0^r}(s)$ with $p = 0.35$.



(b) Graph of $G_{T_0^r}(s)$ with $p = 0.5$.

Fig. 3.4: Probability generating functions of T_0^r for $p = 0.35$ and $p = 0.5$.

ii) In the symmetric case $p = q = 1/2$ we have $\mathbb{P}(T_0^r < \infty \mid S_0 = 0) = 1$ and

$$\mathbb{E}[T_0^r \mid S_0 = 0] = \mathbb{E}\big[T_0^r \mathbb{1}_{\{T_0^r < \infty\}} \,\big|\, S_0 = 0\big] = G'_{T_0^r}(1^-) = \infty \quad (3.4.20)$$

as the slope of $s \mapsto G_{T_0^r}(s)$ in Figure 3.4-b) is infinite at $s = 1$, or by taking the limit as $p, q \to 1/2$ in (3.4.18) or (3.4.19).

When $p = q = 1/2$ the random walk returns to state $\textcircled{0}$ with probability *one* within a *finite* (random) time, while the average of this random time is *infinite*. This yields another example of a random variable T_0^r which is almost surely finite, while its expectation is infinite as in the St. Petersburg paradox.

```
nsim <- 10000; N=1000000; T<-2.0; t <- 0:(N-1); dt <- 1; mean=0.0;
for (i in 1:nsim){signal=0;colour="blue";
Z <- 2*(rbinom(N,1,0.5)-0.5);X <- c(1,N+1);X[1]=0;j=1;
while (j<N && signal==0){j=j+1;X[j]=X[j-1]+Z[j];if (X[j]==0) {signal=1;mean=mean+j-1}}
plot(t[1:j], X[1:j], xlab = "t", ylab = "", type = "p", ylim = c(min(X[1:j])-max(X
[1:j]),-min(X[1:j])+max(X[1:j])), col = colour,main=paste("Time=",j-1,", Mean=
",mean,"/",i,"=",round(mean/i, digits=1)), xaxs="i", xaxt="n",lwd=3)
lines(t[1:j], X[1:j], type = "l",col="blue",lwd=2)
axis(side=1, at=c(0:j), c(0:j));axis(side=1, pos=0, at=c(0:j), c(0:j))
text((j-1)/2,0.5,paste(j-1),cex=5);
readline(prompt = "Pause. Press <Enter>...")}
```

This shows how even a fair game can be risky when the player's wealth is negative, as it will take on average an infinite time to recover the losses.

First return time distribution

Proposition 3.6 can also be obtained from the path counting result of Exercise 3.9, which shows that the number of path linking $\textcircled{0}$ to $\textcircled{0}$ over $2k$ time steps without hitting $\textcircled{0}$ in between is given by $\dfrac{1}{2k-1}\dbinom{2k}{k}$, $k \geqslant 1$.

Proposition 3.6. *The probability distribution* $\mathbb{P}\big(T_0^r = n \mid S_0 = 0\big)$ *of the first return time* T_0^r *to* ⓪ *is given by*

$$\mathbb{P}(T_0^r = 2k \mid S_0 = 0) = \frac{1}{2k-1}\binom{2k}{k}(pq)^k, \qquad k \geqslant 1, \tag{3.4.21}$$

with $\mathbb{P}\big(T_0^r = 2k+1 \mid S_0 = 0\big) = 0$, $k \in \mathbb{N}$.

Proof. By applying a Taylor expansion to $s \longmapsto 1-(1-4pqs^2)^{1/2}$ in (3.4.9), we get

$$\begin{aligned} G_{T_0^r}(s) &= 1-(1-4pqs^2)^{1/2} \\ &= 1-\sum_{k\geqslant 0}\frac{1}{k!}(-4pqs^2)^k\left(\frac{1}{2}-0\right)\left(\frac{1}{2}-1\right)\times\cdots\times\left(\frac{1}{2}-(k-1)\right) \\ &= \frac{1}{2}\sum_{k\geqslant 1}s^{2k}\frac{(4pq)^k}{k!}\left(1-\frac{1}{2}\right)\times\cdots\times\left(k-1-\frac{1}{2}\right), \end{aligned} \tag{3.4.22}$$

where we used (A.8) for $\alpha = 1/2$. By identification of (3.4.22) with the expansion

$$G_{T_0^r}(s) = \sum_{n\geqslant 0} s^n \mathbb{P}(T_0^r = n \mid S_0 = 0), \qquad -1 \leqslant s \leqslant 1,$$

we obtain

$$\begin{aligned} \mathbb{P}(T_0^r = 2k \mid S_0 = 0) &= g(2k) \\ &= \frac{(4pq)^k}{k!}\frac{1}{2}\left(1-\frac{1}{2}\right)\times\cdots\times\left(k-1-\frac{1}{2}\right) \\ &= \frac{(4pq)^k}{2k!}\prod_{m=1}^{k-1}\left(m-\frac{1}{2}\right) \\ &= \frac{1}{2k-1}\binom{2k}{k}(pq)^k, \qquad k \geqslant 1, \end{aligned}$$

while $\mathbb{P}\big(T_0^r = 2k+1 \mid S_0 = 0\big) = g(2k+1) = 0$, $k \in \mathbb{N}$. This conclusion could also be obtained using (1.7.10) from the relation

$$\mathbb{P}(T_0^r = n \mid S_0 = 0) = \frac{1}{n!}\frac{\partial^n}{\partial s^n}G_{T_0^r}(s)_{|s=0}, \qquad n \geqslant 0.$$

□

Exercise: Check that the formula (3.4.21) recovers (3.4.1) and (3.4.2) when $k = 0, 1, 2$.

Using the independence of increments of the random walk $(S_n)_{n\geqslant 0}$, one can also show that the probability generating function of the first passage time

$$T_k := \inf\{n \geqslant 0 \ : \ S_n = k\}$$

to any level $k \geqslant 1$ is given by

$$G_{T_k}(s) = \left(\frac{1-\sqrt{1-4pqs^2}}{2qs}\right)^k, \qquad 4pqs^2 < 1, \quad q \leqslant p, \tag{3.4.23}$$

from which the distribution of T_k can be computed given the series expansion of $G_{T_k}(s)$, cf. Exercise 3.4-(b) and (S.27) below with $i := -k$.

The gambling process of Chapter 2 and the standard random walk $(S_n)_{n\geqslant 0}$ will later be reconsidered as particular cases in the general framework of Markov chains of Chapters 4 and 5.

Exercises

Exercise 3.1 We consider the simple random walk $(S_n)_{n\geqslant 0}$ of Section 3.1 with independent increments and started at $S_0 = 0$, in which the probability of advance is p and the probability of retreat is $1-p$.

a) Enumerate all possible sample paths that conduct to $S_4 = 2$ starting from $S_0 = 0$.
b) Show that

$$\mathbb{P}(S_4 = 2 \mid S_0 = 0) = \binom{4}{3} p^3(1-p) = \binom{4}{1} p^3(1-p).$$

c) Show that we have

$$\mathbb{P}(S_n = k \mid S_0 = 0) \tag{3.4.24}$$
$$= \begin{cases} \displaystyle\binom{n}{(n+k)/2} p^{(n+k)/2}(1-p)^{(n-k)/2}, & n+k \text{ even and } |k| \leqslant n, \\ 0, & n+k \text{ odd or } |k| > n. \end{cases}$$

d) Show, by a direct argument using a "last step" analysis at time $n+1$ on random walks, that $p_{n,k} := \mathbb{P}(S_n = k \mid S_0 = 0)$ satisfies the difference equation

$$p_{n+1,k} = pp_{n,k-1} + qp_{n,k+1}, \tag{3.4.25}$$

subject to the boundary conditions $p_{0,0} = 1$ and $p_{0,k} = 0$, $k \neq 0$.

e) Confirm that $p_{n,k} = \mathbb{P}(S_n = k \mid S_0 = 0)$ given by (3.4.24) satisfies the equation (3.4.25) and its boundary conditions.

Exercise 3.2 Consider a random walk $(S_n)_{n\geqslant 0}$ on $\mathbb{Z}$ with independent increments and probabilities p, resp. $q = 1-p$ of moving up by one step, resp. down by one step. Let

$$T_0 = \inf\{n \geqslant 0 \ : \ S_n = 0\}$$

denote the hitting time of state ⓪.

a) Explain why for any $k \geqslant 1$ we have

$$\mathbb{E}[T_0 \mid S_0 = k] = k\mathbb{E}[T_0 \mid S_0 = 1],$$

and compute $\mathbb{E}[T_0 \mid S_0 = 1]$ using first step analysis when $q > p$. What can we conclude when $p \geqslant q$?

b) Explain why, by the Markov property, we have

$$\mathbb{P}(T_0 < \infty \mid S_0 = k) = (\mathbb{P}(T_0 < \infty \mid S_0 = 1))^k, \qquad k \geqslant 1.$$

c) Using first step analysis for random walks, show that $\alpha := \mathbb{P}(T_0 < \infty \mid S_0 = 1)$ satisfies the quadratic equation

$$p\alpha^2 - \alpha + q = p(\alpha - q/p)(\alpha - 1) = 0,$$

and give the values of $\mathbb{P}(T_0 < \infty \mid S_0 = 1)$ and $\mathbb{P}(T_0 = \infty \mid S_0 = 1)$ in the cases $p < q$ and $p \geqslant q$ respectively.

Exercise 3.3 Consider the random walk

$$S_n := X_1 + \cdots + X_n, \qquad n \geqslant 1,$$

with $\underline{S_0 = 0}$, where $(X_k)_{k\geqslant 1}$ is a sequence of Bernoulli random variables with

$$\mathbb{P}(X_k = 1) = p \in (0,1), \qquad \mathbb{P}(X_k = -1) = q \in (0,1),$$

and $p + q = 1$. Recall that the probability generating function (PGF)

$$G_{T_0^r}(s) = \sum_{k\geqslant 0} s^k \mathbb{P}(T_0^r = k), \qquad s \in [-1,1], \tag{3.4.26}$$

of the first return time $T_0^r := \inf\{S_n = 0 \ : \ n \geqslant 1\}$ to state ⓪ is given by

$$G_{T_0^r}(s) = 1 - \sqrt{1 - 4pqs^2}, \qquad s \in [-1,1]. \tag{3.4.27}$$

a) Compute $\mathbb{P}\big(T_0^r = 0\big)$ and $\mathbb{P}\big(T_0^r < \infty\big)$ from $G_{T_0^r}$.

b) By differentiation of (3.4.26) and (3.4.27), compute $\mathbb{P}(T_0^r = 1)$, $\mathbb{P}(T_0^r = 2)$, $\mathbb{P}(T_0^r = 3)$ and $\mathbb{P}(T_0^r = 4)$ using the probability generating function $G_{T_0^r}$.

c) Compute $\mathbb{E}[T_0^r \mid T_0^r < \infty]$ using the probability generating function $G_{T_0^r}$.

Exercise 3.4 Consider a simple random walk $(X_n)_{n\geqslant 0}$ on $\mathbb{Z}$ with respective probabilities p and q of increment and decrement. Let

$$T_0 := \inf\{n \geqslant 0 \ : \ X_n = 0\}$$

denote the first hitting time of state ⓪, and consider the probability generating function

$$G_i(s) := \mathbb{E}[s^{T_0} \mid X_0 = i], \qquad -1 < s < 1, \quad i \in \mathbb{Z}.$$

a) By a first step analysis argument, find the finite difference equation satisfied by $G_i(s)$, and its boundary condition(s) at $i = 0$ and $i = \pm\infty$.

b) Find the value of $G_i(s)$ for all $i \in \mathbb{Z}$ and $s \in (0,1)$, and recover the result of (3.4.23) on the probability generating function of the hitting time T_0 of ⓪ starting from state ⓘ.

c) Recover Relations (2.2.14)-(3.4.16) using $G_i(s)$.

d) Recover Relations (2.3.13)-(3.4.18) by differentiation of $s \mapsto G_i(s)$.

e) Recover the result of (3.4.9) on the probability generating function of the return time T_0^r to ⓪.

Exercise 3.5 Using the probability distribution (3.4.21) of T_0^r, recover the fact that $\mathbb{E}[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \mid S_0 = 0] = \infty$, when $p = q = 1/2$.

Exercise 3.6 Consider a sequence $(X_k)_{k\geqslant 1}$ of independent Bernoulli random variables with

$$\mathbb{P}(X_k = 1) = p, \quad \text{and} \quad \mathbb{P}(X_k = -1) = q, \qquad k \geqslant 1,$$

where $p + q = 1$, and let the process $(M_n)_{n\geqslant 0}$ be defined by $M_0 := 0$ and

$$M_n := \sum_{k=1}^{n} 2^{k-1} X_k, \qquad n \geqslant 1.$$

a) Compute $\mathbb{E}[M_n]$ for all $n \geqslant 0$.

b) Consider the hitting time $\tau := \inf\{n \geqslant 1 \ : \ M_n = 1\}$ and the stopped process

$$M_{\min(n,\tau)} = M_n \mathbb{1}_{\{n<\tau\}} + \mathbb{1}_{\{\tau\leqslant n\}}, \qquad n \geqslant 0.$$

Determine the possible values of $M_{\min(n,\tau)}$, and the probability distribution of $M_{\min(n,\tau)}$ at any time $n \geqslant 1$.

c) Give an interpretation of the stopped process $(M_{\min(n,\tau)})_{n\geqslant 0}$ in terms of strategy in a game started at $M_0 = 0$.
d) Based on the result of Question (b), compute $\mathbb{E}[M_{\min(n,\tau)}]$ for all $n \geqslant 1$.

Exercise 3.7 Winning streaks. Consider a sequence $(X_n)_{n\geqslant 1}$ of independent Bernoulli random variables with the distribution

$$\mathbb{P}(X_n = 1) = p, \qquad \mathbb{P}(X_n = 0) = q, \qquad n \geqslant 1,$$

with $q := 1 - p$. For some $m \geqslant 1$, let $T^{(m)}$ denote the time of the first appearance of m consecutive "1" in the sequence $(X_n)_{n\geqslant 1}$. For example, for $m = 4$ the following sequence

$$(\overset{\overset{1}{\downarrow}}{0}, \overset{\overset{2}{\downarrow}}{1}, \overset{\overset{3}{\downarrow}}{1}, \overset{\overset{4}{\downarrow}}{0}, \underbrace{\overset{\overset{5}{\downarrow}}{1}, \overset{\overset{6}{\downarrow}}{1}, \overset{\overset{7}{\downarrow}}{1}, \overset{\overset{8}{\downarrow}}{1}}_{\text{4 times}}, 0, 1, 1, 0, \ldots)$$

yields $T^{(4)} = 8$.

a) Compute $\mathbb{P}\big(T^{(m)} < m\big)$, $\mathbb{P}\big(T^{(m)} = m\big)$, $\mathbb{P}\big(T^{(m)} = m+1\big)$, and $\mathbb{P}\big(T^{(m)} = m+2\big)$.
b) Show that the probability generating function

$$G_{T^{(m)}}(s) := \mathbb{E}\big[s^{T^{(m)}} \mathbb{1}_{\{T^{(m)}<\infty\}}\big], \qquad s \in (-1, 1),$$

satisfies

$$G_{T^{(m)}}(s) = p^m s^m + \sum_{k=0}^{m-1} p^k q s^{k+1} G_{T^{(m)}}(s), \qquad s \in (-1, 1). \quad (3.4.28)$$

Hint: Look successively at all possible starting patterns of the form

$$(1, \ldots, 1, \overset{\overset{k}{\downarrow}}{1}, 0, \ldots),$$

where $k = 0, 1, \ldots, m$, compute their respective probabilities, and apply a "k-step analysis" argument.
c) From (3.4.28), compute the probability generating function $G_{T^{(m)}}$ of $T^{(m)}$ for all $s \in (-1, 1)$.

Hint: Recall that we have

$$\sum_{k=0}^{m-1} x^k = \frac{1 - x^m}{1 - x}, \qquad x \in (-1, 1).$$

d) From the probability generating function $G_{T^{(m)}}(s)$, compute $\mathbb{E}\big[T^{(m)}\big]$ for all $m \geqslant 1$.

Hint: It can be simpler to differentiate inside (3.4.28) and to use the relation

$$(1-x)\sum_{k=0}^{m-1}(k+1)x^k + mx^m = \frac{1-x^m}{1-x}, \qquad x \in (-1,1).$$

Exercise 3.8 Consider a sequence $(X_n)_{n\in\mathbb{Z}}$ of independent Bernoulli random variables with the distribution

$$\mathbb{P}(X_n = a) = p, \qquad \mathbb{P}(X_n = b) = q, \qquad n \in \mathbb{Z},$$

where $p \in (0,1]$ and $q := 1-p$, and let $m \geqslant 1$ be a fixed integer.

For $n \geqslant 0$, we let Z_n denote the smallest of m and the number of "a" having appeared up to time n since the last occurrence of "b" in the sequence $(X_k)_{k\leqslant n}$. For example, taking $m := 4$, in the following sequence:

$$(\underset{-5}{\underset{\uparrow}{a}}, \underset{-4}{\underset{\uparrow}{b}}, \underset{-3}{\underset{\uparrow}{a}}, \underset{-2}{\underset{\uparrow}{a}}, \underset{-1}{\underset{\uparrow}{a}}, \underset{0}{\underset{\uparrow}{a}}, \underset{1}{\underset{\uparrow}{a}}, \underset{2}{\underset{\uparrow}{a}}, \underset{3}{\underset{\uparrow}{b}}, \underset{4}{\underset{\uparrow}{a}}, \underset{5}{\underset{\uparrow}{a}}, \underset{6}{\underset{\uparrow}{a}}, \underset{7}{\underset{\uparrow}{a}}, \underset{8}{\underset{\uparrow}{b}}, \underset{9}{\underset{\uparrow}{a}}, \underset{10}{\underset{\uparrow}{a}}, \underset{11}{\underset{\uparrow}{b}}, \ldots),$$

we have

$$Z_0 = 4,\ Z_1 = 5,\ Z_2 = 6,\ Z_3 = 0,\ Z_4 = 1,\ Z_5 = 2,\ Z_6 = 3,\ Z_7 = 4,\ Z_8 = 0,\ Z_9 = 1.$$

a) Show that $(Z_n)_{n\geqslant 0}$ is a Markov chain, give its state space and transition matrix P.
b) Compute the mean hitting time $\mathbb{E}[T_m \mid Z_0 = l]$ of state m by the chain $(Z_n)_{n\geqslant 0}$ after starting from $Z_0 = l$, for $l \in \{0,1,\ldots,m\}$.
c) Give the expected value of the time $T^{(m)}$ of the first appearance of m consecutive "a" in the sequence $(X_n)_{n\geqslant 1}$, and recover the expected value of $T^{(m)}$ obtained in Question (d) of Exercise 3.7.

For example, taking $m := 4$ we have $T^{(4)} = 8$ in the following sequence:

$$(\underset{-3}{\underset{\uparrow}{b}}, \underset{-2}{\underset{\uparrow}{a}}, \underset{-1}{\underset{\uparrow}{a}}, \underset{0}{\underset{\uparrow}{b}}, \overbrace{\underset{1}{\underset{\uparrow}{a}}, \underset{2}{\underset{\uparrow}{a}}, \underset{3}{\underset{\uparrow}{a}}, \underset{4}{\underset{\uparrow}{a}},}^{4\text{ times}} b, a, a, b, \ldots).$$

Exercise 3.9 Consider a random walk $(S_n)_{n\geqslant 0}$ on $\mathbb{Z}$ with increments ± 1, started at $S_0 = 0$. Recall that the number of paths joining states ⓪ and $\boxed{2k}$ over $2m$ time steps is

$$\binom{2m}{m+k}. \tag{3.4.29}$$

a) Compute the total number of paths joining $S_1 = 1$ to $S_{2n-1} = 1$.

Hint: Apply the formula (3.4.29).

b) Compute the total number of paths joining $S_1 = 1$ to $S_{2n-1} = -1$.

Hint: Apply the formula (3.4.29).

c) Show that to every one path joining $S_1 = 1$ to $S_{2n-1} = 1$ by *crossing or hitting* ⓪ we can associate one path joining $S_1 = 1$ to $S_{2n-1} = -1$, in a one-to-one correspondence.

Hint: Draw a sample path joining $S_1 = 1$ to $S_{2n-1} = 1$, and reflect it in such a way that the reflected path then joins $S_1 = 1$ to $S_{2n-1} = -1$.

d) Compute the total number of paths joining $S_1 = 1$ to $S_{2n-1} = 1$ by *crossing or hitting* ⓪.

Hint: Combine the answers to part (b) and part (c).

e) Compute the total number of paths joining $S_1 = 1$ to $S_{2n-1} = 1$ *without* crossing or hitting ⓪.

Hint: Combine the answers to part (a) and (d).

f) Give the total number of paths joining $S_0 = 0$ to $S_{2n} = 0$ *without* crossing or hitting ⓪ between time 1 and time $2n - 1$.

Hint: Apply two times the answer to part (e). A drawing is recommended.

Exercise 3.10 Consider a sequence $(X_n)_{n \geqslant 0}$ of independent $\{0,1\}$-valued Bernoulli random variables with distribution $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = 0) = q$, $n \geqslant 1$.

a) Show that

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] = (q + p\mathrm{e}^t)^n, \quad n \geqslant 0, \quad t \in \mathbb{R}.$$

b) Using the Markov inequality, show that

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - p) \geqslant z\right) \leqslant \mathrm{e}^{-n((p+z)t - \log(q+p\mathrm{e}^t))}, \qquad z > 0, \ t > 0.$$

c) Find the value $t(x)$ of $t > 0$ that maximizes $t \mapsto xt - \log(q + p\mathrm{e}^t)$ for x fixed in $(0, 1)$.

d) Show the bound

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - p) \geqslant z\right) \leqslant \exp\left(-n\left((p+z)\log\frac{(p+z)q}{(q-z)p} - \log\frac{q}{q-z}\right)\right),$$

$0 \leqslant z < q$.

e) Using Taylor's formula with remainder

$$f(t) = f(0) + tf'(0) + \frac{t^2}{2} f''(\theta t), \qquad t \in \mathbb{R},$$

for some $\theta \in [0,1]$, show that $\log(q + pe^t) \leqslant pt + t^2/8$, $t \in \mathbb{R}$.

Hint. Show that for all $\alpha \in \mathbb{R}$ we have $4pq\alpha \leqslant (q + p\alpha)^2$.

f) Find the value $t(z)$ of $t \in \mathbb{R}$ that maximizes $t \mapsto zt - t^2/8$ for $z \in \mathbb{R}$.

g) Show the bound

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - p) \geqslant z\right) \leqslant \mathrm{e}^{-2nz^2}, \qquad z \geqslant 0. \tag{3.4.30}$$

Problem 3.11 Consider a sequence $(X_n)_{n\geqslant 1}$ of independent random variables on $\{1,\ldots,d\}$ with same distribution $\pi = (\pi_1,\ldots,\pi_d)$. In what follows,

$$f : \{1,\ldots,d\} \to \mathbb{R}$$

denotes any function such that $\|f\|_\infty \leqslant 1$ and $\mathbb{E}[f(X_n)] = 0$, $n \geqslant 1$, and we let

$$\lambda_0(\alpha) := \sum_{l=1}^{d} \pi_l e^{\alpha f(l)}, \qquad \alpha \geqslant 0.$$

a) Show that for any $\alpha \in \mathbb{R}$ we have

$$\mathbb{E}\left[\exp\left(\alpha \sum_{l=1}^{n} f(X_l)\right)\right] = (\lambda_0(\alpha))^n, \qquad n \geqslant 0.$$

b) Show that for any $\alpha \in \mathbb{R}$ and $\gamma > 0$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^{n} f(X_l) \geqslant \gamma\right) \leqslant e^{-n(\alpha\gamma - \log \lambda_0(\alpha))}, \qquad n \geqslant 1.$$

Hint. Use the Chernoff argument.

c) Show that

$$\lambda_0(\alpha) = 1 + \sum_{l=1}^{d} \pi_l (e^{\alpha f(l)} - \alpha f(l) - 1), \qquad \alpha \geqslant 0.$$

d) Show that

$$\lambda_0(\alpha) \leqslant 1 + \frac{\alpha^2}{1-\alpha}, \qquad \alpha \in [0,1).$$

e) Show that for any $\alpha \in [0,1)$ and $\gamma > 0$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^{n} f(X_l) \geqslant \gamma\right) \leqslant e^{-n(\alpha\gamma - \frac{\alpha^2}{1-\alpha})}, \qquad n \geqslant 1.$$

f) Find the value of $\alpha \in [0,1)$ which maximizes $\alpha\gamma - \alpha^2/(1-\alpha)$.
g) Show that for all $\gamma > 0$ and $n \geqslant 1$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^{n} f(X_l) \geqslant \gamma\right) \leqslant e^{-n\gamma^2/6}.$$

Problem 3.12 Consider a sequence $(X_n)_{n\geqslant 0}$ of independent identically distributed random variables with distribution $\pi = (\pi_1, \ldots, \pi_d)$ on $\{1, \ldots, d\}$. Our goal is to estimate the distribution π using the estimator $\widehat{\pi}_j(n) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k = j\}}$, $j = 1, \ldots, d$.

a) Show that $\mathbb{E}\left[\sum_{j=1}^{d}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k=j\}} - \pi_j\right|\right] \leqslant \sqrt{\frac{d}{n}}$, $i = 1, \ldots, d$.
b) Show that for any $n \geqslant 1$, the function $f_n : \mathbb{R}^n \to \mathbb{R}$ defined by

$$(x_1, \ldots, x_n) \mapsto \sum_{j=1}^{d}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{x_k=j\}} - \pi_j\right|$$

satisfies the bounded differences property with $c_i = 2/n$, $i = 1, \ldots, n$, *i.e.*

$$\operatorname*{Sup}_{y\in\mathbb{R}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)| \leqslant c_i, \quad x_1, \ldots, x_n \in \mathbb{R}.$$

c) Based on the results of Questions (a)-(b) and McDiarmid's inequality

$$\mathbb{P}(f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \geqslant \varepsilon) \leqslant \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n} c_i^2}\right),$$

show that for all $i = 1, \ldots, d$ we have

$$\mathbb{P}\left(\sum_{j=1}^{d}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k=j\}} - \pi_j\right| > \varepsilon\right) \leqslant \exp\left(-\frac{n}{2}\operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}}\right)^2\right).$$

d) Show that if $n \geqslant 4d/\varepsilon^2$, then we have $\mathbb{P}\left(\sum_{j=1}^{d}\left|\widehat{\pi}_j(n) - \pi_j\right| > \varepsilon\right) \leqslant e^{-n\varepsilon^2/8}$.
e) Show that there is a constant $c > 0$ such that for any $\varepsilon, \delta \in (0,1)$ we have

$$\mathbb{P}\left(\operatorname*{Max}_{j=1,\ldots,d}\left|\widehat{\pi}_j(n)-\pi_j\right|\leqslant\varepsilon\right)\geqslant 1-\delta,$$

for any $n > c\log(1/\delta)/\varepsilon^2$.

Problem 3.13 We consider an N-arm bandit in which the reward of arm $n^\circ i$ at time $n \geqslant 1$ is $X_n^{(i)}$, where for $i = 1,\ldots,N$, $(X_n^{(i)})_{n\geqslant 0}$ is a i.i.d. Bernoulli sequence with $\mathbb{P}(X_n^{(i)} = 1) = p_i \in [0,1]$, $n \geqslant 1$, ordered as $p_1 \leqslant \cdots \leqslant p_N$. We let

$$\widehat{m}_n^{(i,\alpha)} := \frac{1}{T_n^{(i,\alpha)}}\sum_{k=1}^n X_k^{(i)}\mathbb{1}_{\{\alpha_k=i\}}$$

denote the sample average reward obtained from arm $n^\circ i$ until time $n \geqslant 1$ under a given policy $(\alpha_k)_{k\geqslant 1}$. We define the policy $(\alpha_n^*)_{n\geqslant 1}$ by $\alpha_n^* := n$ for $n = 1,\ldots,N$, and for $n > N$ we let α_n^* be the index $i \in \{1,\ldots,N\}$ that maximizes the quantity $\widehat{m}_{n-1}^{(i,\alpha^*)} + \sqrt{2(\log n)/T_{n-1}^{(i,\alpha^*)}}$.

a) Let $1 \leqslant i < N$ and $n \geqslant N$. Show by contradiction that if $\alpha_n^* = i$, then at least one of the three following conditions must hold:

$$\widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log n}{T_{n-1}^{(N,\alpha^*)}}} \leqslant p_N,\ \ \widehat{m}_{n-1}^{(i,\alpha^*)} > p_i + \sqrt{\frac{2\log n}{T_{n-1}^{(i,\alpha^*)}}},\ \ T_{n-1}^{(i,\alpha^*)} < \frac{8\log n}{(p_N-p_i)^2}.$$

b) Show that letting $\widehat{n}_i := \lceil 8(\log n)/(p_N-p_i)^2\rceil$, we have

$$\mathbb{E}\big[T_n^{(i,\alpha^*)}\big] \leqslant \widehat{n}_i + \sum_{\widehat{n}_i<k\leqslant n}\left(\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant p_N\right)\right.$$
$$\left. + \mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > p_i + \sqrt{\frac{2\log k}{T_{k-1}^{(i,\alpha^*)}}}\right)\right),\quad 1 \leqslant 1 < N,\quad n \geqslant N.$$

c) Show that $\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant p_N\right) \leqslant \frac{1}{k^3}$ and

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > p_i + \sqrt{\frac{2\log k}{T_{k-1}^{(i,\alpha^*)}}}\right) \leqslant \frac{1}{k^3},\quad i = 1,\ldots,N,\quad k \geqslant N.$$

Hint. Use the bound (3.4.30) in Question 2 above.

d) Show that the *modified* regret $\overline{\mathcal{R}}n_n^\alpha := \sum_{k=1}^n \mathbb{E}\big[p_N - p_{\alpha_k}\big]$ can be bounded by

$$\overline{\mathcal{R}}n_n^{\alpha^*} \leqslant \sum_{i=1}^{N-1} (p_N - p_i) + 8 \sum_{i=1}^{N-1} \frac{\log n}{p_N - p_i}, \qquad n \geqslant 1.$$

Hint. Use a comparison argument between series and integrals.

Problem 3.14

a) Consider a gambling process $(S_n)_{n\geqslant 0}$ taking values in the discrete interval $\{0, 1, \ldots, L\}$ with respective probabilities p, q of increment and decrement. We let $T_{0,L}$ denote the hitting time of the boundary $\{0, L\}$ by $(S_n)_{n\geqslant 0}$.

   i) Compute the probability generating function

$$G_i(s) := \mathbb{E}[s^{T_{0,L}} \mid S_0 = i], \qquad i = 0, 1, \ldots, L, \quad s \in [-1, 1],$$

   of $T_{0,L}$. Consider the cases $p = q$ and $p \neq q$ separately.

   Hint. See Exercise 3.4.

   ii) Compute the Laplace transform

$$L_i(\lambda) := \mathbb{E}[\mathrm{e}^{-\lambda T_{0,L}} \mid S_0 = i], \qquad i = 0, 1, \ldots, L, \quad \lambda \geqslant 0.$$

   of $T_{0,L}$. Consider the cases $p = q$ and $p \neq q$ separately.

b) We rescale the process $(S_n)_{n\geqslant 1}$ into a continuous-time random walk $(X_t)_{t\in\mathbb{R}_+}$. For this,

- we split the time interval $[0, t]$ into $n \simeq t/\varepsilon$ time steps of length $\varepsilon > 0$,
- we split the space interval $[0, y]$ into $L \simeq y/\sqrt{\varepsilon}$ steps of height $\sqrt{\varepsilon}$,
- we rescale the probabilities p and q as

$$p_\varepsilon := \frac{1}{2}(1 - \mu\sqrt{\varepsilon}) \quad \text{and} \quad q_\varepsilon := \frac{1}{2}(1 + \mu\sqrt{\varepsilon}),$$

for some $\mu \in \mathbb{R}$, see Equation (7.6) in Privault (2023), and we let ε tend to zero. We let $T_{0,y}$ denote the hitting time of the boundary $\{0, y\}$ by $(X_t)_{t\in\mathbb{R}_+}$.

   i) Taking $\mu = 0$, compute the Laplace transform

$$L_x(\lambda) := \mathbb{E}[\mathrm{e}^{-\lambda T_{0,y}} \mid X_0 = x], \qquad x \in [0, y], \quad \lambda \geqslant 0,$$

   of $T_{0,y}$.

   Hint. Your answer should recover Equation (3) in Antal and Redner (2005), see also Equation (2.2.10) in Redner (2001) and Exercise 14.3-a) in Privault (2023).

   ii) Compute the Laplace transform

$$L_x(\lambda) := \mathbb{E}[\mathrm{e}^{-\lambda T_{0,y}} \mid X_0 = x], \qquad x \in [0, y], \quad \lambda \geqslant 0,$$

of $T_{0,y}$ in case $\mu \neq 0$.

Hint. See also Exercise 14.5 in Privault (2023).

c) Repeat Questions a and b above *for the hitting time T_L of the level L* when $(S_n)_{n\geqslant 0}$ is the random walk on $\{0,\ldots,L\}$ reflected at state 0.

Hint. Your answer should recover Equation (5) in Antal and Redner (2005) when $\mu = 0$, see also Equation (2.2.21) in Redner (2001).

Problem 3.15 Consider a random walk $(S_n)_{n\geqslant 0}$ on $\mathbb{Z}$ with independent increments, such that

$$\mathbb{P}(S_{n+1} - S_n = +1) = p \quad \text{and} \quad \mathbb{P}(S_{n+1} - S_n = -1) = q, \qquad n \geqslant 0,$$

with $p + q = 1$. The sequence $(T_0^k)_{k\geqslant 1}$ of return times to 0 of $(S_n)_{n\geqslant 0}$ is defined recursively with

$$T_0^1 := \inf\{n \geqslant 1 \ : \ S_n = 0\}.$$

and

$$T_0^{k+1} := \inf\{n > T_0^k \ : \ S_n = 0\}, \qquad k \geqslant 1.$$

a) Consider the generating function $H_i(s)$ defined as

$$H_i(s) := \mathbb{E}\left[\sum_{k\geqslant 1} s^{T_0^k} \,\middle|\, S_0 = i\right], \qquad i \in \mathbb{Z}, \quad -1 \leqslant s \leqslant 1.$$

Using first step analysis, find the recurrence relations satisfied by $H_i(s)$ for $i \geqslant 2$ and $i \leqslant -2$, and for $i = -1$, $i = 0$, $i = 1$.

b) Find $H_i(s)$ for $i \geqslant 1$, $i = 0$, and $i \leqslant -1$.

Hint. Look for a solution of the form

$$H_i(s) = C(s)\alpha^i(s) \text{ for } i \geqslant 1 \text{ and } i \leqslant -1.$$

c) Consider the probability generating function $G_i(s)$ of the first return time to (0), defined as

$$G_i(s) := \mathbb{E}\big[s^{T_0^1} \mid S_0 = i\big], \qquad i \in \mathbb{Z}, \quad -1 \leqslant s \leqslant 1.$$

Using conditioning based on T_0^1, find a relation between $G_i(s)$, $H_i(s)$ and $H_0(s)$ for $i \geqslant 2$ and $i \leqslant -2$, and for $i = -1$, $i = 0$, $i = 1$.

d) Find $G_i(s)$ for $i \geqslant 1$, $i = 0$, and $i \leqslant -1$.

e) Find the probability $\mathbb{P}(T_0^1 < \infty \mid S_0 = i)$ of hitting state (0) in finite time after starting from state (i).

f) Find the mean number of visits $\mathbb{E}[R_0 \mid S_0 = i]$ to state (0) after starting from state (i), $i \in \mathbb{Z}$.

Problem 3.16 Time spent above zero by a random walk. Consider the *symmetric* random walk $(S_n)_{n\geqslant 0}$ started at $S_0 = 0$ on $\mathbb{S} = \mathbb{Z}$. We let

$$T_{2n}^+ := 2\sum_{r=1}^{n} \mathbb{1}_{\{S_{2r-1}\geqslant 1\}}$$

denote an even estimate of the time spent strictly above the level 0 by the random walk between time 0 and time $2n$. We also let

$$T_0 := \inf\{n \geqslant 1 \ : \ S_n = 0\}$$

denote an even estimate of the time of first return of $(S_n)_{n\geqslant 0}$ to ⓪.

a) Compute $\mathbb{P}(S_{2n} = 2k)$ for $k = 0, 1, \ldots, n$.
b) Show the *convolution equation*

$$\mathbb{P}(S_{2n} = 0) = \sum_{r=1}^{n} \mathbb{P}(T_0 = 2r)\mathbb{P}(S_{2n-2r} = 0), \qquad n \geqslant 1.$$

c) By partitioning the event $\{T_{2n}^+ = 2k\}$ according to all possible times $2r = 2, 4, \ldots, 2n$ of *first* return to state ⓪ until time $2n$, show the convolution equation

$$\begin{aligned}
\mathbb{P}(T_{2n}^+ = 2k) &= \sum_{r=1}^{n} \mathbb{P}\big(T_0 = 2r, T_{2n}^+ = 2k\big) \\
&= \frac{1}{2}\sum_{r=1}^{k} \mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^+ = 2k - 2r\big) \\
&\quad + \frac{1}{2}\sum_{r=1}^{n-k} \mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^+ = 2k\big), \qquad n \geqslant 1.
\end{aligned}$$

d) Show that

$$\mathbb{P}(T_{2n}^+ = 2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0), \qquad 0 \leqslant k \leqslant n,$$

solves the convolution equation of Question (c).
e) Using the Stirling approximation $n! \simeq (n/e)^n\sqrt{2\pi n}$ as n tends to ∞, compute the limit

$$\lim_{n\to\infty} \mathbb{P}\big(T_{2n}^+/(2n) \leqslant x\big) = \lim_{n\to\infty} \sum_{0\leqslant k\leqslant nx} \mathbb{P}\big(T_{2n}^+/(2n) = k/n\big),$$

and find the limiting distribution of $T_{2n}^+/(2n)$ as n tends to infinity.

Problem 3.17 Range process. Consider the random walk $(S_n)_{n\geqslant 0}$ defined by $S_0 = 0$ and

$$S_n := X_1 + \cdots + X_n, \qquad n \geqslant 1,$$

where $(X_k)_{k\geqslant 1}$ is an *i.i.d.** family of $\{-1,+1\}$-valued random variables with distribution

$$\begin{cases} \mathbb{P}(X_k = +1) = p, \\ \\ \mathbb{P}(X_k = -1) = q, \end{cases}$$

$k \geqslant 1$, where $p + q = 1$. We let R_n denote the *range* of $(S_0, S_1, \ldots, S_n)$, *i.e.* the (random) number of distinct values appearing in the sequence $(S_0, S_1, \ldots, S_n)$.

a) Explain why

$$R_n = 1 + \left(\operatorname*{Sup}_{k=0,1,\ldots,n} S_k \right) - \left(\inf_{k=0,1,\ldots,n} S_k \right),$$

and give the value of R_0 and R_1.

b) Show that for all $k \geqslant 1$, $R_k - R_{k-1}$ is a Bernoulli random variable, and that

$$\mathbb{P}(R_k - R_{k-1} = 1) = \mathbb{P}(S_k - S_0 \neq 0, S_k - S_1 \neq 0, \ldots, S_k - S_{k-1} \neq 0).$$

c) Show that for all $k \geqslant 1$ we have

$$\mathbb{P}(R_k - R_{k-1} = 1) = \mathbb{P}(X_1 \neq 0, X_1 + X_2 \neq 0, \ldots, X_1 + \cdots + X_k \neq 0).$$

d) Show why the telescoping identity $R_n = R_0 + \sum_{k=1}^{n} (R_k - R_{k-1})$ holds for all $n \in \mathbb{N}$.

e) Show that $\mathbb{P}(T_0^r = \infty) = \lim_{k\to\infty} \mathbb{P}(T_0^r > k)$.

f) From the results of Questions (c) and (d), show that

$$\mathbb{E}[R_n] = \sum_{k=0}^{n} \mathbb{P}(T_0^r > k), \qquad n \geqslant 0,$$

where $T_0^r = \inf\{n \geqslant 1 \ : \ S_n = 0\}$ is the time of first return to ⓪ of the random walk.

g) From the results of Questions (e) and (f), show that

$$\mathbb{P}(T_0^r = \infty) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[R_n].$$

h) Show that

* independent and identically distributed.

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[R_n] = 0.$$

when $p = 1/2$, and that $\mathbb{E}[R_n] \simeq_{n\to\infty} n|p - q|$, when $p \neq 1/2$.*

Hints and comments on

a) No mathematical computation is needed here, a credible explanation (in words) is sufficient. It may be of interest to also compute $\mathbb{E}[R_2]$.

b) Show first that the two events

$$\{R_k - R_{k-1} = 1\} \quad \text{and} \quad \{S_k - S_0 \neq 0, S_k - S_1 \neq 0, \ldots, S_k - S_{k-1} \neq 0\}$$

are the same.

c) Here, the events

$$\{R_k - R_{k-1} = 1\} \quad \text{and} \quad \{X_1 \neq 0, X_1 + X_2 \neq 0, \ldots, X_1 + \cdots + X_k \neq 0\}$$

are not the same, however we can show the equality between the probabilities.

d) Telescoping identities can be useful in many situations, not restricted to probability or stochastic processes.

e) Use point *2*. on page 14, after identifying what the events A_k are.

f) The basic identity $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ can be used.

g) A mathematically rigorous proof is asked here. The following definition of limit may be used:

A real number a is said to be the limit of the sequence $(x_n)_{n\geqslant 1}$, written $a = \lim_{n\to\infty} x_n$, if and only if for every real number $\varepsilon > 0$, there exists a natural number N such that for every $n > N$ we have $|x_n - a| < \varepsilon$.

Alternatively, we may use the notion of Cesàro mean and state and apply the relevant theorem.

h) Use the formula (3.4.15) giving $\mathbb{P}(T_0^r = \infty)$.

* The meaning of $f(n) \simeq_{n\to\infty} g(n)$ is $\lim_{n\to\infty} f(n)/g(n) = 1$, provided that $g(n) \neq 0$, $n \geqslant 1$.

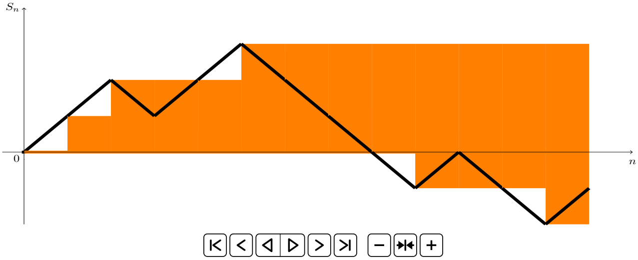


Fig. 3.5: Illustration of the range process.*

In Figure 3.5 the height at time n of the colored area coincides with $R_n - 1$.

Problem 3.18 Recurrent random walk in d dimensions. Let $\{e_1, e_2, \ldots, e_d\}$ denote the canonical basis of $\mathbb{R}^d$, *i.e.*

$$e_k = (0, \ldots, \underset{\substack{\uparrow \\ \mathbf{k}}}{1}, 0, \ldots, 0), \qquad k = 1, 2, \ldots, d.$$

Consider the symmetric $\mathbb{Z}^d$-valued random walk

$$S_n = X_1 + \cdots + X_n, \qquad n \geqslant 0,$$

started at $S_0 = \vec{0} = (0, 0, \ldots, 0)$, where $(X_n)_{n\geqslant 1}$ is a sequence of (mutually) independent uniformly distributed random variables

$$X_n \in \{e_1, e_2, \ldots, e_d, -e_1, -e_2, \ldots, -e_d\}, \qquad n \geqslant 1,$$

with distribution

$$\mathbb{P}(X_n = e_k) = \mathbb{P}(X_n = -e_k) = \frac{1}{2d}, \qquad k = 1, 2, \ldots, d.$$

Let

$$T^r_{\vec{0}} := \inf\{n \geqslant 1 \ : \ S_n = \vec{0}\}$$

denote the first return time of $(S_n)_{n\geqslant 0}$ to $\vec{0} = (0, 0, \ldots, 0)$. The random walk is said to be *recurrent* if $\mathbb{P}(T^r_{\vec{0}} < \infty) = 1$.

a) Show that the probability distribution $\mathbb{P}(T^r_{\vec{0}} = n)$, $n \geqslant 1$, satisfies the equation

* Animated figure (works in Acrobat Reader).

$$\mathbb{P}\big(S_n = \vec{0}\big) = \sum_{k=2}^{n} \mathbb{P}\big(T^r_{\vec{0}} = k\big)\mathbb{P}\big(S_{n-k} = \vec{0}\big), \qquad n \geqslant 1.$$

b) Show that

$$\sum_{n=2}^{m} \mathbb{P}\big(T^r_{\vec{0}} = n\big) \geqslant 1 - \frac{1}{\displaystyle\sum_{n=0}^{m} \mathbb{P}\big(S_n = \vec{0}\big)}, \qquad m \geqslant 1. \tag{3.4.31}$$

Hint: Start by showing that

$$\sum_{n=1}^{m} \mathbb{P}\big(S_n = \vec{0}\big) = \sum_{k=2}^{m} \mathbb{P}\big(T^r_{\vec{0}} = k\big) \sum_{l=0}^{m-k} \mathbb{P}\big(S_l = \vec{0}\big).$$

c) Show that under the condition

$$\sum_{n\geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big) = \infty,$$

we have $\mathbb{P}\big(T^r_{\vec{0}} < \infty\big) = 1$.

d) Show that

$$\sum_{n=2}^{m} \mathbb{P}\big(T^r_{\vec{0}} = n\big) \leqslant \frac{\displaystyle\sum_{n=2}^{2m} \mathbb{P}\big(S_n = \vec{0}\big)}{\displaystyle\sum_{n=0}^{m} \mathbb{P}\big(S_n = \vec{0}\big)}, \qquad m \geqslant 1. \tag{3.4.32}$$

Hint: Start by showing that

$$\sum_{n=1}^{2m} \mathbb{P}\big(S_n = \vec{0}\big) = \sum_{n=2}^{2m} \mathbb{P}\big(T^r_{\vec{0}} = n\big) \sum_{l=0}^{2m-n} \mathbb{P}\big(S_l = \vec{0}\big).$$

e) Show that under the condition

$$\sum_{n\geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big) < \infty,$$

we have $\mathbb{P}\big(T^r_{\vec{0}} = \infty\big) > 0$.

f) When $\underline{d = 1}$, compute $\mathbb{P}(S_{2n} = 0)$, $n \geqslant 1$, and show that the one-dimensional random walk is *recurrent*, *i.e.* we have $\mathbb{P}(T^r_0 < \infty) = 1$.

Hint: Use Stirling's approximation $n! \simeq (n/e)^n\sqrt{2\pi n}$ as n tends to ∞.

g) When $\underline{d = 2}$, show that we have

$$\mathbb{P}\big(S_{2n} = \vec{0}\big) = \frac{(2n)!}{4^{2n}(n!)^2} \sum_{k=0}^{n} \binom{n}{k}^2, \qquad n \geqslant 1, \tag{3.4.33}$$

and that the two-dimensional random walk is *recurrent*, *i.e.* we have $\mathbb{P}\big(T_{\vec{0}}^{r} < \infty\big) = 1$.

Hint: Use the combinatorial identity

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$$

and Stirling's approximation.

h) Given $i_1, i_2, \ldots, i_d \in \mathbb{N}$, count all paths starting from $\vec{0}$ and returning to $\vec{0}$ via i_k "forward" steps in the direction e_k and i_k "backward" steps in the direction $-e_k$, $k = 1, 2, \ldots, d$. Deduce an expression for $\mathbb{P}\big(S_{2n} = \vec{0}\big)$ that generalizes (3.4.33) to all $d \geqslant 2$.

Hint: Use multinomial coefficients.

i) Based on the Euclidean division $n = a_n d + b_n$ where $b_n \in \{0, 1, \ldots, d-1\}$, show that we have

$$\sum_{n \geqslant 1} \mathbb{P}\big(S_{2n} = \vec{0}\big) \leqslant \sum_{n \geqslant 1} \frac{(2n)!}{2^{2n} d^n n! (a_n!)^d a_n^{b_n}}$$

Hint: Use the bound

$$i_1! i_2! \cdots i_d! \geqslant (a_n!)^d (a_n + 1)^{b_n}$$

which is valid for any $n = i_1 + \cdots + i_d$, and the identity

$$d^n = \sum_{\substack{i_1 + \cdots + i_d = n \\ i_1, i_2, \ldots, i_d \geqslant 0}} \frac{n!}{i_1! \cdots i_d!}.$$

j) Applying Stirling's approximation to $n!$, $(2n)!$ and $a_n!$, show that there exists a constant $C > 0$ such that for all n sufficiently large we have

$$\frac{(2n)!}{2^{2n} d^n n! (a_n!)^d} \leqslant \frac{C}{n^{d/2}}.$$

Hint: We have $\lim_{m \to \infty} (1 + x/m)^m = \mathrm{e}^x$ for all $x \in \mathbb{R}$.

k) Is the random walk recurrent when $d \geqslant 3$?

Problem 3.19 (Benjamini and Wilson (2003), Antal and Redner (2005)). Random walks in a cookie environment, also called excited random walks, can be used to model the behavior of primitive organisms.

In the absence of cookies the random walk is symmetric, with probabilities 1/2 of going up and down, and it can *rebound* to ⓵ with probability 1/2 after hitting state ⓪.

When the random walk encounters a cookie, its behavior becomes modified and it restarts with probabilities p and $q = 1 - p$ of moving up, resp. down, where $p \in [0, 1]$. Every encountered cookie is eaten by the organism, and when the random walk reaches an empty spot it restarts with equal probabilities 1/2 of moving up or down. In this case the organism wanders without a preferred direction. The random walk is *attracted* by the cookies when $p > 1/2$, and *repulsed* when $p < 1/2$.

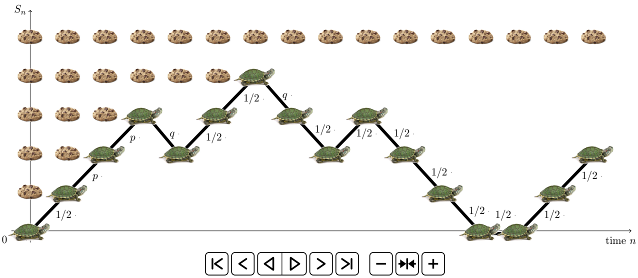


Fig. 3.6: Random walk with cookies.*

a) Does the cookie random walk $(S_n)_{n \geqslant 0}$ have the Markov property? Explain your answer.
b) Suppose that at time $n \geqslant 1$ the random walk has just eaten a cookie at state $x \geqslant 1$, after eating all cookies at states $1, 2, \ldots, x - 1$. Show that in this case, the probability of reaching $\boxed{x+1}$ before reaching ⓪ is given by

$$p + (1 - p)\frac{x - 1}{x + 1} = 1 - \frac{2q}{x + 1}, \qquad x \geqslant 1.$$

Hint: Use first step analysis together with formula (2.2.13) page 60, see also page 65.
c) For any $x \geqslant 1$, let τ_x denote the first hitting time

$$\tau_x := \inf\{n \geqslant 1 \ : \ S_n = x\}.$$

Give the value of $\mathbb{P}(\tau_1 < \tau_0 \mid S_0 = 0)$, and show that for all $x \geqslant 1$ we have[*]

$$\mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) = \frac{\mathbf{1}}{\mathbf{2}} \exp\left(\sum_{l=2}^{x} \log\left(1 - \frac{2q}{l}\right)\right), \qquad x \geqslant 1,$$

where "log" denotes the *natural logarithm* "ln".

d) Show the bound $\mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) \leqslant (x/2)^{-2q}$ for $x \geqslant 2$.

Hint: Use comparison arguments between integrals and series.

e) Show that $\mathbb{P}(\tau_0 < \infty \mid S_0 = 0) = 1$, provided that $p < 1$.[†]

Hint: Consider $\mathbb{P}(\tau_0 \leqslant \tau_x \mid S_0 = 0)$ and use Relation (1.2.5) page 14.

f) Starting from state $\textcircled{0}$, compute the mean time needed by the random walk to reach state $\textcircled{1}$.

Hint: Apply first step analysis, or locate and use the relevant result in Problem 2.7.

g) Suppose that a cookie has just been eaten at state $x \geqslant 1$, after eating all cookies at states $1, 2, \ldots, x-1$. Show that the mean time to reach the next cookie at state $\boxed{x+1}$ is $1 + q(4x+2)$.

Hint: Locate and apply the relevant results in Problem 2.7 together with first step analysis.

h) Show that the mean time to reach state $\textcircled{x}$ starting from $\textcircled{0}$ is given by

$$\mathbb{E}[\tau_x \mid S_0 = 0] = 1 - 2q + x + 2qx^2, \qquad x \geqslant 1.$$

Hint: Perform a summation on the results of Questions (f) and (g).

i) Suppose that a cookie has just been eaten at state $x \geqslant 1$, after eating all cookies at states $1, 2, \ldots, x-1$. Show that, *given one does not hit* $\textcircled{0}$, the probabilities of moving up to $\boxed{x+1}$, resp. down to $\boxed{x-1}$, are

$$\frac{p}{1 - 2q/(x+1)} \quad \text{and} \quad \frac{q(x-1)/(x+1)}{1 - 2q/(x+1)}, \qquad x \geqslant 1.$$

Hint: Proceed similarly to Question (g) of Problem 2.7.

j) Suppose that a cookie has just been eaten at state $x \geqslant 1$, after eating all cookies at states $1, 2, \ldots, x-1$. Show that the mean time to reach the next cookie at state $\boxed{x+1}$ *given one does not hit* $\textcircled{0}$ is

$$1 + \frac{q - 2q/(x+1)}{1 - 2q/(x+1)} \frac{4x}{3}, \qquad x \geqslant 1.$$

---

[*] We use the convention $\sum_{k=2}^{1} a_k = 0$ for any sequence (a_k).

[†] In this case the cookie random walk is said to be *recurrent*.

Hint: Locate and use the relevant result in Problem 2.7.

Problem 3.20 (Antal and Redner (2005)). A random walk $(S_n)_{n\geqslant 0}$ with cookies on $\{1,2,3,\ldots\}$ is symmetric in the absence of cookies, and restarts with probabilities p and $q=1-p$ of moving up, resp. down, when it encounters a cookie, where $p\in[0,1)$. The random walk starts at state $⓪$, which is empty of cookie.

For any $x\geqslant 1$, let τ_x denote the first hitting time

$$\tau_x := \inf\{n\geqslant 1 \ : \ S_n = x\}, \qquad x\geqslant 1.$$

Recall that the probability of eating at least x cookies before returning to the origin $⓪$ is given by

$$\mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) = \frac{1}{2}\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right), \quad x\geqslant 1, \tag{3.4.34}$$

and that the random walk is recurrent, *i.e.* it returns to the origin $⓪$ in finite time whenever $p<1$, that means we have $\mathbb{P}(\tau_0<\infty \mid S_0=0)=1$.

a) Let X denote the number of cookies eaten by the random walk before returning to the origin $⓪$. Show that

$$\mathbb{P}(X=0)=1/2, \qquad \mathbb{P}(X=1)=q/2,$$

and, using (3.4.34), that the distribution of satisfies

$$\mathbb{P}(X=x) = \frac{q}{x+1}\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right), \qquad x\geqslant 2. \tag{3.4.35}$$

b) Show from (3.4.35) that the average number $\mathbb{E}[X]$ of cookies eaten before returning to the origin $⓪$ is finite, *i.e.* $\mathbb{E}[X]<\infty$, if and only if $q>1/2$.

Hint: There exists constants $c_q, C_q>0$ such that

$$\frac{c_q}{x^{2q}} \leqslant \prod_{l=2}^{x}\left(1-\frac{2q}{l}\right) \leqslant \frac{C_q}{x^{2q}}, \qquad x\geqslant 2.$$

Chapter 4
Discrete-Time Markov Chains

In this chapter we start the general study of discrete-time Markov chains by focusing on the Markov property and on the role played by transition probability matrices. We also include a complete study of the time evolution of the two-state chain, which represents the simplest example of Markov chain.

4.1 Markov (1856-1922) Property

We consider a discrete-time stochastic process $(Z_n)_{n\geqslant 0}$ taking values in a discrete state space $\mathbb{S}$, typically $\mathbb{S} = \mathbb{Z}$. The $\mathbb{S}$-valued process $(Z_n)_{n\geqslant 0}$ is said to be *Markov (1909)*, or to have the *Markov property* if, for all $n \geqslant 1$, the probability distribution of Z_{n+1} is determined by the state Z_n of the process at time n, and does not depend on the past values of Z_k for $k = 0, 1, \ldots, n-1$.

Fig. 4.1: NGram Viewer output for the term "Markov chains".

In other words, for all $n \geqslant 1$ and all $i_0, i_1, \ldots, i_n, j \in \mathbb{S}$ we have

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i_n,\ Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) = \mathbb{P}(Z_{n+1} = j \mid Z_n = i_n).$$

In particular, we have

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i_n,\ Z_{n-1} = i_{n-1}) = \mathbb{P}(Z_{n+1} = j \mid Z_n = i_n),$$

and

$$\mathbb{P}(Z_2 = j \mid Z_1 = i_1, Z_0 = i_0) = \mathbb{P}(Z_2 = j \mid Z_1 = i_1).$$

Note that this feature is apparent in the statement of Lemma 2.2. In addition, we have the following facts.

1. *Chain rule.* The first order transition probabilities can be used for the complete computation of the probability distribution of the process as

$$\begin{aligned}
&\mathbb{P}(Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) \\
&\quad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0)\mathbb{P}(Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) \\
&\quad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1})\mathbb{P}(Z_{n-1} = i_{n-1} \mid Z_{n-2} = i_{n-2}, \ldots, Z_0 = i_0) \\
&\qquad \times \mathbb{P}(Z_{n-2} = i_{n-2}, \ldots, Z_0 = i_0) \\
&\quad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1})\mathbb{P}(Z_{n-1} = i_{n-1} \mid Z_{n-2} = i_{n-2}) \\
&\qquad \times \mathbb{P}(Z_{n-2} = i_{n-2} \mid Z_{n-3} = i_{n-3}, \ldots, Z_0 = i_0)\mathbb{P}(Z_{n-3} = i_{n-3}, \ldots, Z_0 = i_0) \\
&\quad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1})\mathbb{P}(Z_{n-1} = i_{n-1} \mid Z_{n-2} = i_{n-2}) \\
&\qquad \times \mathbb{P}(Z_{n-2} = i_{n-2} \mid Z_{n-3} = i_{n-3})\mathbb{P}(Z_{n-3} = i_{n-3}, \ldots, Z_0 = i_0),
\end{aligned}$$

which shows, reasoning by induction, that

$$\begin{aligned}
&\mathbb{P}(Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) \qquad (4.1.1) \\
&\qquad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1}) \cdots \mathbb{P}(Z_1 = i_1 \mid Z_0 = i_0)\mathbb{P}(Z_0 = i_0),
\end{aligned}$$

or

$$\begin{aligned}
&\mathbb{P}(Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_1 = i_1 \mid Z_0 = i_0) \qquad (4.1.2)\\
&\qquad = \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1}) \cdots \mathbb{P}(Z_1 = i_1 \mid Z_0 = i_0),
\end{aligned}$$

$i_0, i_1, \ldots, i_n \in \mathbb{S}$.

2. By the *law of total probability* (1.3.1) applied under $\mathbb{P}$ to the events $A_{i_0} = \{Z_1 = i_1 \text{ and } Z_0 = i_0\}$, $i_0 \in \mathbb{S}$, we also have

$$\begin{aligned}
\mathbb{P}(Z_1 = i_1) &= \sum_{i_0 \in \mathbb{S}} \mathbb{P}(Z_1 = i_1, Z_0 = i_0)\\
&= \sum_{i_0 \in \mathbb{S}} \mathbb{P}(Z_1 = i_1 \mid Z_0 = i_0)\mathbb{P}(Z_0 = i_0), \quad i_1 \in \mathbb{S}, \quad (4.1.3)
\end{aligned}$$

and similarly, under the probability measure $\mathbb{P}(\cdot \mid Z_0 = i_0)$,

$$\begin{aligned}
\mathbb{P}(Z_2 = i_2 \mid Z_0 = i_0) &= \sum_{i_1 \in \mathbb{S}} \mathbb{P}(Z_2 = i_2 \text{ and } Z_1 = i_1 \mid Z_0 = i_0)\\
&= \sum_{i_1 \in \mathbb{S}} \mathbb{P}(Z_2 = i_2 \mid Z_1 = i_1)\mathbb{P}(Z_1 = i_1 \mid Z_0 = i_0), \quad (4.1.4)
\end{aligned}$$

$i_0, i_2 \in \mathbb{S}$.

Example

The random walk

$$S_n := X_1 + X_2 + \cdots + X_n, \qquad n \geqslant 0, \qquad (4.1.5)$$

considered in Chapter 3, where $(X_n)_{n\geqslant 1}$ is a sequence of (mutually) independent $\mathbb{Z}$-valued random increments, is a discrete-time Markov chain with $\mathbb{S} = \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

Indeed, the value of S_{n+1} depends only on S_n and on the value of the next increment X_{n+1}. In other words, for all $j, i_n, \ldots, i_1 \in \mathbb{Z}$ we have (note that $S_0 = 0$ here)

$$\begin{aligned}
&\mathbb{P}(S_{n+1} = j \mid S_n = i_n,\ S_{n-1} = i_{n-1}, \ldots, S_1 = i_1) \qquad (4.1.6)\\
&= \frac{\mathbb{P}(S_{n+1} = j, S_n = i_n, S_{n-1} = i_{n-1}, \ldots, S_1 = i_1)}{\mathbb{P}(S_n = i_n,\ S_{n-1} = i_{n-1}, \ldots, S_1 = i_1)}\\
&= \frac{\mathbb{P}(S_{n+1} - S_n = j - i_n, S_n - S_{n-1} = i_n - i_{n-1}, \ldots, S_2 - S_1 = i_2 - i_1, S_1 = i_1)}{\mathbb{P}(S_n - S_{n-1} = i_n - i_{n-1}, \ldots, S_2 - S_1 = i_2 - i_1, S_1 = i_1)}\\
&= \frac{\mathbb{P}(X_{n+1} = j - i_n, X_n = i_n - i_{n-1}, \ldots, X_2 = i_2 - i_1, X_1 = i_1)}{\mathbb{P}(X_n = i_n - i_{n-1}, \ldots, X_2 = i_2 - i_1, X_1 = i_1)}\\
&= \frac{\mathbb{P}(X_{n+1} = j - i_n)\mathbb{P}(X_n = i_n - i_{n-1}, \ldots, X_2 = i_2 - i_1, X_1 = i_1)}{\mathbb{P}(X_n = i_n - i_{n-1}, \ldots, X_2 = i_2 - i_1, X_1 = i_1)}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(X_{n+1} = j - i_n) \\
&= \frac{\mathbb{P}(X_{n+1} = j - i_n)\mathbb{P}(X_n + \cdots + X_1 = i_n)}{\mathbb{P}(X_1 + \cdots + X_n = i_n)} \\
&= \frac{\mathbb{P}(X_{n+1} = j - i_n, X_n + \cdots + X_1 = i_n)}{\mathbb{P}(X_1 + \cdots + X_n = i_n)} \\
&= \frac{\mathbb{P}(X_{n+1} = j - i_n \text{ and } S_n = i_n)}{\mathbb{P}(S_n = i_n)} = \frac{\mathbb{P}(S_{n+1} = j \text{ and } S_n = i_n)}{\mathbb{P}(S_n = i_n)} \\
&= \mathbb{P}(S_{n+1} = j \mid S_n = i_n).
\end{aligned}
$$

In addition, the Markov chain $(S_n)_{n\geqslant 0}$ is *time homogeneous* if the random sequence $(X_n)_{n\geqslant 1}$ is identically distributed.

In particular, we have

$$
\mathbb{P}(S_{n+1} = j \mid S_n = i) = \mathbb{P}(X_{n+1} = j - i),
$$

hence the transition probability from state $\textcircled{i}$ to state $\textcircled{j}$ of a random walk with independent increments depends only on the difference $j - i$ and on the distribution of X_{n+1}.

More generally, all processes with independent increments are Markov processes. However,*not all Markov chains have independent increments*. In fact, the Markov chains of interest in this chapter do not have independent increments.

4.2 Transition matrix

In what follows we will assume that the Markov chain $(Z_n)_{n\geqslant 0}$ is *time homogeneous*, *i.e.* the probability $\mathbb{P}(Z_{n+1} = j \mid Z_n = i)$ is independent of $n \in \mathbb{N}$. In this case, the random evolution of a Markov chain $(Z_n)_{n\geqslant 0}$ is determined by the data of

$$
P_{i,j} := \mathbb{P}(Z_1 = j \mid Z_0 = i), \qquad i, j \in \mathbb{S}, \tag{4.2.1}
$$

which coincides with the probability $\mathbb{P}(Z_{n+1} = j \mid Z_n = i)$ for all $n \in \mathbb{N}$. These data can be encoded into a matrix indexed by $\mathbb{S}^2 = \mathbb{S} \times \mathbb{S}$, called the *transition matrix* of the Markov chain:

$$
[\ P_{i,j}\ ]_{i,j\in\mathbb{S}} = [\ \mathbb{P}(Z_1 = j \mid Z_0 = i)\ ]_{i,j\in\mathbb{S}},
$$

also written on $\mathbb{S} := \mathbb{Z}$ as

$$P = \left[P_{i,j} \right]_{i,j \in \mathbb{S}} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & ⋰ \\ \cdots & P_{-2,-2} & P_{-2,-1} & P_{-2,0} & P_{-2,1} & P_{-2,2} & \cdots \\ \cdots & P_{-1,-2} & P_{-1,-1} & P_{-1,0} & P_{-1,1} & P_{-1,2} & \cdots \\ \cdots & P_{0,-2} & P_{0,-1} & P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ \cdots & P_{1,-2} & P_{1,-1} & P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ \cdots & P_{2,-2} & P_{2,-1} & P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ ⋰ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The notion of transition matrix is related to that of (weighted) adjacency matrix in graph theory.

Note the inversion of the order of indices (i, j) between $\mathbb{P}(Z_{n+1} = j \mid Z_n = i)$ and $P_{i,j}$. In particular, the initial state $\textcircled{i}$ is a *row number* in the matrix, while the final state $\textcircled{j}$ corresponds to a *column number*.

By the *law of total probability* (1.3.1) applied to the probability measure $\mathbb{P}(\cdot \mid Z_0 = i)$, we also have the equality

$$\sum_{j \in \mathbb{S}} \mathbb{P}(Z_1 = j \mid Z_0 = i) = \mathbb{P}\left(\bigcup_{j \in \mathbb{S}} \{Z_1 = j\} \mid Z_0 = i \right) = \mathbb{P}(\Omega) = 1, \quad i \in \mathbb{S}, \tag{4.2.2}$$

i.e. the *rows* of the transition matrix satisfy the condition

$$\sum_{j \in \mathbb{S}} P_{i,j} = 1,$$

for every row index $i \in \mathbb{S}$.

Using the matrix notation $P = (P_{i,j})_{i,j \in \mathbb{S}}$, and Relation (4.1.1) we find

$$\mathbb{P}(Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) = P_{i_{n-1},i_n} \cdots P_{i_0,i_1} \mathbb{P}(Z_0 = i_0),$$

$i_0, i_1, \ldots, i_n \in \mathbb{S}$, and we rewrite (4.1.3) as

$$\mathbb{P}(Z_1 = i) = \sum_{j \in \mathbb{S}} \mathbb{P}(Z_1 = i \mid Z_0 = j)\mathbb{P}(Z_0 = j) = \sum_{j \in \mathbb{S}} P_{j,i} \mathbb{P}(Z_0 = j), \quad i \in \mathbb{S}. \tag{4.2.3}$$

A state $k \in \mathbb{S}$ is said to be *absorbing* if $P_{k,k} = 1$.

In what follows we will often consider $\mathbb{S} = \mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbb{N}$-valued Markov chains, in which case the *transition matrix* $\left[\mathbb{P}(Z_{n+1} = j \mid Z_n = i) \right]_{i,j \in \mathbb{N}}$ of the chain is written as

$$[\ P_{i,j}\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

From (4.2.2) we have

$$\sum_{j\geqslant 0} P_{i,j} = 1,$$

for all $i \in \mathbb{N}$.

In case the Markov chain $(Z_k)_{k\geqslant 0}$ takes values in the finite state space $\mathbb{S} = \{0, 1, \ldots, N\}$ its $(N+1)\times(N+1)$ transition matrix will simply have the form

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant N} = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix}.$$

Still on the finite state space $\mathbb{S} = \{0, 1, \ldots, N\}$, Relation (4.2.3) can be restated in the language of matrix and vector products using the shorthand notation:

$$\eta = \pi P, \tag{4.2.4}$$

where

$$\eta := [\mathbb{P}(Z_1 = 0), \ldots, \mathbb{P}(Z_1 = N)] = [\eta_0, \eta_1, \ldots, \eta_N] \in \mathbb{R}^{N+1}$$

is the row vector "distribution of Z_1",

$$\pi := [\mathbb{P}(Z_0 = 0), \ldots, \mathbb{P}(Z_0 = N)] = [\pi_0, \ldots, \pi_N] \in \mathbb{R}^{N+1}$$

is the row vector representing the probability distribution of Z_0, and

$$[\eta_0, \eta_1, \ldots, \eta_N] = [\pi_0, \ldots, \pi_N] \times \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix}. \qquad (4.2.5)$$

Invariant vectors

A row vector π such that $\pi = \pi P$ is said to be *invariant* or *stationary* by the transition matrix P.

For example, in case the matrix P takes the form

$$P = [\, P_{i,j} \,]_{0 \leqslant i,j \leqslant N} = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \end{bmatrix},$$

with all rows equal and $\pi_0 + \pi_1 + \cdots + \pi_N = 1$, then we have $\pi = \pi P$, *i.e.* π is an invariant (or stationary) distribution for P.

4.3 Examples of Markov Chains

The wide range of applications of Markov chains to engineering, physics and biology has already been mentioned in the introduction. Here we consider some more specific examples.

i) Random walk.

The transition matrix $[\, P_{i,j} \,]_{i,j \in \mathbb{S}}$ of the unrestricted random walk (4.1.5) is given by

$$
[\, P_{i,j} \,]_{i,j\in \mathbb{S}} = \begin{array}{c} \\ \\ i-2 \\ i-1 \\ i \\ i+1 \\ i+2 \\ \\ \end{array}\begin{array}{c} \begin{array}{ccc} i-1 & i & i+1 \end{array} \\ \left[\begin{array}{cccccc} \ddots \vdots & \vdots & \vdots & \vdots & \vdots \iddots \\ \cdots 0 & p & 0 & 0 & 0 \cdots \\ \cdots q & 0 & p & 0 & 0 \cdots \\ \cdots 0 & q & 0 & p & 0 \cdots \\ \cdots 0 & 0 & q & 0 & p \cdots \\ \cdots 0 & 0 & 0 & q & 0 \cdots \\ \iddots \vdots & \vdots & \vdots & \vdots & \vdots \ddots \end{array}\right] \end{array}. \tag{4.3.1}
$$

ii) Gambling process.

The transition matrix $[\, P_{i,j} \,]_{0\leqslant i,j\leqslant S}$ of the gambling process on $\{0, 1, \ldots, S\}$ with absorbing states ⓪ and Ⓢ, see Chapter 2, is given by

$$
P = [\, P_{i,j} \,]_{0\leqslant i,j\leqslant S} = \left[\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 1 \end{array}\right].
$$

iii) Credit rating.

[Transition probabilities are expressed in %].

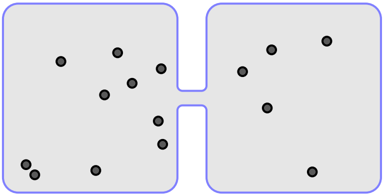
Rating at the start of a year / Rating at the end of the year

	AAA	AA	A	BBB	BB	B	CCC	D	N.R.	Total
AAA	90.34	5.62	0.39	0.08	0.03	0	0	0	3.5	100
AA	0.64	88.78	6.72	0.47	0.06	0.09	0.02	0.01	3.21	100
A	0.07	2.16	87.94	4.97	0.47	0.19	0.01	0.04	4.16	100
BBB	0.03	0.24	4.56	84.26	4.19	0.76	0.15	0.22	5.59	100
BB	0.03	0.06	0.4	6.09	76.09	6.82	0.96	0.98	8.58	100
B	0	0.09	0.29	0.41	5.11	74.62	3.43	5.3	10.76	100
CCC	0.13	0	0.26	0.77	1.66	8.93	53.19	21.94	13.14	100
D	0	0	0	0	1.0	3.1	9.29	51.29	37.32	100
N.R.	0	0	0	0	0	0.1	8.55	74.06	17.07	100

We note that higher ratings are more stable since the diagonal coefficients of the matrix go decreasing. On the other hand, starting from an AA rating it is easier to be downgraded to A (probability 6.72%) than to be upgraded to AAA (probability 0.64%).

iv) Ehrenfest chain.

Two volumes of air (left and right), containing a total of N balls, are connected by a pipe.



At each time step, one picks a ball at random and moves it to the other side. Let $Z_n \in \{0, 1, \ldots, N\}$ denote the number of balls in the left side at time n. The transition probabilities $\mathbb{P}(Z_{n+1} = j \mid Z_n = i)$, $0 \leqslant i, j \leqslant N$, are given by

$$\mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k) = \frac{N-k}{N}, \qquad k = 0, 1, \ldots, N-1, \quad (4.3.2)$$

and

$$\mathbb{P}(Z_{n+1} = k-1 \mid Z_n = k) = \frac{k}{N}, \qquad k = 1, 2, \ldots, N, \qquad (4.3.3)$$

with

$$P = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 1/N & 0 & (N-1)/N & \cdots & \cdots & 0 & 0 \\ 0 & 2/N & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 3/N & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \ddots & 3/N & 0 & 0 \\ 0 & 0 & \cdots & \ddots & 0 & 2/N & 0 \\ 0 & 0 & \cdots & \cdots & (N-1)/N & 0 & 1/N \\ 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix},$$

cf. Exercise 6.7, Exercise 7.4, Problem 7.32 on the (modified) Ehrenfest chain, and Exercise 4.9 on the Bernoulli-Laplace chain.

v) Markov chains in music.

By a statistical analysis of note transitions, every type of music can be encoded into a Markov chain. An example of such an analysis is presented in the next transition matrix.

	A	A♯	B	C	D	E	F	G	G♯
A	4/19	0	3/19	0	2/19	1/19	0	6/19	3/19
A♯	1	0	0	0	0	0	0	0	0
B	7/15	0	1/15	4/15	0	3/15	0	0	0
C	0	0	6/15	3/15	6/15	0	0	0	0
D	0	0	0	3/11	3/11	5/11	0	0	0
E	4/19	1/19	0	3/19	0	5/19	4/19	1/19	1/19
F	0	0	0	1/5	0	0	1/5	0	3/5
G	1/5	0	1/5	2/5	0	0	0	1/5	0
G♯	0	0	3/4	0	0	1/4	0	0	0

Fig. 4.2: Mozart: variations KV 265 for Piano.*

The transitions of Mozart's variations, cf. Figure 4.2 above, on this famous tune 🔊 have been analyzed to form a transition matrix.† Then that transition matrix was used for random melody 🔊 generation.Hear also this arrangement 🔊 and see *here* for recent examples.

vi) Text generation.

Markov chains can be used to generate sentences in a given language, based on a statistical analysis on the transition between words in a sample text. The state space of the Markov chain can be made of different

* Click on the figure to play the video (works in Acrobat Reader).

† Try *here* if it does not work.

word sequences as in *e.g.* Problem 4.15 or Problem 5.33. See *here* for an example of the use of Markov chain in random text generation.

Other applications of Markov chains include:

- Memory management in computer science,
- Logistics, supply chain management, and waiting queues,
- Modeling of insurance claims (Problem 4.14),
- Board games, *e.g.* Snakes and Ladders,
- Genetics, cf. the Wright-Fisher model (Problem 5.32).
- Random fields in imaging,
- Artificial intelligence, learning theory and machine learning (Problems 4.15, 5.28 and 5.33).

Graph representation

Whenever possible we will represent a Markov chain using a graph, as in the following example with transition matrix

$$P = \begin{bmatrix} 0 & 0.2 & 0.8 & 0 & 0 \\ 0.4 & 0 & 0 & 0.6 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.4 & 0.6 & 0 & 0 \end{bmatrix}, \tag{4.3.4}$$

see Figure 4.3.

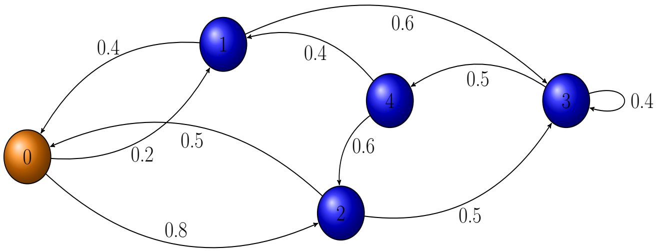


Fig. 4.3: Graph of a five-state Markov chain.*

* Animated figure (works in Acrobat Reader).

4.4 Higher-Order Transition Probabilities

As noted above, the transition matrix P is a convenient way to record $\mathbb{P}(Z_{n+1} = j \mid Z_n = i)$, $i, j \in \mathbb{S}$, into an array of data.

However, it is *much more than that*, as already hinted at in Relation (4.2.4). Suppose for example that we are interested in the two-step transition probability

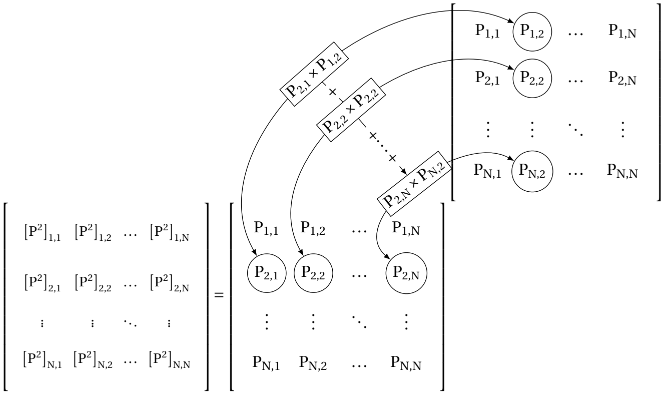
$$\mathbb{P}(Z_{n+2} = j \mid Z_n = i).$$

This probability does not appear in the transition matrix P, but it can be computed by first step analysis, applying the *law of total probability* (1.3.1) to the probability measure $\mathbb{P}(\,\cdot \mid Z_n = i)$ as follows.

$i)$ 2-step transitions. Denoting by $\mathbb{S}$ the state space of the process we have, using (4.1.4),

$$\begin{aligned}
\mathbb{P}(Z_{n+2} = j \mid Z_n = i) &= \sum_{l \in \mathbb{S}} \mathbb{P}(Z_{n+2} = j \text{ and } Z_{n+1} = l \mid Z_n = i) \\
&= \sum_{l \in \mathbb{S}} \frac{\mathbb{P}(Z_{n+2} = j,\ Z_{n+1} = l,\ Z_n = i)}{\mathbb{P}(Z_n = i)} \\
&= \sum_{l \in \mathbb{S}} \frac{\mathbb{P}(Z_{n+2} = j,\ Z_{n+1} = l,\ Z_n = i)}{\mathbb{P}(Z_{n+1} = l \text{ and } Z_n = i)} \frac{\mathbb{P}(Z_{n+1} = l \text{ and } Z_n = i)}{\mathbb{P}(Z_n = i)} \\
&= \sum_{l \in \mathbb{S}} \mathbb{P}(Z_{n+2} = j \mid Z_{n+1} = l \text{ and } Z_n = i) \mathbb{P}(Z_{n+1} = l \mid Z_n = i) \\
&= \sum_{l \in \mathbb{S}} \mathbb{P}(Z_{n+2} = j \mid Z_{n+1} = l) \mathbb{P}(Z_{n+1} = l \mid Z_n = i) \\
&= \sum_{l \in \mathbb{S}} P_{i,l} P_{l,j} \\
&= [P^2]_{i,j}, \qquad i, j \in \mathbb{S},
\end{aligned}$$

where we used (4.2.1), in agreement with the matrix multiplication mechanism described below.



Hence, using matrix product notation, we find

$$
\begin{aligned}
&(\mathbb{P}(Z_{n+2} = j \mid Z_n = i))_{0 \leqslant i,j \leqslant N} \\
&= \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix} \times \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix}.
\end{aligned}
$$

ii) k-step transitions. More generally, we have the following result.

Proposition 4.1. *For all* $k \in \mathbb{N}$ *we have the relation*

$$
\big[\ \mathbb{P}(Z_{n+k} = j \mid Z_n = i) \ \big]_{i,j \in \mathbb{S}} = \Big[\ [P^k]_{i,j} \ \Big]_{i,j \in \mathbb{S}} = P^k. \tag{4.4.1}
$$

Proof. We prove (4.4.1) by induction. Clearly, the statement holds for $k = 0$ and $k = 1$. Next, for all $k \in \mathbb{N}$, we have

$$
\begin{aligned}
\mathbb{P}(Z_{n+k+1} = j \mid Z_n = i) &= \sum_{l \in \mathbb{S}} \mathbb{P}(Z_{n+k+1} = j \text{ and } Z_{n+k} = l \mid Z_n = i) \\
&= \sum_{l \in \mathbb{S}} \frac{\mathbb{P}(Z_{n+k+1} = j, \ Z_{n+k} = l, \ Z_n = i)}{\mathbb{P}(Z_n = i)}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{l\in\mathbb{S}} \frac{\mathbb{P}(Z_{n+k+1}=j,\ Z_{n+k}=l,\ Z_n=i)}{\mathbb{P}(Z_{n+k}=l \text{ and } Z_n=i)} \frac{\mathbb{P}(Z_{n+k}=l \text{ and } Z_n=i)}{\mathbb{P}(Z_n=i)} \\
&= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_{n+k+1}=j \mid Z_{n+k}=l \text{ and } Z_n=i)\mathbb{P}(Z_{n+k}=l \mid Z_n=i) \\
&= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_{n+k+1}=j \mid Z_{n+k}=l)\mathbb{P}(Z_{n+k}=l \mid Z_n=i) \\
&= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_{n+k}=l \mid Z_n=i)P_{l,j}.
\end{aligned}
$$

We have just checked that the family of matrices

$$
[\ \mathbb{P}(Z_{n+k}=j \mid Z_n=i)\ ]_{i,j\in\mathbb{S}}, \qquad k \geqslant 1,
$$

satisfies the same induction relation as the *matrix power* P^k, *i.e.*

$$
[P^{k+1}]_{i,j} = \sum_{l\in\mathbb{S}} [P^k]_{i,l} P_{l,j},
$$

and the same initial condition, hence by induction on $k \geqslant 0$ the equality

$$
[\ \mathbb{P}(Z_{n+k}=j \mid Z_n=i)\ ]_{i,j\in\mathbb{S}} = \Big[\ [P^k]_{i,j}\ \Big]_{i,j\in\mathbb{S}} = P^k
$$

holds not only for $k=0$ and $k=1$, but also for all $k \in \mathbb{N}$. □

The matrix product relation

$$
P^{m+n} = P^m P^n = P^n P^m,
$$

which reads

$$
[P^{m+n}]_{i,j} = \sum_{l\in\mathbb{S}} [P^m]_{i,l}[P^n]_{l,j} = \sum_{l\in\mathbb{S}} [P^n]_{i,l}[P^m]_{l,j}, \qquad i,j \in \mathbb{S},
$$

can now be interpreted as

$$
\begin{aligned}
\mathbb{P}(Z_{n+m}=j \mid Z_0=i) &= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_m=j \mid Z_0=l)\mathbb{P}(Z_n=l \mid Z_0=i) \\
&= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_n=j \mid Z_0=l)\mathbb{P}(Z_m=l \mid Z_0=i),
\end{aligned}
$$

$i,j \in \mathbb{S}$, which is called the *Chapman-Kolmogorov* equation, cf. also the triple (1.2.4).

Example. The gambling process $(Z_n)_{n\geqslant 0}$.

Taking $S = 4$ and $p = 40\%$, the transition matrix of the gambling process on $\mathbb{S} = \{0, 1, \ldots, 4\}$ of Chapter 2 reads

$$P = \left[\, P_{i,j} \,\right]_{0 \leqslant i,j \leqslant 4} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad (4.4.2)$$

and we can check by hand calculation that:

$$P^2 = P \times P$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0.24 & 0 & 0.16 & 0 \\ 0.36 & 0 & 0.48 & 0 & 0.16 \\ 0 & 0.36 & 0 & 0.24 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Exercise: From the above matrix (4.4.2), check that

$$\mathbb{P}(Z_2 = 4 \mid Z_0 = 2) = [P^2]_{2,4} = 0.16,$$

$$\mathbb{P}(Z_2 = 1 \mid Z_0 = 2) = [P^2]_{2,1} = 0, \qquad \text{and}$$

$$\mathbb{P}(Z_2 = 2 \mid Z_0 = 2) = [P^2]_{2,2} = 0.48.$$

Example. The fifth order transitions of the chain with Markov matrix (4.3.4) can be computed from the fifth matrix power

$$P^5 = \begin{bmatrix} 0.14352 & 0.09600 & 0.25920 & 0.30160 & 0.19968 \\ 0.15840 & 0.10608 & 0.24192 & 0.30400 & 0.18960 \\ 0.17040 & 0.10920 & 0.23280 & 0.30880 & 0.17880 \\ 0.17664 & 0.11520 & 0.22800 & 0.30928 & 0.17088 \\ 0.14904 & 0.09600 & 0.25440 & 0.30520 & 0.19536 \end{bmatrix},$$

cf. *e.g.* *here*. Note that for large transition orders (*e.g.* for 1000 time steps) we get

$$P^{1000} = \begin{bmatrix} 0.16273 & 0.10613 & 0.24056 & 0.30660 & 0.18396 \\ 0.16273 & 0.10613 & 0.24056 & 0.30660 & 0.18396 \\ 0.16273 & 0.10613 & 0.24056 & 0.30660 & 0.18396 \\ 0.16273 & 0.10613 & 0.24056 & 0.30660 & 0.18396 \\ 0.16273 & 0.10613 & 0.24056 & 0.30660 & 0.18396 \end{bmatrix},$$

cf. *e.g.* *here*, which suggests a *convergence phenomenon* in large time for the Markov chain, see Chapter 7 for details.

Example. For the simple random walk of Chapter 3, computing the probability to travel from $\textcircled{0}$ to $\boxed{2k}=\boxed{10}$ in $2n = 20$ time steps involves a summation over $\binom{20}{10+5} = \binom{2n}{n+k} = 15504$ paths, which can be evaluated by computing $[P^{20}]_{0,10}$, cf. also Figure 3.1.

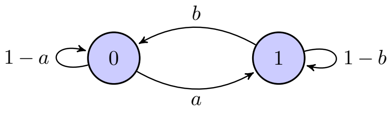
4.5 The Two-State Discrete-Time Markov Chain

The above discussion shows that there is some interest in computing the *n-th* order transition matrix P^n. Although this is generally difficult, this is actually possible when the number of states equals two, *i.e.* $\mathbb{S} = \{0, 1\}$.

We close this chapter with a complete study of the two-state Markov chain, whose transition matrix has the form

$$P = \begin{bmatrix} 1-a & a \\ \\ b & 1-b \end{bmatrix}, \qquad (4.5.1)$$

where $a \in [0, 1]$ and $b \in [0, 1]$.



We also have

$$\mathbb{P}(Z_{n+1} = 1 \mid Z_n = 0) = a, \qquad \mathbb{P}(Z_{n+1} = 0 \mid Z_n = 0) = 1 - a,$$

and

$$\mathbb{P}(Z_{n+1} = 0 \mid Z_n = 1) = b, \qquad \mathbb{P}(Z_{n+1} = 1 \mid Z_n = 1) = 1 - b.$$

The matrix power

$$P^n = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}^n = \underbrace{\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \times \cdots \times \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}}_{n \text{ times}}$$

of the transition matrix P is computed for all $n \geqslant 0$ in the next Proposition 4.2. We always exclude the case $a = b = 0$ since it corresponds to the trivial case where $P = \mathrm{I}$ is the identity matrix (constant chain).

Proposition 4.2. *We have*

$$P^n = \frac{1}{a+b} \begin{bmatrix} b + a(1-a-b)^n & a(1-(1-a-b)^n) \\ b(1-(1-a-b)^n) & a + b(1-a-b)^n \end{bmatrix}, \qquad n \geqslant 0.$$

Proof. This result will be proved by a *diagonalization* argument. The matrix P has two eigenvectors*

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -a \\ b \end{bmatrix},$$

with respective eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - a - b$, see for example the commands Eigensystem[1-a,a,b,1-b] and Diagonalize[1-a,a,b,1-b].

Hence P can be written in the *diagonal form*

$$P = M \times D \times M^{-1}, \tag{4.5.2}$$

i.e.

$$P = \begin{bmatrix} 1 & -a \\ 1 & b \end{bmatrix} \times \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \times \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ -\frac{1}{a+b} & \frac{1}{a+b} \end{bmatrix}.$$

As a consequence of (4.5.2), we have

$$\begin{aligned} P^n &= (M \times D \times M^{-1})^n = (M \times D \times M^{-1}) \cdots (M \times D \times M^{-1}) \\ &= M \times D \times \cdots \times D \times M^{-1} = M \times D^n \times M^{-1}, \end{aligned}$$

where

$$D^n = \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2^n \end{bmatrix}, \qquad n \geqslant 0.$$

hence

* Please refer to MH1201 - Linear Algebra II for more on "eigenvectors, eigenvalues, diagonalization".

$$P^n = \begin{bmatrix} 1 & -a \\ 1 & b \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2^n \end{bmatrix} \times \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ -\frac{1}{a+b} & \frac{1}{a+b} \end{bmatrix}$$
$$= \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{\lambda_2^n}{a+b} \begin{bmatrix} a & -a \\ -b & b \end{bmatrix}$$
$$= \frac{1}{a+b} \begin{bmatrix} b + a\lambda_2^n & a(1-\lambda_2^n) \\ b(1-\lambda_2^n) & a + b\lambda_2^n \end{bmatrix}, \quad (4.5.3)$$

see also the command

MatrixPower[1-a,a,b,1-b,n].

□

For an alternative proof of Proposition 4.2, see also Exercise 1.4.1 page 5 of Norris (1998) in which P^n is written as

$$P^n = \begin{bmatrix} 1-a_n & -a_n \\ b_n & 1-b_n \end{bmatrix}$$

and the relation $P^{n+1} = P \times P^n$ is used to find induction relations for a_n and b_n, cf. the solution of Exercise 7.19 for a similar analysis.

From the result of Proposition 4.2 we may now compute the probabilities

$$\mathbb{P}(Z_n = 0 \mid Z_0 = 0) = \frac{b + a\lambda_2^n}{a+b}, \quad \mathbb{P}(Z_n = 1 \mid Z_0 = 0) = \frac{a(1-\lambda_2^n)}{a+b} \quad (4.5.4)$$

and

$$\mathbb{P}(Z_n = 0 \mid Z_0 = 1) = \frac{b(1-\lambda_2^n)}{a+b}, \quad \mathbb{P}(Z_n = 1 \mid Z_0 = 1) = \frac{a + b\lambda_2^n}{a+b}. \quad (4.5.5)$$

As an example, the value of $\mathbb{P}(Z_3 = 0 \mid Z_0 = 0)$ could also be computed using pathwise analysis as

$$\mathbb{P}(Z_3 = 0 \mid Z_0 = 0) = (1-a)^3 + ab(1-b) + 2(1-a)ab,$$

which coincides with (4.5.4), *i.e.*

$$\mathbb{P}(Z_3 = 0 \mid Z_0 = 0) = \frac{b + a(1-a-b)^3}{a+b},$$

for $n = 3$. Under the condition

$$-1 < \lambda_2 = 1 - a - b < 1,$$

which is equivalent to $(a, b) \neq (0, 0)$ and $(a, b) \neq (1, 1)$, we can let n go to infinity in (4.5.3) to derive the large time behavior, or limiting distribution, of the Markov chain:

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} \begin{bmatrix} \mathbb{P}(Z_n = 0 \mid Z_0 = 0) & \mathbb{P}(Z_n = 1 \mid Z_0 = 0) \\ \mathbb{P}(Z_n = 0 \mid Z_0 = 1) & \mathbb{P}(Z_n = 1 \mid Z_0 = 1) \end{bmatrix} = \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix}.$$

Note that convergence will be faster when $a + b$ is closer to 1.

Hence we have

$$\lim_{n\to\infty} \mathbb{P}(Z_n = 1 \mid Z_0 = 0) = \lim_{n\to\infty} \mathbb{P}(Z_n = 1 \mid Z_0 = 1) = \frac{a}{a+b} \tag{4.5.6}$$

and

$$\lim_{n\to\infty} \mathbb{P}(Z_n = 0 \mid Z_0 = 0) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0 \mid Z_0 = 1) = \frac{b}{a+b}. \tag{4.5.7}$$

Consequently,

$$\pi = [\pi_0, \pi_1] := \left[\frac{b}{a+b}, \frac{a}{a+b}\right] \tag{4.5.8}$$

is a *limiting distribution* as n goes to infinity, provided that $(a, b) \neq (1, 1)$. In other words, whatever the initial state Z_0, the probability of being at ① after a "large" time becomes close to $a/(a + b)$, while the probability of being at ⓪ becomes close to $b/(a + b)$.

In case $a = b = 0$, we have

$$P = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

the chain is constant and it clearly admits its initial distribution as limiting distribution. In case $a = b = 1$, we have

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

and there is no limiting distribution as the chain switches indefinitely between state ⓪ and ①

The notions of limiting and invariant (or stationary) distributions will be treated in Chapter 7 in the general framework of Markov chains, see for example Proposition 7.7.

Remarks

i) The limiting distribution π in (4.5.8) is *invariant* (or stationary) by P in the sense that

$$\pi P = \frac{1}{a+b}[b,a]\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \frac{1}{a+b}\begin{bmatrix} b(1-a)+ab \\ ab+a(1-b) \end{bmatrix}^\top$$
$$= \frac{1}{a+b}[b,a] = \pi,$$

i.e. π is invariant (or stationary) with respect to P, and the invariance relation (4.2.4):

$$\pi = \pi P,$$

which means that $\mathbb{P}(Z_1 = k) = \pi_k$ if $\mathbb{P}(Z_0 = k) = \pi_k$, $k = 0, 1$. For example, the distribution $\pi = [1/2, 1/2]$ is clearly invariant (or stationary) for the swapping chain with $a = b = 1$ and transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

while $\pi := [1/3, 2/3]$ will *not* be invariant (or stationary) for this chain. This is a two-state particular case of the circular chain of Example (7.2.4).

ii) If $a + b = 1$, one sees that

$$P^n = \begin{bmatrix} b & a \\ b & a \end{bmatrix} = P$$

for all $n \in \mathbb{N}$ and we find

$$\mathbb{P}(Z_n = 1 \mid Z_k = 0) = \mathbb{P}(Z_n = 1 \mid Z_k = 1) = \mathbb{P}(Z_n = 1) = a$$

and

$$\mathbb{P}(Z_n = 0 \mid Z_k = 0) = \mathbb{P}(Z_n = 0 \mid Z_k = 1) = \mathbb{P}(Z_n = 0) = b$$

for all $k = 0, 1, \ldots, n-1$, regardless of the initial distribution $[\mathbb{P}(Z_0 = 0), \mathbb{P}(Z_0 = 1)]$. In this case, Z_n is independent of Z_k as in (S.31), as we have

$$\mathbb{P}(Z_n = i,\ Z_k = j) = \mathbb{P}(Z_n = i \mid Z_k = j)\mathbb{P}(Z_k = j) = \mathbb{P}(Z_n = i)\mathbb{P}(Z_k = j),$$

$i, j = 0, 1$, $0 \leqslant k < n$, and $(Z_n)_{n\geqslant 0}$ is an *i.i.d* sequence of random variables with distribution $(1-a, a) = (b, a)$ over $\{0, 1\}$, cf. (S.31) in the general case.

iii) A given proportion $p = a/(a+b) \in (0,1)$ of visits to state ① in the long run can be reached by any $a \in (0,p]$ and $b \in (0,1-p]$ satisfying $a = bp/(1-p)$. Smaller values of a and b will lead to increased *stickiness*, see also *here*. The case $(a,b) = (p,1-p)$ satisfies $a+b=1$ and corresponds to minimal stickiness, *i.e.* to the independence of the sequence $(Z_n)_{n\geqslant 0}$.

iv) When $a = b = 1$ in (4.5.1) the limit $\lim_{n\to\infty} P^n$ does not exist as we have

$$P^n = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & n = 2k, \\ \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & n = 2k+1, \end{cases}$$

and the chain is indefinitely switching at each time step from one state to the other.

In Figures 4.4 and 4.5 we consider a simulation of the two-state random walk with transition matrix

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix},$$

i.e. $a = 0.2$ and $b = 0.4$. Figure 4.4 represents a sample path $(x_n)_{n=0,1,\ldots,100}$ of the chain, while Figure 4.5 represents the sample average

$$y_n = \frac{1}{n+1}(x_0 + x_1 + \cdots + x_n), \qquad n = 0,1,\ldots,100,$$

which counts the proportion of values of the chain in the state ①. This proportion is found to converge to $a/(a+b) = 1/3$. This is actually a consequence of the Ergodic Theorem, cf. Theorem 7.12 in Chapter 7.

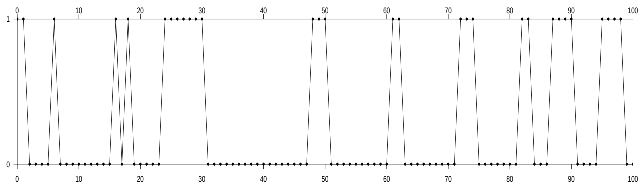


Fig. 4.4: Sample path of a two-state chain in continuous time with $a = 0.2$ and $b = 0.4$.

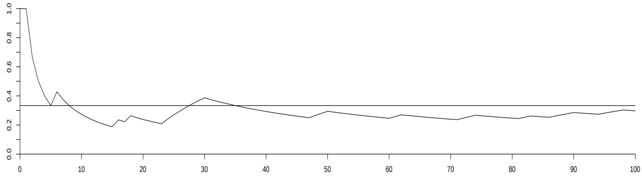


Fig. 4.5: The proportion of chain values at state 1 tends to $1/3 = a/(a+b)$.

The source code of the above program, written in R, is given below.*

```
dim=2 # Dimension of the transition matrix
a=0.2; b=0.4; # Parameter definition
# Definition of the transition matrix
P=matrix(c(1-a,a,b,1-b),nrow=dim,ncol=dim,byrow=TRUE)
N=100 # Number of time steps
Z=array(N+1); # Encoding of chain values
for(ll in seq(1,N)) {Z[1]=sample(dim,size=1,prob=P[2,])
# Random simulation of Z[j+1] given Z[j]
for (j in seq(1,N)) Z[j+1]=sample(dim,size=1,prob=P[Z[j],]);Y=array(N+1);S=0;
# Computation of the average over the l first steps
for(l in seq(1,N+1)) { Z[l]=Z[l]-1; S=S+Z[l]; Y[l]=S/l;}
X=array(N+1); for(l in seq(1,N+1)) {X[l]=l-1;};par(mfrow=c(2,1))
plot(X,Y,type="l",yaxt="n",xaxt="n",xlim=c(0,N),xlab="", ylim=c(0,1),ylab="",xaxs="i",
      col="black", main="",bty="n")
segments(0,a/(a+b),N,a/(a+b))
axis(2,pos=0,at=c(0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0))
axis(1,pos=0,at=seq(0,N,10),outer=TRUE)
plot(X,Z,type="o",xlab="",ylab="",xlim=c(0,N),yaxt="n",xaxt="n",xaxs="i",
col="black",main="",pch=20,bty="n")
axis(1,pos=1,at=seq(0,N+1,10),outer=TRUE,padj=-4,tcl=0.5)
axis(1,pos=0,at=seq(0,N+1,10),outer=TRUE);axis(2,las=2,at=0:1)
readline(prompt = "Pause. Press <Enter> to continue...")}
```

We close this chapter with two other Markov chain sample paths presented in Figures 4.6 and 4.7 (use the controls to see the animations under Acrobat Reader). In the next Figure 4.6 we check again that the proportion of chain values in the state (1) converges to $1/3$ for a two-state Markov chain.

* Download the corresponding R **code** or the **IPython notebook** that can be run ***here***.

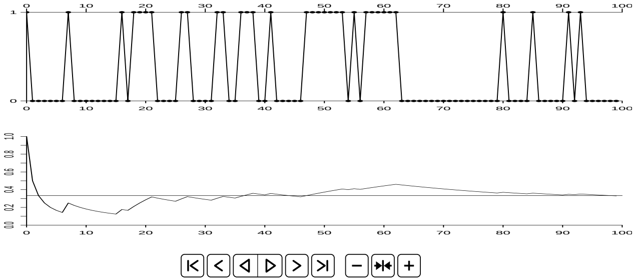


Fig. 4.6: Convergence graph for the two-state Markov chain with $a = 0.2$ and $b = 0.4$.*

In Figure 4.7 we draw a sample path of a five-state Markov chain.[†]

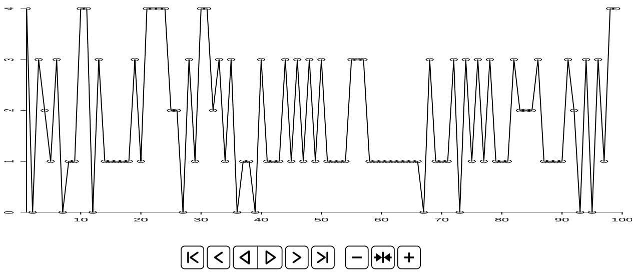


Fig. 4.7: Sample path of a five-state Markov chain.[‡]

Exercises

Exercise 4.1 Consider the *symmetric* random walk $(S_n)_{n\geqslant 0}$ on $\mathbb{S} = \mathbb{Z}$ with independent increments ± 1 chosen with equal probability $1/2$, started at $S_0 = 0$.

a) Is the process $Z_n := 2S_n + 1$ a Markov chain?

[†] Download the corresponding R **code** or the **IPython notebook** that can be run ***here***.

[‡] Animated figure (works in Acrobat Reader).

b) Is the process $Z_n := (S_n)^2$ a Markov chain?

Exercise 4.2 Consider the Markov chain $(Z_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{1, 2\}$ and transition matrix

$$P = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \end{matrix}.$$

a) Compute $\mathbb{P}(Z_7 = 1 \text{ and } Z_5 = 2 \mid Z_4 = 1 \text{ and } Z_3 = 2)$.
b) Compute $\mathbb{E}[Z_2 \mid Z_1 = 1]$.

Exercise 4.3 Consider a transition probability matrix P of the form (S.30), *i.e.*

$$P = \left[P_{i,j} \right]_{0\leqslant i,j\leqslant N} = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \end{bmatrix},$$

where $\pi = [\pi_0, \pi_1, \ldots, \pi_N] \in [0, 1]^{N+1}$ is a vector such that

$$\pi_0 + \pi_1 + \cdots + \pi_N = 1.$$

a) Compute the matrix power P^n for all $n \geqslant 2$.
b) Show that the vector π is an invariant (or stationary) distribution for P.
c) Show that if $\mathbb{P}(Z_0 = i) = \pi_i$, $i = 0, 1, \ldots, N$, then Z_n is independent of Z_k for all $0 \leqslant k < n$, and $(Z_n)_{n\geqslant 0}$ is an *i.i.d* sequence of random variables with distribution $\pi = [\pi_0, \pi_1, \ldots, \pi_N]$ over $\{0, 1, \ldots, N\}$.

Exercise 4.4 Consider a $\{0, 1\}$-valued "hidden" two-state Markov chain $(X_n)_{n\geqslant 0}$ with *transition probability matrix*

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} \\ P_{1,0} & P_{1,1} \end{bmatrix} = \begin{bmatrix} \mathbb{P}(X_1 = 0 \mid X_0 = 0) & \mathbb{P}(X_1 = 1 \mid X_0 = 0) \\ \mathbb{P}(X_1 = 0 \mid X_0 = 1) & \mathbb{P}(X_1 = 1 \mid X_0 = 1) \end{bmatrix},$$

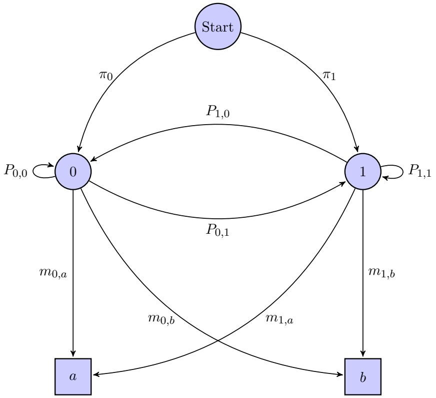
and initial distribution

$$\pi = [\pi_0, \pi_1] = [\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1)].$$

We observe a two-state process $(O_k)_{k\geqslant 0}$ whose state $O_k \in \{a, b\}$ at every time $k \in \mathbb{N}$ has a conditional distribution given $X_k \in \{0, 1\}$ denoted by

$$M = \begin{bmatrix} m_{0,a} & m_{0,b} \\ m_{1,a} & m_{1,b} \end{bmatrix} = \begin{bmatrix} \mathbb{P}(O_k = a \mid X_k = 0) & \mathbb{P}(O_k = b \mid X_k = 0) \\ \mathbb{P}(O_k = a \mid X_k = 1) & \mathbb{P}(O_k = b \mid X_k = 1) \end{bmatrix},$$

called the *emission probability* matrix.



a) Using the matrix entries of π, P and M, compute the four conditional probabilities

$$\begin{cases} \mathbb{P}\big((O_0, O_1) = (a, b) \mid (X_0, X_1) = (0, 0)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \mid (X_0, X_1) = (0, 1)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \mid (X_0, X_1) = (1, 0)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \mid (X_0, X_1) = (1, 1)\big). \end{cases}$$

Hint: By independence, the conditional probability of observing $(O_0, O_1) = (a, b)$ given that $(X_0, X_1) = (x, y)$ splits as

$$\mathbb{P}((O_0, O_1) = (a, b) \mid (X_0, X_1) = (x, y)) \\ = \mathbb{P}(O_0 = a \mid X_0 = x)\mathbb{P}(O_1 = b \mid X_1 = y).$$

b) Using π, P and M, compute $\mathbb{P}(X_0 = 1, X_1 = 1)$ and the four probabilities

$$\begin{cases} \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (0, 0)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (0, 1)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (1, 0)\big), \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (1, 1)\big). \end{cases}$$

c) Find the probability $\mathbb{P}((O_0, O_1) = (a, b))$ that the observed sequence is (a, b).

Hint: Use the law of total probability based on all possible values of (X_0, X_1).

d) Compute the hidden state probabilities

$$\begin{cases} \mathbb{P}((X_0, X_1) = (0, 1) \mid (O_0, O_1) = (a, b)), \\ \\ \mathbb{P}((X_0, X_1) = (1, 0) \mid (O_0, O_1) = (a, b)), \end{cases}$$

and

$$\mathbb{P}(X_1 = 1 \mid (O_0, O_1) = (a, b)) \text{ and } \mathbb{P}(X_1 = 0 \mid (O_0, O_1) = (a, b)),$$

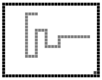
given that the observed sequence is (a, b).

Exercise 4.5 Consider the two-dimensional random walk $(S_n)_{n \geqslant 0}$ started at $S_0 = (0, 0)$ on $\mathbb{S} = \mathbb{Z}^2$, where, starting from a location $S_n = (i, j)$ the chain can move to any of the points $(i + 1, j + 1)$, $(i + 1, j - 1)$, $(i - 1, j + 1)$, $(i - 1, j - 1)$ with equal probability $1/4$. The next Questions (a) and (b) are independent.

a) We denote by $S_n = (X_n, Y_n)$ the coordinates of S_n at time $n \geqslant 0$ and let $Z_n := X_n^2 + Y_n^2$. Does the process $(Z_n)_{n \geqslant 0}$ have the Markov property? Justify your answer.

Hint: Use the fact that a same value of Z_n may correspond to different locations of (X_n, Y_n) on the circle, for example $(X_n, Y_n) = (5, 0)$ and $(X_n, Y_n) = (4, 3)$ when $Z_n = 25$.

b) Suppose in addition that the random walk cannot visit any site more than once, as in a *snake game*. In this case, does the random walk $(S_n)_{n \geqslant 0}$ have the Markov property? Justify your answer.



Questions (a) and (b) above are independent.

Exercise 4.6 The Elephant Random Walk $(S_n)_{n\geqslant 0}$ (Schütz and Trimper (2008)) is a discrete-time $\mathbb{Z}$-valued random walk

$$S_n := X_1 + \cdots + X_n, \qquad n \in \mathbb{N},$$

whose increments $X_k = S_k - S_{k-1}$, $k \geqslant 1$, are recursively defined as follows:

- At time $n = 1$, X_1 is a Bernoulli $\{-1, +1\}$-valued random variable with

$$\mathbb{P}(X_1 = +1) = p \quad \text{and} \quad \mathbb{P}(X_1 = -1) = q = 1 - p \in (0, 1).$$

- At any subsequent time $n \geqslant 2$, one randomly draws an integer time index $k \in \{1, 2, \ldots, n-1\}$ with uniform probability, and lets $X_n := X_k$ with probability p, and $X_n := -X_k$ with probability $q := 1 - p$.

Does the Elephant Random Walk $(S_n)_{n\geqslant 0}$ have the Markov property?

Exercise 4.7 Consider a two-state Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0, 1\}$ and transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \end{array},$$

where $a, b \in (0, 1)$, and define a new stochastic process $(Z_n)_{n\geqslant 1}$ by $Z_n = (X_{n-1}, X_n)$, $n \geqslant 1$. Argue that $(Z_n)_{n\geqslant 1}$ is a Markov chain, and write down its transition matrix. Start by determining the state space of $(Z_n)_{n\geqslant 1}$. See Problem 5.33 for the use of this construction in pattern recognition.

Exercise 4.8 Given $p \in [0, 1)$, consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\{0, 1, 2\}$ having the transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} p & q & 0 \\ 0 & p & q \\ 0 & 0 & 1 \end{bmatrix} \end{array},$$

with $q := 1 - p$.

a) Give the probability distribution of the first hitting time

$$T_2 := \inf\{n \geqslant 0 \ : \ X_n = 2\}.$$

of state ② starting from $X_0 = ⓪$.

Hint: The sum $Z = X_1 + \cdots + X_n$ of n independent geometrically distributed random variables $X_1, \ldots X_n$ on $\{1, 2, \ldots\}$ has the negative binomial distribution

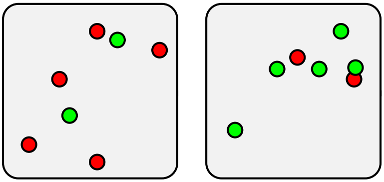
$$\mathbb{P}(Z = k \mid X_0 = 1) = \binom{k-1}{k-d}(1-p)^d p^{k-d}, \qquad k \geqslant d.$$

b) Compute the mean hitting time $\mathbb{E}[T_2 \mid X_0 = 0]$ of state ② starting from $X_0 = 0$.

Hint: We have

$$\sum_{k\geqslant 1} kp^{k-1} = \frac{1}{(1-p)^2} \quad \text{and} \quad \sum_{k\geqslant 2} k(k-1)p^{k-2} = \frac{2}{(1-p)^3}, \qquad 0 \leqslant p < 1.$$

Exercise 4.9 Bernoulli-Laplace chain. Consider two boxes and a total of $2N$ balls made of N red balls and N green balls. At time 0, a number $k = X_0$ of red balls and a number $N - k$ of green balls are placed in the first box, while the remaining $N - k$ red balls and k green balls are placed in the second box.



At each unit of time, one ball is chosen randomly out of N in each box, and the two balls are interchanged. Write down the transition matrix of the Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0, 1, 2, \ldots, N\}$, representing the number of red balls in the first box. Start for example from $N = 5$.

Exercise 4.10

a) After winning k dollars, a gambler either receives $k + 1$ dollars with probability p, or has to *quit* the game and lose everything with probability $q = 1 - p$. Starting from *one* dollar, find a model for the time evolution of the wealth of the player using a Markov chain whose transition probability matrix P will be described explicitly along with its powers P^n of all orders $n \geqslant 1$.
b) (Success runs Markov chain). We modify the model of Question (a) by allowing the gambler to start playing again and win with probability p after reaching state ⓪. Write down the corresponding transition probability matrix P, and compute P^n for all $n \geqslant 2$.

Exercise 4.11 Let $(X_k)_{k\geqslant 0}$ be the Markov chain with transition matrix

$$P = \begin{bmatrix} 1/4 & 0 & 1/2 & 1/4 \\ 0 & 1/5 & 0 & 4/5 \\ 0 & 1 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \end{bmatrix}.$$

A new process is defined by letting

$$Z_n := \begin{cases} 0 & \text{if } X_n = 0 \text{ or } X_n = 1, \\ \\ X_n & \text{if } X_n = 2 \text{ or } X_n = 3, \end{cases}$$

i.e.

$$Z_n = X_n \mathbb{1}_{\{X_n \in \{2,3\}\}}, \qquad n \geqslant 0.$$

a) Compute

$$\mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 2)$$

and

$$\mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 3),$$

$n \geqslant 1$.

b) Is $(Z_n)_{n \geqslant 0}$ a Markov chain?

Exercise 4.12 (Olaleye et al. (2009)). Abeokuta, one of the major towns of the defunct Western Region of Nigeria, has recently seen an astronomic increase in vehicular activities. The intensity of vehicle traffic at the Lafenwa intersection which consists of Ayetoro, Old Bridge and Ita-Oshin routes, is modeled according to three states L/M/H = { Low / Moderate / High }.

a) During year 2005, low intensity incoming traffic has been observed at Lafenwa intersection for $\eta_L = 50\%$ of the time, moderate traffic has been observed for $\eta_M = 40\%$ of the time, while high traffic has been observed during $\eta_H = 10\%$ of the time.

Fig. 4.8: Lafenwa intersection.

Given the correspondence table

Incoming traffic	Vehicles per hour
L (low intensity)	360
M (medium intensity)	505
H (high intensity)	640

compute the average incoming traffic per hour in year 2005.

b) The analysis of incoming daily traffic volumes at Lafenwa intersection between years 2004 and 2005 shows that the probability of switching states within {L, M, H} is given by the Markov transition probability matrix

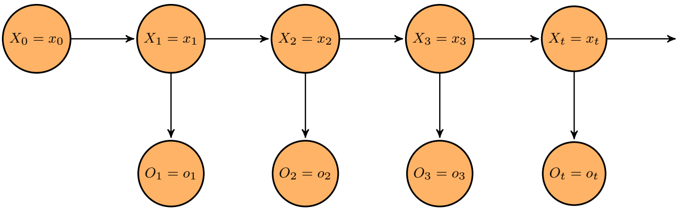
$$P = \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}.$$

Based on the knowledge of P and $\eta = [\eta_L, \eta_M, \eta_H]$, give a projection of the respective proportions of traffic in the states L/M/H for year 2006.

c) Based on the result of Question (b), give a projected estimate for the average incoming traffic per hour in year 2006.

d) By solving the equation $\pi = \pi P$ for the invariant (or stationary) probability distribution $\pi = [\pi_L, \pi_M, \pi_H]$, give a long term projection of steady traffic at Lafenwa intersection. *Hint*: we have $\pi_L = 11/24$.

Exercise 4.13 Consider the graphical hidden Markov model



with the relation

$$\begin{aligned} &\mathbb{P}(X_t = i_t, \ldots, X_0 = i_0,\ O_t = o_t, \ldots, O_1 = o_1) \\ &= \mathbb{P}(O_t = o_t \mid X_t = i_t) \cdots \mathbb{P}(O_1 = o_1 \mid X_1 = i_1) \\ &\quad \times \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}) \cdots \mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_0 = i_0), \quad t \geqslant 0. \end{aligned}$$

a) Show that

$$\mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}),$$

$t \geqslant 1$.

b) Show that

$$\mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1},\ O_{t-1} = o_{t-1}, \ldots, O_1 = o_1) = \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}),$$

$t \geqslant 1$.

Problem 4.14 Phase-type distributions are used in insurance to model the heavy-tailed random claim sizes appearing in reserve and surplus processes.

a) Given $p \in [0,1]$, construct the transition matrix of a two-state Markov chain on the state space $\{0,1\}$ satisfying the following two conditions:

   i) State ⓪ is absorbing, i.e. $\mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = 1$, *and*
   ii) The first hitting time

$$T_0 := \inf\{n \geqslant 0 \ : \ X_n = 0\}$$

   of state ⓪ starting from state ① has the geometric distribution p given by

$$\mathbb{P}(T_0 = k \mid X_0 = 1) = (1-p)p^{k-1}, \qquad k \geqslant 1.$$

b) Given $p \in [0,1]$ and $d \geqslant 1$, construct the transition matrix of a $d+1$-state Markov chain on the state space $\{0,1,\ldots,d\}$ satisfying the following two conditions:

   i) State ⓪ is absorbing, i.e. $\mathbb{P}(X_{k+1} = 0 \mid X_k = 0) = 1$, *and*
   ii) The first hitting time T_0 of state ⓪ starting from state d has the negative binomial distribution

$$\mathbb{P}(T_0 = k \mid X_0 = d) = \binom{k-1}{k-d}(1-p)^d p^{k-d}, \qquad k \geqslant d.$$

c) In what follows we consider a discrete-time Markov chain on $\{0,1,\ldots,d\}$ having d transient* states $\{1,2,\ldots,d\}$, and 0 as absorbing state. Show that its transition matrix P takes the form

$$P = \left[\, P_{i,j} \,\right]_{0 \leqslant i,j \leqslant d} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_1 & Q_{1,1} & \cdots & Q_{1,d} \\ \alpha_2 & Q_{2,1} & \cdots & Q_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_d & Q_{d,1} & \cdots & Q_{d,d} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha & Q \end{bmatrix},$$

where α is the column vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_d]^\top$ and Q is the $d \times d$ matrix

* Here the transience condition implies that $\mathbb{P}(T_0 < \infty \mid X_0 = i) = 1$ for all $i = 1,2,\ldots,d$, it will be ensured by assuming that $I - Q$ is invertible, cf. § 5.4 and § 6.3 for details.

$$Q = \begin{bmatrix} Q_{1,1} & \cdots & Q_{1,d} \\ \vdots & \ddots & \vdots \\ Q_{d,1} & \cdots & Q_{d,d} \end{bmatrix},$$

and give the relation between the vector α and the matrix Q using the column vector $e := [1, 1, \ldots, 1]^\top$.

d) Show by induction on $n \geqslant 0$ that we have

$$P^n = \begin{bmatrix} 1 & 0 \\ (I - Q^n)e & Q^n \end{bmatrix}, \tag{4.5.9}$$

where I is the $d \times d$ identity matrix

$$I := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

e) Compute the probability distribution of T_0 using the vector α and the matrix Q^{n-1}, where T_0 is the first hitting time of state ⓪ starting from state $i \geqslant 1$.

Hint: Partition the event $\{T_0 = n\}$ as

$$\{T_0 = n\} = \bigcup_{k=1}^{d} \{X_n = 0 \text{ and } X_{n-1} = k\}.$$

f) From now on we assume that the initial distribution of X_0 is given by the d-dimensional vector

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix},$$

i.e. $\beta_i = \mathbb{P}(X_0 = i)$, $i = 1, 2, \ldots, d$, with $\mathbb{P}(X_0 = 0) = 0$. Compute the probability distribution of T_0 using the vectors α, β, and the matrix Q^{n-1}.

g) Compute the cumulative distribution function $P(T_0 \leqslant n)$ of T_0 the vectors β, e, and the matrix Q^n.

h) Recover the result of Question (g) by rewriting $\mathbb{P}(T_0 \leqslant n)$ as the probability of not being in any state $i = 1, 2, \ldots, d$ at time n.

i) Compute the probability generating function

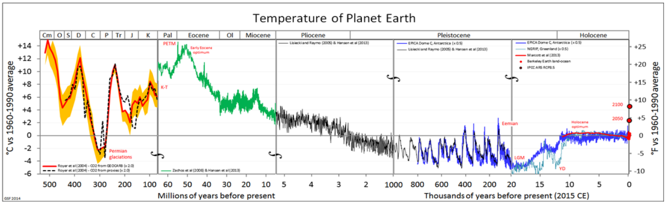
$$G_{T_0}(s) := \sum_{k \geqslant 0} s^k \mathbb{P}(T_0 = k)$$

of T_0, assuming the existence of the matrix inverse $(I - sQ)^{-1}$ given by the series

$$(I - sQ)^{-1} = \sum_{k \geqslant 0} s^k Q^k, \qquad s \in (-1, 1].$$

j) Using the probability generating function $s \mapsto G_{T_0}(s)$, compute the first and second moments $E[T_0]$ and $E[T_0^2]$ of T_0.

Problem 4.15 Hidden Markov models have applications to speech recognition, face recognition, emotion recognition, genomics and biological sequence analysis, sentiment analysis, unsupervised learning, climate change studies, etc. Consider a $\{0,1\}$-valued "hidden" two-state Markov chain $(X_n)_{n \geqslant 0}$ with



transition probability matrix

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} \\ \\ P_{1,0} & P_{1,1} \end{bmatrix} = \begin{bmatrix} \mathbb{P}(X_1 = 0 \mid X_0 = 0) & \mathbb{P}(X_1 = 1 \mid X_0 = 0) \\ \\ \mathbb{P}(X_1 = 0 \mid X_0 = 1) & \mathbb{P}(X_1 = 1 \mid X_0 = 1) \end{bmatrix},$$

and initial distribution

$$\pi = [\pi_0, \pi_1] = [\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1)].$$

We observe a stochastic process $(O_k)_{k \geqslant 0}$ whose state $O_k \in \{a, b, c\}$ at every time $k \in \mathbb{N}$ has a conditional distribution given $X_k \in \{0, 1\}$ denoted by

$$\begin{aligned} M &= \begin{bmatrix} m_{0,a} & m_{0,b} & m_{0,c} \\ m_{1,a} & m_{1,b} & m_{1,c} \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{P}(O_k = a \mid X_k = 0) & \mathbb{P}(O_k = b \mid X_k = 0) & \mathbb{P}(O_k = c \mid X_k = 0) \\ \mathbb{P}(O_k = a \mid X_k = 1) & \mathbb{P}(O_k = b \mid X_k = 1) & \mathbb{P}(O_k = c \mid X_k = 1) \end{bmatrix}, \end{aligned}$$

called the *emission probability* matrix.

a) Using the matrix entries of π, P and M, compute $\mathbb{P}(X_0 = 1, X_1 = 1, X_2 = 0)$ and the probability

$$\mathbb{P}((O_0, O_1, O_2) = (c, a, b) \text{ and } (X_0, X_1, X_2) = (1, 1, 0))$$

of observing the sequence $(O_0, O_1, O_2) = (c, a, b)$ when $(X_0, X_1, X_2) = (1, 1, 0)$.

Hint: By independence, the conditional probability of observing $(O_0, O_1, O_2) = (c, a, b)$ given that $(X_0, X_1, X_2) = (1, 1, 0)$ splits as

$$\mathbb{P}((O_0, O_1, O_2) = (c, a, b) \mid (X_0, X_1, X_2) = (1, 1, 0)) \\ = \mathbb{P}(O_0 = c \mid X_0 = 1)\mathbb{P}(O_1 = a \mid X_1 = 1)\mathbb{P}(O_2 = b \mid X_2 = 0).$$

b) Find the probability $\mathbb{P}((O_0, O_1, O_2) = (c, a, b))$ that the observed sequence is (c, a, b).

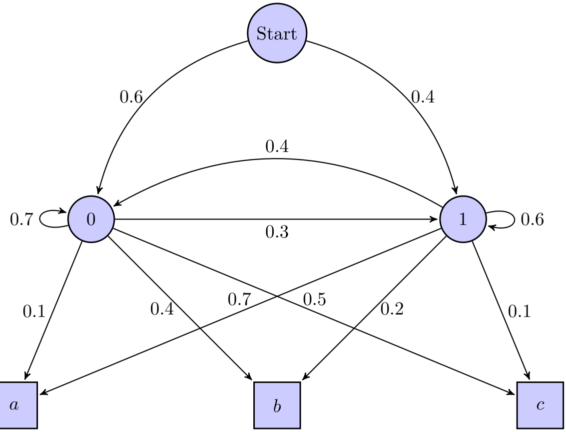
Hint: Use the law of total probability based on all possible values of (X_0, X_1, X_2).

c) Compute the probabilities

$$\mathbb{P}(X_1 = 1 \mid (O_0, O_1, O_2) = (c, a, b)), \quad \text{and} \quad \mathbb{P}(X_1 = 0 \mid (O_0, O_1, O_2) = (c, a, b)).$$

From Question (d) to Question (i) below we take $\pi = [\pi_0, \pi_1] := [0.6, 0.4]$ **and**

$$P := \begin{bmatrix} 0.7 & 0.3 \\ \\ 0.4 & 0.6 \end{bmatrix}, \quad M := \begin{bmatrix} m_{0,a} & m_{0,b} & m_{0,c} \\ m_{1,a} & m_{1,b} & m_{1,c} \end{bmatrix} = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}. \quad (4.5.10)$$



d) Compute* the eight probabilities

* Give a numerical value with exactly three significant digits.

$$\mathbb{P}((X_0, X_1, X_2) = (x, y, z) \text{ and } (O_0, O_1, O_2) = (c, a, b))$$

for all $x, y, z \in \{0, 1\}$, and identify the most likely sample sequence of values for (X_0, X_1, X_2).

e) Compute[1] the probability $\mathbb{P}((O_0, O_1, O_2) = (c, a, b))$ that the observed sequence is (c, a, b).

f) Compute[1] the six probabilities

$$\mathbb{P}(X_k = 1 \mid (O_0, O_1, O_2) = (c, a, b)), \quad \mathbb{P}(X_k = 0 \mid (O_0, O_1, O_2) = (c, a, b)),$$

$k = 0, 1, 2$. What is the most likely sequence for (X_0, X_1, X_2) according to this computation?

Based on the knowledge of $\mathbb{P}((X_0, X_1, X_2) = (1, 1, 0) \mid (O_0, O_1, O_2) = (c, a, b))$, one can build an estimator $\widehat{P}$, $\hat{\pi}$, $\widehat{M}$ of the model parameters P, π and M, as

$$\begin{cases} \hat{\pi}_i := \mathbb{P}(X_0 = i \mid (O_0, O_1, O_2) = (c, a, b)), & \text{(4.5.11a)} \\ \widehat{P}_{i,j} := \dfrac{\displaystyle\sum_{t=0}^{N-1} \mathbb{P}(X_t = i, X_{t+1} = j \mid (O_0, O_1, O_2) = (c, a, b))}{\displaystyle\sum_{t=0}^{N-1} \mathbb{P}(X_t = i \mid (O_0, O_1, O_2) = (c, a, b))} & \text{(4.5.11b)} \\ \widehat{M}_{i,k} := \dfrac{\displaystyle\sum_{t=0}^{N} \mathbb{1}_{\{O_t = k\}} \mathbb{P}(X_t = i \mid (O_0, O_1, O_2) = (c, a, b))}{\displaystyle\sum_{t=0}^{N} \mathbb{P}(X_t = i \mid (O_0, O_1, O_2) = (c, a, b))}, & \text{(4.5.11c)} \end{cases}$$

with here $N = 2$. Note that (4.5.11c) averages the number of times the observed state is "k" given that the hidden state is "i", which gives an estimate of the conditional probability $M_{i,k}$.

g) Compute* the vector estimate $\hat{\pi} = [\hat{\pi}_0, \hat{\pi}_1]$ using (4.5.11a) and the data of Equation (4.5.10).

h) Compute[1] the matrix estimate $\widehat{P}$ using (4.5.11b) and the data of Equation (4.5.10).

i) Compute[1] the matrix estimate $\widehat{M}$ using (4.5.11c) and the data of Equation (4.5.10).

j) Iterating the estimates (4.5.11a)-(4.5.11c) is computationally intensive, however this procedure admits an efficient recursive implementation via

* Give a numerical value with exactly three significant digits.

the Baum-Welch algorithm which is based on the Expectation-Maximization (EM) algorithm.

Imagine an alien trying to analyse an English manuscript without any prior knowledge of English. Using a simple two-state hidden chain $(X_n)_{n\geqslant 0}$ he will try to uncover some *features* of the language, starting with a *binary classification* of the alphabet.

```
install.packages("HMM")
library (HMM)
library (lattice)
text = readChar("my_own_text_file.txt",nchars=10000)
data <- unlist (strsplit (gsub ("[^a-z]", "_", tolower (text)), ""))
pi=c(0.4,0.6)
P=t(matrix(c(c(0.6177499,0.3822501),c(0.8826096,0.1173904)),nrow=2,ncol=2))
M=t(matrix(c(c(0.037192964,0.009902360,0.032833978,0.044882670,0.057331132,
0.052143890,0.013665015,0.036187536,0.072293323,0.044793972,0.060008388,
0.004256270,0.024770706,0.053520546,0.014232306,0.046981769,0.053733382,
0.066355203,0.046817817,0.006912535,0.016201697,0.013425499,0.024694447,
0.064902148,0.046170421,0.033586536,0.022203489),
c(0.0389931197,0.0697183142,0.0239154174,0.0512772632,0.0404732634,0.0059687348,
0.0211687193,0.0625229746,0.0039632091,0.0567828864,0.0468108656,0.0168355418,
0.0627882213,0.0286478204,0.0389215263,0.0064318198,0.0001698078,0.0493758725,
0.0652709152,0.0069580806,0.0093043072,0.0028807932,0.0521827110,0.0608822385,
0.0645417465,0.0555249876,0.0576888424)),nrow=27,ncol=2))
model <- initHMM (c("0", "1"),c("_", letters), pi, P, M)
system.time (estimate <- baumWelch (model, data, 100)) # 100 iterations
xyplot(estimate$hmm$emissionProbs[1,] ~ c(1:27), scales=list(x=list(at=1:27,
labels=c("_", letters))),type="h", lwd=5, xlab="", ylab="")
```

Using a text of your choice and a $\{0,1\}$-valued hidden Markov chain $(X_n)_{n\geqslant 0}$, estimate the corresponding emission probability matrix M by running the above R code which relies on the HMM (Hidden Markov Model) package. A text length of $N \simeq 10,000$ characters can be a minimum. The initial values of π, P and M have been set randomly.

What do you conclude?

Instructions:

Specify your choice of document used.
Attach a graph with a discussion of your output.

Suggestions and possible variations:

Try a language different from English.

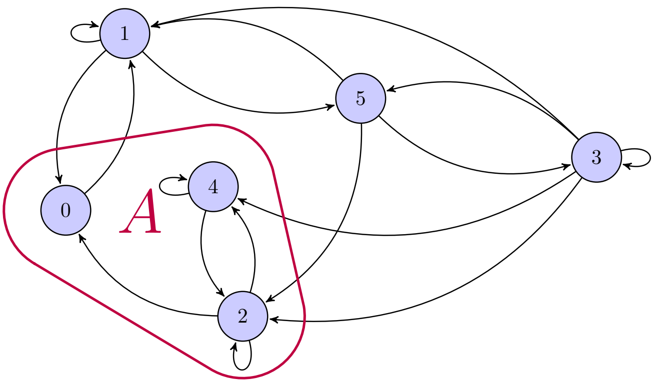
Increase the state space of $(X_n)_{n\geqslant 0}$ in order to uncover more features of the chosen language.

Chapter 5
First Step Analysis

Starting with this chapter we introduce the systematic use of the first step analysis technique, in a general framework that covers the examples of random walks already treated in Chapters 2 and 3. The main applications of first step analysis are the computation of hitting probabilities, mean hitting and absorption times, mean first return times, and average number of returns to a given state.

5.1 Hitting Probabilities

Let us consider a Markov chain $(Z_n)_{n\geqslant 0}$ with state space $\mathbb{S}$, and let $A \subset \mathbb{S}$ denote a subset of $\mathbb{S}$, as in the following example with $\mathbb{S} = \{0, 1, 2, 3, 4, 5\}$ and $A := \{0, 2, 4\}$.



We are interested in the first time T_A the chain hits the subset A, where

$$T_A := \inf\{n \geqslant 0 \ : \ Z_n \in A\}, \tag{5.1.1}$$

with $T_A = 0$ if $Z_0 \in A$ and

$$T_A = +\infty \quad \text{if} \quad \{n \geqslant 0 \ : \ Z_n \in A\} = \emptyset,$$

i.e. if $Z_n \notin A$ for all $n \in \mathbb{N}$. In case the transition matrix P satisfies

$$P_{k,l} = \mathbb{1}_{\{k=l\}} \quad \text{for all} \quad k, \ l \in A, \tag{5.1.2}$$

the set $A \subset \S$ is said to be *absorbing*.

Similarly to the gambling problem of Chapter 2, we would like to compute the probabilities

$$g_l(k) = \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = k)$$

of hitting the set $A \subset \mathbb{S}$ through state $l \in A$ starting from $k \in \mathbb{S}$, where Z_{T_A} represents the location of the chain $(Z_n)_{n \geqslant 0}$ at the hitting time T_A.

This computation can be achieved by first step analysis, using the *law of total probability* (1.3.1) for the probability measure $\mathbb{P}(\cdot \mid Z_0 = k)$ and the Markov property, as follows.

Proposition 5.1. *Assume that* (5.1.2) *holds. The hitting probabilities*

$$g_l(k) := \mathbb{P}(Z_{T_A} = l \textit{ and } T_A < \infty \mid Z_0 = k), \qquad k \in \mathbb{S}, \ l \in A,$$

satisfy the equation

$$g_l(k) = \sum_{m\in\mathbb{S}} P_{k,m} g_l(m) = P_{k,l} + \sum_{m\in\mathbb{S}\setminus A} P_{k,m} g_l(m), \qquad (5.1.3)$$

$k \in \mathbb{S} \setminus A$, $l \in A$, subject to the boundary conditions

$$g_l(k) = \mathbb{P}(Z_{T_A} = l \textit{ and } T_A < \infty \mid Z_0 = k) = \mathbb{1}_{\{k=l\}} = \begin{cases} 1 \textit{ if } k = l, \\ 0 \textit{ if } k \neq l, \end{cases}$$

$k, l \in A$.

which hold since $T_A = 0$ whenever one starts from $Z_0 \in A$.

Proof. For all $k \in \mathbb{S} \setminus A$ we have $T_A \geqslant 1$ given that $Z_0 = k$, hence we can write

$$\begin{aligned} g_l(k) &= \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = k) \\ &= \sum_{m\in\mathbb{S}} \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_1 = m \text{ and } Z_0 = k)\mathbb{P}(Z_1 = m \mid Z_0 = k) \\ &= \sum_{m\in\mathbb{S}} \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_1 = m)\mathbb{P}(Z_1 = m \mid Z_0 = k) \\ &= \sum_{m\in\mathbb{S}} P_{k,m}\mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_1 = m) \\ &= \sum_{m\in\mathbb{S}} P_{k,m}\mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = m) \\ &= \sum_{m\in\mathbb{S}} P_{k,m} g_l(m), \qquad k \in \mathbb{S} \setminus A, \quad l \in A, \end{aligned}$$

where the relation

$$\mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_1 = m) = \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = m)$$

follows from the fact that the probability of ruin does not depend on the initial time the counter is started. □

Equation (5.1.3) can be rewritten in matrix form as

$$g_l = P g_l, \qquad l \in A, \qquad (5.1.4)$$

where g_l is a column vector, *i.e.*

$$\begin{bmatrix} g_l(0) \\ \vdots \\ g_l(N) \end{bmatrix} = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix} \times \begin{bmatrix} g_l(0) \\ \vdots \\ g_l(N) \end{bmatrix}, \qquad l \in A,$$

subject to the boundary condition

$$g_l(k) = \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = k) = \mathbb{1}_{\{l\}}(k) = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

for all $k, l \in A$. See *e.g.* Theorem 3.4 page 40 of Karlin and Taylor (1981) for a uniqueness result for the solution of such equations, and Theorem 2.1 in Goldberg (1986) for the uniqueness of solutions to difference equations in general.

In addition, the hitting probabilities

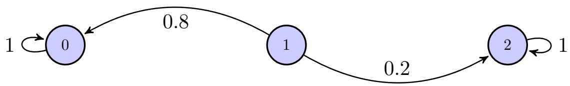
$$g_l(k) = \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = k), \qquad k \in \mathbb{S},$$

satisfy the condition

$$\begin{aligned} 1 &= \mathbb{P}(T_A = \infty \mid Z_0 = k) + \sum_{l \in A} \mathbb{P}(Z_{T_A} = l \text{ and } T_A < \infty \mid Z_0 = k) \\ &= \mathbb{P}(T_A = \infty \mid Z_0 = k) + \sum_{l \in A} g_l(k), \end{aligned} \tag{5.1.5}$$

for all $k \in \mathbb{S}$. Note that we may have $\mathbb{P}(T_A = \infty \mid Z_0 = k) > 0$, for example in the following chain with $A = \{0\}$ and $k = 1$ we have

$$\mathbb{P}(T_0 = \infty \mid Z_0 = 1) = 0.2.$$



More generally, if $f : A \longrightarrow \mathbb{R}$ is a function on the domain A and $\mathbb{P}(T_A < \infty \mid Z_0 = k) = 1$, letting

$$g_A(k) := \mathbb{E}\big[f(Z_{T_A}) \,\big|\, Z_0 = k\big]$$

$$= \sum_{l \in A} f(l) \mathbb{P}(Z_{T_A} = l \mid Z_0 = k), \qquad k \in \mathbb{S},$$

by linearity we find the *Dirichlet problem*

$$(I - P) g_A = 0,$$

subject to the boundary condition

$$g_A(k) = f(k), \qquad k \in A,$$

see, *e.g.*, Theorem 5.3 in Privault (2008) for a continuous-time analog.

The next lemma will be used in Chapter 8 on branching Processes.

Lemma 5.2. *Assume that state* $\textcircled{j} \in \mathbb{S}$ *is absorbing. Then for all* $\textcircled{i} \in \mathbb{S}$ *we have*

$$\mathbb{P}(T_j < \infty \mid Z_0 = i) = \lim_{n \to \infty} \mathbb{P}(Z_n = j \mid Z_0 = i).$$

Proof. We have

$$\{T_j < \infty\} = \bigcup_{n \geqslant 1} \{Z_n = j\},$$

because the finiteness of T_j means that Z_n becomes equal to $\textcircled{j}$ for some $n \in \mathbb{N}$. In addition, since $\textcircled{j} \in \mathbb{S}$ is absorbing it holds that

$$\{Z_n = j\} \subset \{Z_{n+1} = j\}, \qquad n \geqslant 0,$$

hence given that $\{Z_0 = i\}$, by (1.2.5) we have

$$\alpha_1 = \mathbb{P}(T_j < \infty \mid Z_0 = i) = \mathbb{P}\left(\bigcup_{n \geqslant 1} \{Z_n = j\} \,\middle|\, Z_0 = i \right) \tag{5.1.6}$$
$$= \mathbb{P}\left(\lim_{n \to \infty} \{Z_n = j\} \,\middle|\, Z_0 = i \right) = \lim_{n \to \infty} \mathbb{P}(Z_n = j \mid Z_0 = i).$$

□

Block triangular transition matrices

Assume now that the state space is $\mathbb{S} = \{0, 1, \ldots, N\}$ and the transition matrix P has the form

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}, \tag{5.1.7}$$

where Q is a square $(r+1)\times(r+1)$ matrix, R is a $(r+1)\times(N-r)$ matrix, and

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

is the $(N-r)\times(N-r)$ identity matrix, in which case the states in $\{r+1, r+2, \ldots, N\}$ are *absorbing*.

If the set $A := \{r+1, r+2, \ldots, N\}$ is made of the absorbing states of the chain, we have the boundary conditions

$$g_l(m) = \mathbb{1}_{\{m=l\}}, \qquad l = 0,1,\ldots,N, \quad m = r+1, r+2, \ldots, N, \tag{5.1.8}$$

hence the equation (5.1.3) can be rewritten as

$$\begin{aligned} g_l(k) &= \sum_{m=0}^{N} P_{k,m} g_l(m) \\ &= \sum_{m=0}^{r} P_{k,m} g_l(m) + \sum_{m=r+1}^{N} P_{k,m} g_l(m) \\ &= \sum_{m=0}^{r} P_{k,m} g_l(m) + P_{k,l} \\ &= \sum_{m=0}^{r} Q_{k,m} g_l(m) + R_{k,l}, \qquad k = 0,1,\ldots,r, \quad l = r+1,\ldots,N, \end{aligned}$$

from (5.1.8) and since $P_{k,l} = R_{k,l}$, $k = 0,1,\ldots,r$, $l = r+1,\ldots,N$. Hence we have

$$g_l(k) = \sum_{m=0}^{r} Q_{k,m} g_l(m) + R_{k,l}, \qquad k = 0,1,\ldots,r, \quad l = r+1,\ldots,N.$$

Remark 5.3. *In the case of the two-state Markov chain with transition matrix (4.5.1) and $A = \{0\}$, we simply find $g_0(0) = 1$ and*

$$g_0(1) = b + (1-b) \times g_0(1),$$

hence $g_0(1) = 1$ if $b > 0$ and $g_0(1) = 0$ if $b = 0$.

Example

Consider a Markov chain on the state space $\mathbb{S} = \{0,1,2,3\}$ with transition matrix of the form

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a & b & c & d \\ \alpha & \beta & \gamma & \eta \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.1.9}$$

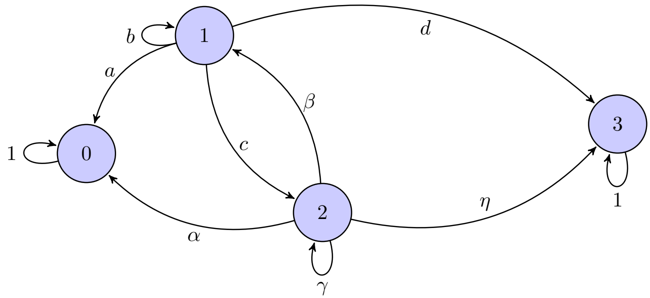
Let $A = \{0, 3\}$ denote the absorbing states of the chain, and let

$$T_{0,3} = \inf\{n \geqslant 0 \ : \ Z_n = 0 \text{ or } Z_n = 3\}$$

and compute the probabilities

$$g_0(k) = \mathbb{P}(Z_{T_{0,3}} = 0 \mid Z_0 = k)$$

of hitting state ⓪ first within $\{0, 3\}$ starting from $k = 0, 1, 2, 3$. The chain has the following graph:



(5.1.10)

Noting that ⓪ and ③ are absorbing states, and writing the relevant rows of the first step analysis matrix equation $g = Pg$, we have

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = a \times 1 + b g_0(1) + c g_0(2) + d \times 0 \\ g_0(2) = \alpha \times 1 + \beta g_0(1) + \gamma g_0(2) + \eta \times 0 \\ g_0(3) = 0, \end{cases}$$

i.e.

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = a + b g_0(1) + c g_0(2) \\ g_0(2) = \alpha + \beta g_0(1) + \gamma g_0(2) \\ g_0(3) = 0, \end{cases}$$

which has for solution

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = \dfrac{c\alpha + a(1-\gamma)}{(1-b)(1-\gamma) - c\beta} \\ g_0(2) = \dfrac{a\beta + \alpha(1-b)}{(1-b)(1-\gamma) - c\beta} \\ g_0(3) = 0. \end{cases} \tag{5.1.11}$$

We have $g_l(0) = g_l(3) = 0$ for $l = 1, 2$, and by a similar analysis, letting

$$g_3(k) := \mathbb{P}(Z_{T_{0,3}} = 3 \mid Z_0 = k), \qquad k = 0, 1, 2, 3,$$

we find

$$\begin{cases} g_3(0) = 0 \\ g_3(1) = \dfrac{c\eta + d(1-\gamma)}{(1-b)(1-\gamma) - c\beta} \\ g_3(2) = \dfrac{\beta d + \eta(1-b)}{(1-b)(1-\gamma) - c\beta} \\ g_3(3) = 1, \end{cases}$$

and we note that

$$g_0(1) + g_3(1) = \frac{c\alpha + a(1-\gamma)}{(1-b)(1-\gamma) - c\beta} + \frac{c\eta + d(1-\gamma)}{(1-b)(1-\gamma) - c\beta} = 1,$$

since $\alpha + \eta = 1 - \gamma - \beta$ and $a + d = 1 - b - c$, and similarly

$$g_0(2) + g_3(2) = \frac{a\beta + \alpha(1-b)}{(1-b)(1-\gamma) - c\beta} + \frac{\beta d + \eta(1-b)}{(1-b)(1-\gamma) - c\beta} = 1.$$

We also check that in case $a = d$ and $\alpha = \eta$ we have

$$g_0(1) = \frac{c\alpha + a(\beta + 2\alpha)}{(c + 2a)(\beta + 2\alpha) - c\beta} = \frac{c\alpha + a\beta + 2a\alpha}{2c\alpha + 2a\beta + 4a\alpha} = g_0(2) = \frac{1}{2}, \tag{5.1.12}$$

and

$$g_0(1) = g_3(1) = g_0(2) = g_3(2) = \frac{1}{2}.$$

Note that, letting

$$T_0 := \inf\{n \geqslant 0 \ : \ Z_n = 0\} \quad \text{and} \quad T_3 := \inf\{n \geqslant 0 \ : \ Z_n = 0\},$$

we also have

$$g_0(k) = \mathbb{P}(Z_{T_{0,3}} = 0 \mid Z_0 = k) = \mathbb{P}(T_0 < \infty \mid Z_0 = k)$$

and

$$g_3(k) = \mathbb{P}(Z_{T_{0,3}} = 3 \mid Z_0 = k) = \mathbb{P}(T_3 < \infty \mid Z_0 = k)$$

$k = 0, 1, 2, 3$.

5.2 Mean Hitting and Absorption Times

We are now interested in the mean hitting time

$$h_A(k) := \mathbb{E}[T_A \mid Z_0 = k]$$

it takes for the chain to hit the set $A \subset \mathbb{S}$ starting from a state $k \in \mathbb{S}$. In case the set A is absorbing we refer to $h_A(k)$ as the *mean absorption time* into A starting from the state $\textcircled{k}$. Clearly, since $T_A = 0$ whenever $Z_0 = k \in A$, we have

$$h_A(k) = 0, \qquad \text{for all } k \in A.$$

Proposition 5.4. *The mean hitting times*

$$h_A(k) := \mathbb{E}[T_A \mid Z_0 = k], \qquad k \in \mathbb{S},$$

satisfy the equations

$$h_A(k) = 1 + \sum_{l \in \mathbb{S}} P_{k,l} h_A(l) = 1 + \sum_{l \in \mathbb{S} \setminus A} P_{k,l} h_A(l), \quad k \in \mathbb{S} \setminus A, \quad (5.2.1)$$

subject to the boundary conditions

$$h_A(k) = \mathbb{E}[T_A \mid Z_0 = k] = 0, \qquad k \in A.$$

Proof. For all $k \in \mathbb{S} \setminus A$, by first step analysis using the *law of total expectation* applied to the probability measure $\mathbb{P}(\cdot \mid Z_0 = l)$, and the Markov property we have

$$\begin{aligned}
h_A(k) &= \mathbb{E}[T_A \mid Z_0 = k] \\
&= \sum_{l \in \mathbb{S}} \mathbb{E}\big[T_A \mathbb{1}_{\{Z_1 = l\}} \mid Z_0 = k\big] \\
&= \frac{1}{\mathbb{P}(Z_0 = k)} \sum_{l \in \mathbb{S}} \mathbb{E}\big[T_A \mathbb{1}_{\{Z_1 = l\}} \mathbb{1}_{\{Z_0 = k\}}\big] \\
&= \frac{1}{\mathbb{P}(Z_0 = k)} \sum_{l \in \mathbb{S}} \mathbb{E}\big[T_A \mathbb{1}_{\{Z_1 = l \text{ and } Z_0 = k\}}\big] \\
&= \sum_{l \in \mathbb{S}} \mathbb{E}\big[T_A \mid Z_1 = l \text{ and } Z_0 = k\big] \frac{\mathbb{P}(Z_1 = l \text{ and } Z_0 = k)}{\mathbb{P}(Z_0 = k)} \\
&= \sum_{l \in \mathbb{S}} \mathbb{E}[T_A \mid Z_1 = l \text{ and } Z_0 = k] \mathbb{P}(Z_1 = l \mid Z_0 = k)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l\in\mathbb{S}} \mathbb{E}[1+T_A \mid Z_0=l]\mathbb{P}(Z_1=l \mid Z_0=k) \\
&= \sum_{l\in\mathbb{S}} (1+\mathbb{E}[T_A \mid Z_0=l])\mathbb{P}(Z_1=l \mid Z_0=k) \\
&= \sum_{l\in\mathbb{S}} \mathbb{P}(Z_1=l \mid Z_0=k) + \sum_{l\in\mathbb{S}} \mathbb{P}(Z_1=l \mid Z_0=k)\mathbb{E}[T_A \mid Z_0=l] \\
&= 1+\sum_{l\in\mathbb{S}} \mathbb{P}(Z_1=l \mid Z_0=k)\mathbb{E}[T_A \mid Z_0=l] \\
&= 1+\sum_{l\in\mathbb{S}} P_{k,l}h_A(l), \qquad k\in\mathbb{S}\setminus A,
\end{aligned}
$$

with the relation

$$
\mathbb{E}[T_A \mid Z_1=l,\ Z_0=k] = 1+\mathbb{E}[T_A \mid Z_0=l].
$$

Hence we have

$$
h_A(k) = 1+\sum_{l\in\mathbb{S}} P_{k,l}h_A(l), \quad k\in\mathbb{S}\setminus A, \tag{5.2.2}
$$

subject to the boundary conditions

$$
h_A(k) = \mathbb{E}[T_A \mid Z_0=k] = 0, \qquad k\in A, \tag{5.2.3}
$$

the Condition (5.2.3) implies that (5.2.2) becomes

$$
h_A(k) = 1+\sum_{l\in\mathbb{S}\setminus A} P_{k,l}h_A(l), \qquad k\in\mathbb{S}\setminus A.
$$

□

The equations (5.2.1) can be rewritten in matrix form as

$$
h_A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + Ph_A,
$$

by considering only the rows with index $k\in A^c=\mathbb{S}\setminus A$, subject to the boundary conditions

$$
h_A(k)=0, \qquad k\in A.
$$

Block triangular transition matrices

When the transition matrix P has the form (5.1.7) and $A = \{r+1, r+2, \ldots, N\}$, Equation (5.2.2) rewrites as

$$\begin{aligned} h_A(k) &= 1 + \sum_{l=0}^{N} P_{k,l} h_A(l) \\ &= 1 + \sum_{l=0}^{r} P_{k,l} h_A(l) + \sum_{l=r+1}^{N} P_{k,l} h_A(l) \\ &= 1 + \sum_{l=0}^{r} P_{k,l} h_A(l), \qquad 0 \leqslant k \leqslant r, \end{aligned}$$

since $h_A(l) = 0$, $l = r+1, r+2, \ldots, N$, *i.e.*

$$h_A(k) = 1 + \sum_{l=0}^{r} P_{k,l} h_A(l), \qquad 0 \leqslant k \leqslant r,$$

with $h_A(k) = 0$, $k = r+1, \ldots, n$.

Two-state chain

In the case of the two-state Markov chain with transition matrix (4.5.1) and $A = \{0\}$, we simply find $h_{\{0\}}(0) = h_{\{1\}}(1) = 0$ and

$$h_{\{0\}}(1) = b \times 1 + (1-b)(1 + h_{\{0\}}(1)) = 1 + (1-b) h_{\{0\}}(1), \qquad (5.2.4)$$

with solution

$$h_{\{0\}}(1) = b \sum_{k \geqslant 1} k(1-b)^{k-1} = \frac{1}{b},$$

and similarly we find

$$h_{\{1\}}(0) = a \sum_{k \geqslant 1} k(1-a)^{k-1} = \frac{1}{a},$$

cf. also (5.3.3) below.

Utility functionals

The above can be generalized to derive an equation for an expectation of the form

$$h_A(k) := \mathbb{E}\left[\sum_{i=0}^{T_A} f(Z_i) \;\Big|\; Z_0 = k\right], \qquad k = 0, 1, \ldots, N,$$

where $f(\cdot)$ is a given utility function, as follows:

$$\begin{aligned} h_A(k) &= \mathbb{E}\left[\sum_{i=0}^{T_A} f(Z_i) \;\Big|\; Z_0 = k\right] \\ &= \sum_{l=0}^{N} P_{k,l}\left(f(k) + \mathbb{E}\left[\sum_{i=1}^{T_A} f(Z_i) \Big| Z_1 = l\right]\right) \\ &= \sum_{l=0}^{N} P_{k,l} f(k) + \sum_{l=0}^{N} P_{k,l}\mathbb{E}\left[\sum_{i=1}^{T_A} f(Z_i) \Big| Z_1 = l\right] \\ &= f(k)\sum_{l=0}^{N} P_{k,l} + \sum_{l=0}^{r} P_{k,l}\mathbb{E}\left[\sum_{i=0}^{T_A} f(Z_i) \Big| Z_0 = l\right] \\ &= f(k) + \sum_{l=0}^{r} P_{k,l} h_A(l), \qquad k \in A^c := \{0, 1, \ldots, r\}, \end{aligned}$$

with $A := \{r+1, \ldots, N\}$, hence

$$h_A(k) = f(k) + \sum_{l=0}^{r} P_{k,l} h_A(l), \qquad k \in A^c = \{0, 1, \ldots, r\},$$

which can be rewritten as

$$\begin{bmatrix} h_A(0) \\ \vdots \\ h_A(r) \end{bmatrix} = \begin{bmatrix} f(0) \\ \vdots \\ f(r) \end{bmatrix} + \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,r} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,r} \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots & P_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{r,0} & P_{r,1} & P_{r,2} & \cdots & P_{r,r} \end{bmatrix} \times \begin{bmatrix} h_A(0) \\ \vdots \\ h_A(r) \end{bmatrix},$$

with the boundary condition

$$h_A(k) = f(k), \qquad k \in A = \{r+1, \ldots, n\},$$

see also Exercise 5.28 and Problem 5.39.

For example:

- When $f = \mathbb{1}_{A^c} = \mathbb{1}_{\{0,1,\ldots,r\}}$ is the indicator function over the set A^c, *i.e.*

$$f(Z_i) = \mathbb{1}_{A^c}(Z_i) = \begin{cases} 1 & \text{if } Z_i \notin A, \\ \\ 0 & \text{if } Z_i \in A, \end{cases}$$

the quantity $h_A(k)$ coincides with the mean hitting time of the set A starting from the state (k). In particular, when $A = \{m\}$ this recovers the equation

$$h_{\{m\}}(k) = 1 + \sum_{\substack{l \in \mathbb{S} \\ l \neq m}} P_{k,l} h_{\{m\}}(l), \qquad k \in \mathbb{S} \setminus \{m\}, \tag{5.2.5}$$

with $h_{\{m\}}(m) = 0$.

- When f is the indicator function $f = \mathbb{1}_{\{l\}}$, *i.e.*

$$f(Z_i) = \mathbb{1}_{\{l\}}(Z_i) = \begin{cases} 1 & \text{if } Z_i = l, \\ \\ 0 & \text{if } Z_i \neq l, \end{cases}$$

with $l \in A^c$, the quantity $h_A(k)$ counts the mean number of visits to state (l) starting from (k) before hitting the set A.

- See Exercise 5.26, Exercise 5.28, and also Problem 5.34 for a complete solution in case $f(k) := k$ and $(Z_k)_{k \geqslant 0}$ is the gambling process of Chapter 2.

Poisson equation

Taking

$$h_A(k) := \mathbb{E}\left[\sum_{i=0}^{T_A - 1} f(Z_i) \,\middle|\, Z_0 = k \right], \qquad k = 0, 1, \ldots, N,$$

leads to the same first step analysis equation

$$h_A(k) = f(k) + \sum_{m=0}^{r} P_{k,m} h_A(m), \qquad k \in A^c := \{0, 1, \ldots, r\},$$

or

$$(I - [P_{k,m}]_{k,m=0,1,\ldots,r}) \times h_A = f,$$

with the boundary condition

$$h_A(k) = 0, \qquad k \in A = \{r+1, \ldots, n\},$$

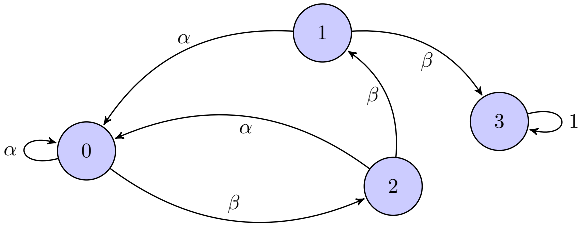
see *e.g.* Theorem 5.5 in Privault (2008) for a continuous-time analog.

Example

Consider a Markov chain whose transition probability matrix is given by

$$P = \left[P_{i,j} \right]_{0 \leqslant i,j \leqslant 3} = \begin{bmatrix} \alpha & 0 & \beta & 0 \\ \alpha & 0 & 0 & \beta \\ \alpha & \beta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\alpha, \beta \geqslant 0$ and $\alpha + \beta = 1$. Taking $A := \{3\}$, determine the mean time it takes to reach state ③ starting from state ⓪. We observe that state ③ is absorbing:



Let

$$h_3(k) := \mathbb{E}[T_3 \mid Z_0 = k]$$

denote the mean (hitting) time needed to reach state ③, after starting from state $k = 0, 1, 2, 3$. We get

$$\begin{cases} h_3(0) = \alpha(1 + h_3(0)) + \beta(1 + h_3(2)) = 1 + \alpha h_3(0) + \beta h_3(2) \\ \\ h_3(1) = \alpha(1 + h_3(0)) + \beta(1 + h_3(3)) = 1 + \alpha h_3(0) \\ \\ h_3(2) = \alpha(1 + h_3(0)) + \beta(1 + h_3(1)) = 1 + \alpha h_3(0) + \beta h_3(1) \\ \\ h_3(3) = 0, \end{cases}$$

which, using the relation $\alpha = 1 - \beta$, yields

$$h_3(3) = 0, \qquad h_3(1) = \frac{1}{\beta^3}, \qquad h_3(2) = \frac{1+\beta}{\beta^3}, \qquad h_3(0) = \frac{1+\beta+\beta^2}{\beta^3}.$$

Since state ③ can only be reached from state ① with probability β, it is natural that the hitting times go to infinity as β goes to zero. We also check that $h_3(3) < h_3(1) < h_3(2) < h_3(0)$, as can be expected from the above graph. In addition, $(h_3(1), h_3(2), h_3(0))$ converge to $(1, 2, 3)$ as β goes to 1, as can be expected.

5.3 First Return Times

Consider now the first *return* time T_j^r to state $j \in \mathbb{S}$, defined by

$$T_j^r := \inf\{n \geqslant 1 \ : \ Z_n = j\},$$

with

$$T_j^r = +\infty \quad \text{if } Z_n \neq j \text{ for all } n \geqslant 1.$$

Note that in contrast with the definition (5.1.1) of the hitting time T_j, the infimum is taken here for $n \geqslant 1$ as it takes at least one step out of the initial state in order to *return* to state $\textcircled{j}$. Nevertheless we have $T_j = T_j^r$ if the chain is started from a state $\textcircled{i}$ different from $\textcircled{j}$.

We denote by

$$\mu_j(i) := \mathbb{E}[T_j^r \mid Z_0 = i] \geqslant 1$$

the *mean return time* to state $j \in \mathbb{S}$ after starting from state $i \in \mathbb{S}$.

Mean return times can also be computed by first step analysis. We have

$$\begin{aligned}
\mu_j(i) &= \mathbb{E}[T_j^r \mid Z_0 = i] \\
&= 1 \times \mathbb{P}(Z_1 = j \mid Z_0 = i) \\
&\quad + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} \mathbb{P}(Z_1 = l \mid Z_0 = i)(1 + \mathbb{E}[T_j^r \mid Z_0 = l]) \\
&= P_{i,j} + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l}(1 + \mu_j(l)) \\
&= P_{i,j} + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l} + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l}\mu_j(l) \\
&= \sum_{l \in \mathbb{S}} P_{i,l} + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l}\mu_j(l) \\
&= 1 + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l}\mu_j(l),
\end{aligned}$$

hence

$$\mu_j(i) = 1 + \sum_{\substack{l \in \mathbb{S} \\ l \neq j}} P_{i,l}\mu_j(l), \qquad i, j \in \mathbb{S}. \tag{5.3.1}$$

See *e.g.* Theorem 5.9 page 49 of Karlin and Taylor (1981) for a uniqueness result for the solution of such equations.

Hitting times *vs.* return times

Note that the mean return time equation in (5.3.1) does not include any boundary condition, in contrast with the mean hitting time equation (5.2.5) in Section 5.2. In addition, the time T_i^r to return to state $\textcircled{i}$ is always at least one by construction, hence $\mu_i(i) \geqslant 1$ and cannot vanish, while we always have $h_i(i) = 0$ as boundary condition, $i \in \mathbb{S}$. On the other hand, for $i \neq j$ we have by definition

$$\mu_i(j) = \mathbb{E}[T_i^r \mid Z_0 = j] = \mathbb{E}[T_i \mid Z_0 = j] = h_i(j),$$

and for $i = j$ the mean return time $\mu_j(j)$ can be computed from the hitting times $h_j(l)$, $l \neq j$, by first step analysis as

$$\begin{aligned}
\mu_j(j) &= \sum_{l \in S} P_{j,l}(1 + h_j(l)) \\
&= P_{j,j} + \sum_{l \neq j} P_{j,l}(1 + h_j(l)) \\
&= \sum_{l \in S} P_{j,l} + \sum_{l \neq j} P_{j,l} h_j(l) \\
&= 1 + \sum_{l \neq j} P_{j,l} h_j(l), \qquad j \in \mathbb{S}, \qquad (5.3.2)
\end{aligned}$$

which is in agreement with (5.3.1) when $i = j$.

In practice we may prefer to compute first the hitting times $h_i(j) = 0$ under the boundary conditions $h_i(i) = 0$, and then to recover the return time $\mu_i(i)$ from (5.3.2), $i, j \in \mathbb{S}$.

Examples

i) Mean return times for the two-state Markov chain.

The mean return time $\mu_0(i) = \mathbb{E}[T_0^r \mid Z_0 = i]$ to state $\textcircled{0}$ starting from state $\textcircled{i} \in \{0, 1\}$ satisfies

$$\begin{cases} \mu_0(0) = (1-a) \times 1 + a(1 + \mu_0(1)) = 1 + a\mu_0(1) \\ \\ \mu_0(1) = b \times 1 + (1-b)(1 + \mu_0(1)) = 1 + (1-b)\mu_0(1) \end{cases}$$

which yields

$$\mu_0(0) = 1 + \frac{a}{b} \quad \text{and} \quad \mu_0(1) = h_0(1) = \frac{1}{b}, \qquad (5.3.3)$$

cf. also (5.2.4) above for the computation of $\mu_0(1) = h_0(1) = 1/b$ as a mean hitting time. In the two-state case, the distribution of T_0^r given $Z_0 = 0$ is given by

$$f_{0,0}^{(n)} := \mathbb{P}(T_0^r = n \mid Z_0 = 0) = \begin{cases} 0 & \text{if } n = 0, \\ 1-a & \text{if } n = 1, \\ ab(1-b)^{n-2} & \text{if } n \geqslant 2, \end{cases} \tag{5.3.4}$$

hence (5.3.3) can be directly recovered as*

$$\begin{aligned} \mu_0(0) &= \mathbb{E}[T_0^r \mid Z_0 = 0] \\ &= \sum_{n\geqslant 0} n\mathbb{P}(T_0^r = n \mid Z_0 = 0) \\ &= \sum_{n\geqslant 0} n f_{0,0}^{(n)} \\ &= 1 - a + ab\sum_{n\geqslant 2} n(1-b)^{n-2} \\ &= 1 - a + ab\sum_{n\geqslant 0} (n+2)(1-b)^{n} \\ &= 1 - a + ab(1-b)\sum_{n\geqslant 0} n(1-b)^{n-1} + 2ab\sum_{n\geqslant 0}(1-b)^n \\ &= \frac{a+b}{b} = 1 + \frac{a}{b}, \end{aligned} \tag{5.3.5}$$

where we used the identity (A.4). Similarly, we check that

$$\begin{cases} \mu_1(0) = 1 + (1-a)\mu_1(0) \\ \mu_1(1) = 1 + b\mu_1(0), \end{cases}$$

which yields

$$\mu_1(0) = h_1(0) = \frac{1}{a} \quad \text{and} \quad \mu_1(1) = 1 + \frac{b}{a},$$

and can be directly recovered by

$$\mu_1(1) = 1 - b + ab\sum_{n\geqslant 0}(n+2)(1-a)^n = \frac{a+b}{a} = 1 + \frac{b}{a}, \tag{5.3.6}$$

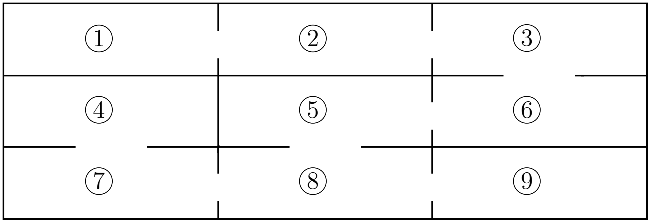
as in (5.3.3) and (5.3.5) above, by swapping a with b and state ⓪ with state ①.

ii) Maze problem.

* We are using the identities $\sum_{k\geqslant 0} r^k = (1-r)^{-1}$ and $\sum_{k\geqslant 1} kr^{k-1} = (1-r)^{-2}$, cf. (A.3) and (A.4).

Mazes provide natural examples of Markovian systems as their users tend rely on their current positions and to forget past information. More generally, Markovian systems can be used as an approximation of a non-Markovian reality.

Consider a fish placed in an aquarium with 9 compartments:



(5.3.7)

The fish moves randomly: at each time step it changes compartments and if it finds $k \geqslant 1$ exit doors from one compartment, it will choose one of them with probability $1/k$, *i.e.* the transition matrix is

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Question: Find the average time to come back to state ① after starting from state ①.

Answer: Letting

$$T_l^r = \inf\{n \geqslant 1 \ : \ Z_n = l\}$$

denote the first return time to state Ⓛ, and defining

$$\mu_1(k) := \mathbb{E}[T_1^r \mid Z_0 = k]$$

the mean return time to state ① starting from Ⓚ, we have

$$\begin{cases} \mu_1(1) = 1 + \mu_1(2) \\ \mu_1(2) = \frac{1}{2}(1+0) + \frac{1}{2}(1+\mu_1(3)) = 1 + \frac{1}{2}\mu_1(3) \\ \mu_1(3) = \frac{1}{2}(1+\mu_1(2)) + \frac{1}{2}(1+\mu_1(6)) = 1 + \frac{1}{2}\mu_1(2) + \frac{1}{2}\mu_1(6) \\ \mu_1(4) = 1 + \mu_1(7) \\ \mu_1(5) = \frac{1}{2}(1+\mu_1(8)) + \frac{1}{2}(1+\mu_1(6)) = 1 + \frac{1}{2}\mu_1(8) + \frac{1}{2}\mu_1(6) \\ \mu_1(6) = \frac{1}{2}(1+\mu_1(3)) + \frac{1}{2}(1+\mu_1(5)) = 1 + \frac{1}{2}\mu_1(3) + \frac{1}{2}\mu_1(5) \\ \mu_1(7) = 1 + \frac{1}{2}\mu_1(4) + \frac{1}{2}\mu_1(8) = \frac{1}{2}(1+\mu_1(4)) + \frac{1}{2}(1+\mu_1(8)) \\ \mu_1(8) = \frac{1}{3}(1+\mu_1(7)) + \frac{1}{3}(1+\mu_1(5)) + \frac{1}{3}(1+\mu_1(9)) \\ \qquad = 1 + \frac{1}{3}\left(\mu_1(7) + \mu_1(5) + \mu_1(9)\right) \\ \mu_1(9) = 1 + \mu_1(8), \end{cases}$$

or

$$\begin{cases} \mu_1(1) = 1 + \mu_1(2), \quad \mu_1(2) = 1 + \frac{1}{2}\mu_1(3), \quad \mu_1(3) = 2 + \frac{2}{3}\mu_1(6), \\ \mu_1(4) = 1 + \mu_1(7), \quad \mu_1(5) = 1 + \frac{1}{2}\mu_1(8) + \frac{1}{2}\mu_1(6), \quad 0 = 30 + 3\mu_1(8) - 5\mu_1(6), \\ \mu_1(7) = 3 + \mu_1(8), \quad 0 = 80 + 5\mu_1(6) - 5\mu_1(8), \quad \mu_1(9) = 1 + \mu_1(8), \end{cases}$$

which yields

$$\begin{aligned} &\mu_1(1) = 16,\ \mu_1(2) = 15,\ \mu_1(3) = 28,\ \mu_1(4) = 59,\ \mu_1(5) = 48, \\ &\mu_1(6) = 39,\ \mu_1(7) = 58,\ \mu_1(8) = 55,\ \mu_1(9) = 56. \end{aligned} \tag{5.3.8}$$

Consequently, it takes on average 16 steps to come back to ① starting from ①, and 59 steps to reach ① starting from ④. These data are illustrated in the following picture in which the numbers represent the average time it takes to return to ① starting from a given state.

$\mu_1(1) = 16$	$\mu_1(2) = 15$	$\mu_1(3) = 28$
$\mu_1(4) = 59$	$\mu_1(5) = 48$	$\mu_1(6) = 39$
$\mu_1(7) = 58$	$\mu_1(8) = 55$	$\mu_1(9) = 56$

The next Figure 5.1 represents the mean return times to state (0) according to the initial state on the maze (5.3.7).

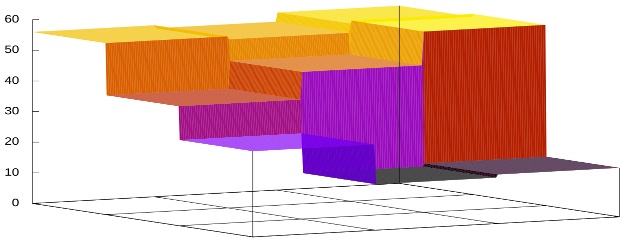


Fig. 5.1: Mean return times to state 0 on the maze (5.3.7).

5.4 Mean Number of Returns

Return probabilities

In what follows, we let

$$p_{ij} = \mathbb{P}(T_j^r < \infty \mid Z_0 = i) = \mathbb{P}(Z_n = j \text{ for some } n \geqslant 1 \mid Z_0 = i), \quad i, j \in \mathbb{S},$$

denote the probability of return to state (j) in finite time* starting from state (i). The probability p_{ii} of return to state (i) within a finite time starting from state (i) can be computed as follows:

$$\begin{aligned} p_{ii} &= \mathbb{P}(Z_n = i \text{ for some } n \geqslant 1 \mid Z_0 = i) \\ &= \mathbb{P}\Big(\bigcup_{n \geqslant 1} \{Z_n = i\} \Bigm| Z_0 = i \Big) \end{aligned}$$

* When $(i) \neq (j)$, p_{ij} is the probability of *visiting* state (j) in finite time after starting from state (i).

$$= \sum_{n\geqslant 1} \mathbb{P}(Z_n = i, Z_{n-1} \neq i, \ldots, Z_1 \neq i \mid Z_0 = i)$$
$$= \sum_{n\geqslant 1} f_{i,i}^{(n)}, \tag{5.4.1}$$

where

$$f_{i,j}^{(n)} := \mathbb{P}(T_j^r = n \mid Z_0 = i) = \mathbb{P}(Z_n = j, Z_{n-1} \neq j, \ldots, Z_1 \neq j \mid Z_0 = i),$$

$i, j \in \mathbb{S}$, is the probability distribution of T_j^r given that $Z_0 = i$, with

$$f_{i,i}^{(0)} = \mathbb{P}(T_i^r = 0 \mid Z_0 = i) = 0.$$

Note that we have

$$f_{i,i}^{(1)} = \mathbb{P}(Z_1 = i \mid Z_0 = i) = P_{i,i}, \qquad i \in \mathbb{S}.$$

Convolution equation

By conditioning on the first return time $k \geqslant 1$, the return time probability distribution $f_{i,i}^{(k)} = \mathbb{P}(T_i^r = k \mid Z_0 = i)$ satisfies the convolution equation

$$\begin{aligned}
[P^n]_{i,i} &= \mathbb{P}(Z_n = i \mid Z_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(Z_k = i, Z_{k-1} \neq i, \ldots, Z_1 \neq i \mid Z_0 = i)\mathbb{P}(Z_n = i \mid Z_k = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(Z_k = i, Z_{k-1} \neq i, \ldots, Z_1 \neq i \mid Z_0 = i)\mathbb{P}(Z_{n-k} = i \mid Z_0 = i) \\
&= \sum_{k=1}^{n} f_{i,i}^{(k)} [P^{n-k}]_{i,i},
\end{aligned}$$

which extends the convolution equation (3.4.7) from random walks to the more general setting of Markov chains.

The return probabilities p_{ij} will be used below to compute the average number of returns to a given state, and the distribution $f_{i,j}^{(k)}$, $k \geqslant 1$, of T_j^r given that $Z_0 = i$ will be useful in Section 6.4 on positive and null recurrence.

Number of returns

Let

$$R_j := \sum_{n\geqslant 1} \mathbb{1}_{\{Z_n=j\}} \tag{5.4.2}$$

denote the number of returns* to state $\textcircled{j}$ by the chain $(Z_n)_{n\geqslant 0}$. The next proposition shows that, given $\{Z_0 = i\}$, R_j has a zero-modified geometric distribution with initial mass $1 - p_{ij}$.

Proposition 5.5. *The probability distribution of the number of returns R_j to state j given that $\{Z_0 = i\}$ is given by*

$$\mathbb{P}(R_j = m \mid Z_0 = i) = \begin{cases} 1 - p_{ij}, & m = 0, \\ \\ p_{ij} \times (p_{jj})^{m-1} \times (1 - p_{jj}), & m \geqslant 1, \end{cases}$$

Proof. When the chain never visits state $\textcircled{j}$ starting from $Z_0 = i$ we have $R_j = 0$, and this happens with probability

$$\begin{aligned} \mathbb{P}(R_j = 0 \mid Z_0 = i) &= \mathbb{P}(T_j^r = \infty \mid Z_0 = i) \\ &= 1 - \mathbb{P}(T_j^r < \infty \mid Z_0 = i) \\ &= 1 - p_{ij}. \end{aligned}$$

Next, when the chain $(Z_n)_{n\geqslant 0}$ makes a number $R_j = m \geqslant 1$ of visits to state $\textcircled{j}$ starting from state $\textcircled{i}$, it makes a first visit to state $\textcircled{j}$ with probability $\underline{p_{ij}}$ and then makes $\underline{m-1}$ returns to state $\textcircled{j}$, each with probability $\underline{p_{jj}}$. After those m visits, it never returns to state $\textcircled{j}$, and this event occurs with probability $\underline{1 - p_{jj}}$. Hence, given that $\{Z_0 = i\}$ we have

$$\mathbb{P}(R_j = m \mid Z_0 = i) = \begin{cases} p_{ij} \times (p_{jj})^{m-1} \times (1 - p_{jj}), & m \geqslant 1, \\ \\ 1 - p_{ij}, & m = 0, \end{cases}$$

by the same argument as in (5.3.4) above. □

In case $i = j$, R_i is simply the number of returns to state $\textcircled{i}$ starting from state $\textcircled{i}$, and it has the geometric distribution

$$\mathbb{P}(R_i = m \mid Z_0 = i) = (1 - p_{ii})(p_{ii})^m, \qquad m \geqslant 0. \tag{5.4.3}$$

Proposition 5.6. *We have*

$$\mathbb{P}(R_j < \infty \mid Z_0 = i) = \begin{cases} 1 - p_{ij}, & \text{if } p_{jj} = 1, \\ \\ 1, & \text{if } p_{jj} < 1. \end{cases}$$

Proof. We note that

* Here, R_j is called a number of returns because the time counter is started at $n = 1$ and excludes the initial state.

$$\begin{aligned}\mathbb{P}(R_j < \infty \mid Z_0 = i) &= \mathbb{P}(R_j = 0 \mid Z_0 = i) + \sum_{m \geqslant 1} \mathbb{P}(R_j = m \mid Z_0 = i) \\ &= 1 - p_{ij} + p_{ij}(1 - p_{jj}) \sum_{m \geqslant 1} (p_{jj})^{m-1} \\ &= \begin{cases} 1 - p_{ij}, & \text{if } p_{jj} = 1, \\ \\ 1, & \text{if } p_{jj} < 1. \end{cases}\end{aligned}$$

□

We also have

$$\mathbb{P}(R_j = \infty \mid Z_0 = i) = \begin{cases} p_{ij}, & \text{if } p_{jj} = 1, \\ \\ 0, & \text{if } p_{jj} < 1. \end{cases}$$

In particular, if $p_{jj} = 1$, *i.e.* state $\textcircled{j}$ is recurrent, we have

$$\mathbb{P}(R_j = m \mid Z_0 = i) = 0, \qquad m \geqslant 1,$$

and in this case,

$$\begin{cases} \mathbb{P}(R_j < \infty \mid Z_0 = i) = \mathbb{P}(R_j = 0 \mid Z_0 = i) = 1 - p_{ij}, \\ \\ \mathbb{P}(R_j = \infty \mid Z_0 = i) = 1 - \mathbb{P}(R_j < \infty \mid Z_0 = i) = p_{ij}. \end{cases}$$

On the other hand, when $i = j$, by (1.5.13) we find

$$\begin{aligned}\mathbb{P}(R_i < \infty \mid Z_0 = i) &= \sum_{m \geqslant 0} \mathbb{P}(R_i = m \mid Z_0 = i) \\ &= (1 - p_{ii}) \sum_{m \geqslant 0} (p_{ii})^m \\ &= \begin{cases} 0, & \text{if } p_{ii} = 1, \\ \\ 1, & \text{if } p_{ii} < 1, \end{cases}\end{aligned} \tag{5.4.4}$$

hence

$$\mathbb{P}(R_i = \infty \mid Z_0 = i) = \begin{cases} 1, & \text{if } p_{ii} = 1, \\ \\ 0, & \text{if } p_{ii} < 1, \end{cases} \tag{5.4.5}$$

i.e. the number of returns to a recurrent state is infinite with probability one. This is an example of a 0-1 law.

Mean number of returns

The notion of *mean number of returns* will be needed for the classification of states of Markov chains in Chapter 6.

Proposition 5.7. *The mean number of returns to state* $\textcircled{j}$ *is given by*

$$\mathbb{E}[R_j \mid X_0 = i] = \frac{p_{ij}}{1 - p_{jj}},$$

and it is finite, i.e. $\mathbb{E}[R_j \mid X_0 = i] < \infty$*, if and only if* $p_{jj} < 1$*,* $i, j \in \mathbb{S}$*.*

Proof. By (A.4), when $p_{jj} < 1$ we have $\mathbb{P}(R_j < \infty \mid X_0 = i) = 1$ and

$$\begin{aligned}\mathbb{E}[R_j \mid X_0 = i] &= \sum_{m \geqslant 1} m \mathbb{P}(R_j = m \mid X_0 = i) \qquad (5.4.6)\\ &= (1 - p_{jj}) p_{ij} \sum_{m \geqslant 1} m (p_{jj})^{m-1} \\ &= \frac{p_{ij}}{1 - p_{jj}}, \qquad (5.4.7)\end{aligned}$$

hence

$$\mathbb{E}[R_j \mid X_0 = i] < \infty \quad \text{if} \quad p_{jj} < 1.$$

If $p_{jj} = 1$ then $\mathbb{E}[R_j \mid X_0 = i] = \infty$ unless $p_{ij} = 0$, in which case $\mathbb{P}(R_j = 0 \mid X_0 = i) = 1$ and $\mathbb{E}[R_j \mid X_0 = i] = 0$. □

By analogy with (A.3), the matrix inverse

$$G := (I - P)^{-1} = \sum_{n \geqslant 0} P^n = I + \sum_{n \geqslant 1} P^n \qquad (5.4.8)$$

of $I - P$ is called the *potential kernel*, or the *resolvent* of P. See Althoen et al. (1993) for an application of the next proposition to the Snakes and Ladders game.

Proposition 5.8. *If* $\textcircled{m}$ *is the* only *absorbing state in* $\mathbb{S}$*, we have the expression*

$$\mathbb{E}[T_m \mid Z_0 = i] = \sum_{\substack{j \in \mathbb{S} \\ j \neq m}} \left[(I - P)^{-1} \right]_{i,j}, \quad i \neq m. \qquad (5.4.9)$$

Proof. By (5.4.2) and (5.4.8) we have

$$\begin{aligned}\mathbb{E}[R_j \mid Z_0 = i] &= \mathbb{E}\left[\sum_{n \geqslant 1} \mathbb{1}_{\{Z_n = j\}} \,\middle|\, Z_0 = i \right] \\ &= \sum_{n \geqslant 1} \mathbb{E}[\mathbb{1}_{\{Z_n = j\}} \mid Z_0 = i]\end{aligned}$$

$$= \sum_{n \geqslant 1} \mathbb{P}(Z_n = j \mid Z_0 = i)$$
$$= \sum_{n \geqslant 1} [P^n]_{i,j}$$
$$= -\mathbb{1}_{\{i=j\}} + \left[(I-P)^{-1}\right]_{i,j}.$$

On the other hand, after starting from $i \neq m$ we have

$$T_m = 1 + \sum_{\substack{j \in \mathbb{S} \\ j \neq m}} R_j,$$

hence

$$\mathbb{E}[T_m \mid Z_0 = i] = 1 + \sum_{\substack{j \in \mathbb{S} \\ j \neq m}} \mathbb{E}[R_j \mid Z_0 = i]$$
$$= 1 + \sum_{\substack{j \in \mathbb{S} \\ j \neq m}} \left(-\mathbb{1}_{\{i=j\}} + \left[(I-P)^{-1}\right]_{i,j}\right)$$
$$= \sum_{\substack{j \in \mathbb{S} \\ j \neq m}} \left[(I-P)^{-1}\right]_{i,j}.$$

□

Exercises

Exercise 5.1 Consider the Markov chain $(X_n)_{n \geqslant 0}$ with state space $\mathbb{S} = \{0,1,2,3\}$ and transition probabilities

$$\mathbb{P}(X_1 = 0 \mid X_0 = 0) = 1, \qquad \mathbb{P}(X_1 = 3 \mid X_0 = 3) = 1,$$
$$\mathbb{P}(X_1 = 0 \mid X_0 = 1) = 1/2, \qquad \mathbb{P}(X_1 = 2 \mid X_0 = 1) = 1/2,$$
$$\mathbb{P}(X_1 = 1 \mid X_0 = 2) = 1/3, \qquad \mathbb{P}(X_1 = 3 \mid X_0 = 2) = 2/3.$$

a) Draw the graph of the chain and write down its transition matrix.
b) Compute $\alpha := \mathbb{P}(T_3 < \infty \mid X_0 = 1)$ and $\beta := \mathbb{P}(T_3 < \infty \mid X_0 = 2)$, where
$$T_3 := \inf\{n \geqslant 0 \ : \ X_n = 3\}.$$
c) Letting
$$T_{0,3} := \inf\{n \geqslant 0 \ : \ X_n = 0 \text{ or } X_n = 3\},$$
compute $\mathbb{E}[T_{0,3} \mid X_0 = 1]$ and $\mathbb{E}[T_{0,3} \mid X_0 = 2]$.

Exercise 5.2 Consider the two-state Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0,1\}$ and transition matrix

$$P = \begin{bmatrix} 0.4 & 0.6 \\ \\ 0.8 & 0.2 \end{bmatrix}.$$

Compute the mean duration between two visits to state ①.

Exercise 5.3 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on $\mathbb{S} = \{0,1,2\}$ whose transition probability matrix P is given by

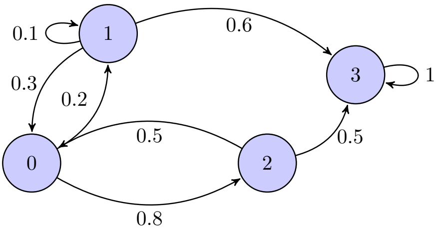
$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1 & 0 \end{bmatrix} \end{matrix}.$$

a) Draw a graph of the chain and find the probability $g_0(k)$ that the chain is absorbed into state ⓪ given that it started from states $k = 0,1,2$.
b) Determine the mean time $h_0(k)$ it takes until the chain is absorbed into state ⓪, after starting from $k = 0,1,2$.

Exercise 5.4 A box contains red balls and green balls. At each time step we pick a ball uniformly at random and without replacement. If the ball is red we lose \$1, and if the ball is green we gain +\$1. The game ends when the box becomes empty. We let $f(x,y)$ denote the value of the game when the game starts with $x \geqslant 0$ red balls and $y \geqslant 0$ green balls in the box.

a) Find the boundary conditions $f(x,0)$, $x \geqslant 0$, and $f(0,y)$, $y \geqslant 0$.
b) Using first step analysis, derive the finite difference equation satisfied by $f(x,y)$ for $x,y \geqslant 1$.
c) Solve the equation of Question (b) for $f(x,y)$, $x,y = 1,2,3$.
d) Find $f(x,y)$ for all $x,y \geqslant 0$.

Exercise 5.5 Consider the Markov chain with graph



(5.4.10)

and let

$$T_k^r := \inf\{n \geqslant 1 \ : \ X_n = k\}$$

denote the *return* time to state $k = 0, 1, 2, 3$.

a) Find the probabilities

$$p_2(k) := \mathbb{P}\big(T_2^r < \infty \mid X_0 = k\big),$$

of returning to state ② in finite time after starting from $k = 0, 1, 2, 3$.

b) Find the probabilities

$$p_1(k) := \mathbb{P}\big(T_1^r < \infty \mid X_0 = k\big), \qquad k = 0, 1, 2, 3.$$

of *returning* to state ① in finite time after starting from $k = 0, 1, 2, 3$.

c) Compute the mean return times $\mu_2(k) := \mathbb{E}\big[T_2^r \mid X_0 = k\big]$ to state ② after starting from $k = 0, 1, 2, 3$.

Exercise 5.6 Consider the Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0, 1, 2, 3\}$ and transition probability matrix given by

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 3} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) What are the absorbing states of the chain $(X_n)_{n\geqslant 0}$?

b) Given that the chain starts at ①, find its probability of absorption $g_0(k) = \mathbb{P}(T_0 < \infty \mid X_0 = k)$ into state ⓪ for $k = 0, 1, 2, 3$.

c) Find the mean hitting times $h_1(k) = \mathbb{E}[T_1 \mid X_0 = k]$ of state ① starting from state ⓚ, for $k = 0, 1, 2, 3$.

Exercise 5.7 Consider the random walk with Markov transition matrix given by

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 3} = \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Compute the average time it takes to reach state ③ given that the chain is started at state ⓪.

Exercise 5.8 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on $\{0, 1, 2, 3\}$ whose transition probability matrix P is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) Draw the graph of this chain.
b) Find the probability $g_0(k)$ that the chain is absorbed into state ⓪ given that it started from state $k = 0, 1, 2, 3$.
c) Determine the mean time $h(k)$ it takes until the chain hits an absorbing state, after starting from $k = 0, 1, 2, 3$.

Exercise 5.9 Consider a discrete-time, time-homogeneous Markov chain $(X_n)_{n\geqslant 0}$ on a state space $\mathbb{S}$, and the first hitting time

$$T_A = \inf\{n \geqslant 0 \ : \ X_n \in A\},$$

of a subset A of $\mathbb{S}$. Show that $(X_n)_{n\geqslant 0}$ has the *strong Markov property* with respect to T_A, *i.e.* show that for all $n, m \geqslant 0$, $j \in \mathbb{S}$, and $(i_k)_{k\geqslant 0} \subset \mathbb{S}$ we have

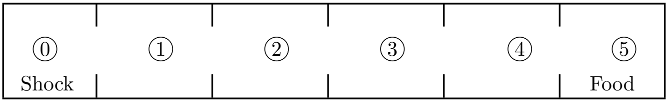
$$\mathbb{P}(X_{T_A+n} = j \mid X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty) = \mathbb{P}(X_n = j \mid X_0 = i_0).$$

Exercise 5.10 We consider the simple random walk $(S_n)_{n\geqslant 0}$ of Chapter 3.

a) Using first step analysis, recover the formula (3.4.16) for the probability $\mathbb{P}(T_0 < \infty \mid S_0 = k)$ of hitting state ⓪ in finite time starting from any state ⓚ $\geqslant 0$ when $q < p$.
b) Using first step analysis, recover the formula (3.4.18) giving the mean hitting time $\mathbb{E}[T_0 \mid S_0 = k]$ of state ⓪ from any state ⓚ $\geqslant 0$ when $q > p$.

Exercise 5.11 A player tosses a fair six-sided die and records the number appearing on the uppermost face. The die is then tossed again, and the second result is added to the first one. This procedure is repeated until the sum of all results becomes strictly greater than 10. Compute the probability that the game finishes with a cumulative sum equal to 13.

Exercise 5.12 A fish is placed into the linear maze as shown, and its state at time n is denoted by $X_n \in \{0, 1, 2, 3, 4, 5\}$:



Starting from any state $k \in \{1, 2, 3, 4\}$, the fish moves to the right with probability p and to the left with probability q such that $p + q = 1$ and $p \in (0, 1)$. Consider the hitting times

$$T_0 = \inf\{n \geqslant 0 \ : \ X_n = 0\}, \quad \text{and} \quad T_5 = \inf\{n \geqslant 0 \ : \ X_n = 5\},$$

and $g(k) = \mathbb{P}(T_5 < T_0 \mid X_0 = k)$, $k = 0, 1, \ldots, 5$.

a) Using first step analysis, write down the equation satisfied by $g(k)$, $k = 0, 1, \ldots, 5$, and give the values of $g(0)$ and $g(5)$.
b) Assume that the fish is equally likely to move right or left at each step. Compute the probability that starting from state $\textcircled{k}$ it finds the food before getting shocked, for $k = 0, 1, \ldots, 5$.

Exercise 5.13 Starting from a state $m \geqslant 1$ at time k, the next state of a random device at time $k + 1$ is uniformly distributed among $\{0, 1, \ldots, m - 1\}$, with $\textcircled{0}$ as an absorbing state.

a) Model the time evolution of this system using a Markov chain whose transition probability matrix will be given explicitly.
b) Let $h_0(m)$ denote the mean time until the system reaches the state zero for the first time after starting from state $\textcircled{m}$. Using first step analysis, write down the equation satisfied by $h_0(m)$, $m \geqslant 1$ and give the values of $h_0(0)$ and $h_0(1)$.
c) Show that $h_0(m)$ is given by $h_0(m) = h_0(m-1) + \dfrac{1}{m}$, $m \geqslant 1$, and that

$$h_0(m) = \sum_{k=1}^{m} \frac{1}{k},$$

for all $m \geqslant 0$.

Exercise 5.14 An individual is placed in a castle tower having three exits. Exit A leads to a tunnel that returns to the tower after three days of walk. Exit B leads to a tunnel that returns to the tower after one day of walk. Exit C leads to the outside. Since the inside of the tower is dark, each exit is chosen at random with probability $1/3$. The individual decides to remain outside after exiting the tower, and you may choose the number of steps it takes from Exit C to the outside of the tower, *e.g.* take it equal to 0 for simplicity.

a) Show that this problem can be modeled using a Markov chain $(X_n)_{n\geqslant 0}$ with four states. Draw the graph of the chain $(X_n)_{n\geqslant 0}$.
b) Write down the transition matrix of the chain $(X_n)_{n\geqslant 0}$.
c) Starting from inside the tower, find the average time it takes to exit the tower.

Exercise 5.15 A mouse is trapped in a maze. Initially, it has to choose one of two directions. If it goes to the right, then it will wander around in the maze for three minutes and will then return to its initial position. If it goes to the left, then with probability 1/3 it will depart the maze after two minutes of travelling, and with probability 2/3 it will return to its initial position after five minutes of travelling. Assuming that the mouse is at all times equally likely to go to the left or to the right, what is the expected number of minutes that it will remain trapped in the maze?

Exercise 5.16 Consider a two-state $\{0,1\}$-valued Markov chain $(X_n)_{n\geqslant 0}$ on the state space with transition matrix

$$P = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \end{matrix},$$

where $a, b \in (0,1)$. This question is to be treated via explicit computations for two-state Markov chains, without referring to general results.

a) Give the stationary distribution $\pi = (\pi_0, \pi_1)$ of the chain $(X_n)_{n\geqslant 0}$.
b) Compute the mean return times $\mu_0(0)$, $\mu_1(1)$ and the mean hitting times $h_0(1)$, $h_1(0)$ of the chain $(X_n)_{n\geqslant 0}$.
c) Compute the conditional expected values $\mathbb{E}[\tau \mid X_0 = 0]$ and $\mathbb{E}[\tau \mid X_0 = 1]$ of the cycle length

$$\tau := \inf\{l > 1 \ : \ X_l = X_1\}.$$

d) Compute the four expected values

$$\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l = i\}} \,\middle|\, X_0 = j\right], \qquad i,j = 0,1.$$

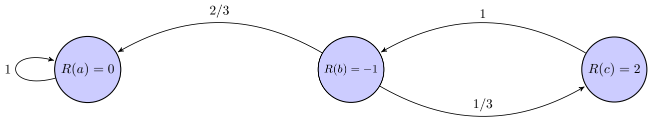
e) Show that for any initial distribution $(\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1))$ we have

$$\pi_0 = \frac{\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=0\}}\right]}{\mathbb{E}[\tau - 1]}, \quad \pi_1 = \frac{\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=1\}}\right]}{\mathbb{E}[\tau - 1]}.$$

Exercise 5.17 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\mathbb{S} = \{a, b, c\}$ whose transition probability matrix P is given by

$$P = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 1 & 0 \end{array}\right] \end{array},$$

with the following graph:

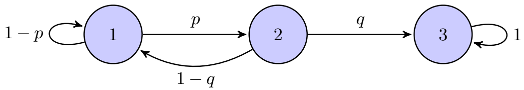


Given the following reward function:

$$R(a) = 0, \qquad R(b) = -1, \qquad R(c) = 2,$$

determine the average accumulated reward $V_a(k) = \mathbb{E}\left[\sum_{n=0}^{\infty} R(X_n) \,\middle|\, X_0 = k\right]$ until the chain is absorbed into state ⓐ after starting from $k = a, b, c$, assuming a discount factor $\gamma = 1$.

Exercise 5.18 Let $(X_n)_{n\geqslant 0}$ be a three-state Markov chain with the following transition probability graph.



By first step analysis, compute the value function

$$V(k) = \mathbb{E}\left[\sum_{n\geqslant 0} \gamma^n R(X_n) \,\middle|\, X_0 = k\right], \qquad k = 1, 2, 3,$$

where $\gamma \in (0, 1)$ is a discount factor and $R : \mathbb{S} \to \mathbb{R}$ is the reward function given by

$$R(1) := -\$2, \; R(2) := \$3, \; R(3) := \$1.$$

Exercise 5.19 Two buffalos are traveling in opposite directions on a one-dimensional road $\{0, 1, \ldots, S\}$, one step at a time. Buffalo A starts from ⓪, moving up by $+1$ at every time step, and Buffalo B starts at the same time from Ⓢ, moving down by -1 at every time step.

a) How many time steps does it take for Buffalo A to travel up from ⓪ to Ⓢ, and for Buffalo B to travel down from Ⓢ to ⓪ ?

b) Next, we assume that when the buffalos collide, they either both continue the same ways with probability p, or they both turn back and continue in opposite directions with probability $q = 1 - p$. How many time steps does it take for the buffalos to reach any of the boundaries $\textcircled{0}$ or $\textcircled{S}$?

Exercise 5.20 This exercise is a particular case of (5.1.9). Consider the Markov chain whose transition probability matrix P is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) Find the probability that the chain finishes at $\textcircled{0}$ given that it was started at state $\textcircled{1}$.
b) Determine the mean time it takes until the chain reaches an absorbing state.

Exercise 5.21 This exercise is stated in the framework of (5.1.9). Consider the Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0, 1, 2, 3\}$ and transition probability matrix given by

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 3} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0 & 0.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) What are the absorbing states of the chain $(X_n)_{n\geqslant 0}$?
b) Denoting by $T_k := \inf\{n \geqslant 0 \ : \ X_n = k\}$ the first hitting time of state $\textcircled{k}$, find the probabilities $g_1(k) = \mathbb{P}(T_1 < \infty \mid X_0 = k)$ of hitting state $\textcircled{1}$ in finite time after starting from state $\textcircled{k}$, for $k = 0, 1, 2, 3$.
c) Denoting by $T_1^r := \inf\{n \geqslant 1 \ : \ X_n = 1\}$ the first *return time* to state $\textcircled{1}$, find the probabilities $p_1(k) = \mathbb{P}(T_1^r < \infty \mid X_0 = k)$ of returning to state $\textcircled{1}$ in finite time after starting from state $\textcircled{k}$, for $k = 0, 1, 2, 3$.
d) Find the mean hitting times $h_1(k) = \mathbb{E}[T_1 \mid X_0 = k]$ of state $\textcircled{1}$ and the mean return times $\mu_1(k) = \mathbb{E}[T_1 \mid X_0 = k]$ to state $\textcircled{1}$ after starting from state $\textcircled{k}$, for $k = 0, 1, 2, 3$.

Exercise 5.22 Consider the Markov chain on $\{0, 1, 2\}$ with transition matrix

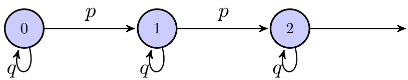
$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

a) Compute the probability $\mathbb{P}(T_2 < \infty \mid X_0 = 1)$ of *hitting* state ② in finite time starting from state ①, and the probability $\mathbb{P}\big(T_1^r < \infty \mid X_0 = 1\big)$ of *returning* to state ① in finite time.
b) Compute the mean *return time* $\mu_1(1) = \mathbb{E}[T_1^r \mid X_0 = 1]$ to state ① and the mean *hitting time* $h_2(1) = \mathbb{E}[T_2 \mid X_0 = 1]$ of state ② starting from state ①.

Exercise 5.23 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the countably infinite state space $\mathbb{S} = \mathbb{N} = \{0, 1, 2, 3, \ldots\}$, with the infinite transition matrix

$$P = [\ P_{i,j}\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} q & p & 0 & 0 & 0 & \cdots \\ 0 & q & p & 0 & 0 & \cdots \\ 0 & 0 & q & p & 0 & \cdots \\ 0 & 0 & 0 & q & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $p, q \in (0, 1)$ are such that $p + q = 1$.



a) By a recurrence using Pascal's identity

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},$$

compute $[P^n]_{i,j}$, $n \geqslant 1$, in the cases (1) $j - i \leqslant n$, (2) $n < j - i$, (3) $i > j$.
b) Show that for all $i, j \geqslant 0$ we have

$$\lim_{n\to\infty} [P^n]_{i,j} = 0.$$

c) Compute

$$\sum_{n\geqslant 0} [P^n]_{i,j}$$

in the cases (1) $i \leqslant j$, (2) $i > j$.
d) Letting $T_j := \inf\{n \geqslant 0 \ : \ X_n = j\}$, determine the value of

$$p_{i,j} := \mathbb{P}(T_j < \infty \mid X_0 = i)$$

in the cases (1) $i < j$, (2) $i = j$, (3) $i > j$.
e) Is the chain $(X_n)_{n\geqslant 0}$ recurrent or transient?

f) Compute the mean number of returns $\mathbb{E}[R_j \mid X_0 = i]$ from state (i) to state (j) in the cases (1) $i < j$, (2) $i = j$, (3) $i > j$.
g) Show that the matrix $I - P$ is invertible, and compute its inverse $(I - P)^{-1}$.

Exercise 5.24 Let $\mathbb{N} := \{0, 1, 2, \ldots\}$ and consider the two-dimensional random walk $(Z_k)_{k \geqslant 0} = (X_k, Y_k)_{k \geqslant 0}$ on $\mathbb{N} \times \mathbb{N}$ with the transition probabilities

$$\mathbb{P}\big((X_{k+1}, Y_{k+1}) = (x+1, y) \mid (X_k, Y_k) = (x, y)\big)$$
$$= \mathbb{P}\big((X_{k+1}, Y_{k+1}) = (x, y+1) \mid (X_k, Y_k) = (x, y)\big) = \frac{1}{2}, \quad (x, y) \in \mathbb{N} \times \mathbb{N},$$

$k \geqslant 0$, and let

$$A := \mathbb{N}^2 \setminus \{0, 1, 2\}^2 = \big\{(x, y) \in \mathbb{N} \times \mathbb{N} \ : \ x \geqslant 3 \text{ or } y \geqslant 3\big\}.$$

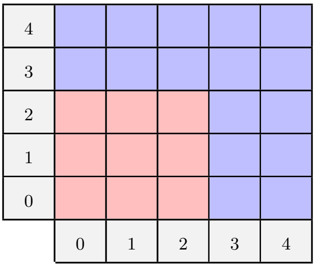


Table 5.1: Domain A with $N = 3$ (in blue).

Let also

$$T_A := \inf\big\{n \geqslant 0 \ : \ (X_n, Y_n) \in A\big\}$$

denote the first hitting time of the set A by the random walk $(Z_k)_{k \geqslant 0} = (X_k, Y_k)_{k \geqslant 0}$, and consider the mean hitting times

$$\mu_A(x, y) := \mathbb{E}\big[T_A \mid (X_0, Y_0) = (x, y)\big], \quad (x, y) \in \mathbb{N} \times \mathbb{N}.$$

a) Give the values of $\mu_A(x, y)$ when $(x, y) \in A$.
b) By applying first step analysis, find an equation satisfied by $\mu_A(x, y)$ on the domain
$$A^c = \big\{(x, y) \in \mathbb{N} \times \mathbb{N} \ : \ 0 \leqslant x, y \leqslant 3\big\}.$$
c) Find the values of $\mu_A(x, y)$ for all $x, y \leqslant 3$ by solving the equation of Question (b).

d) In each round of a ring toss game, a ring is thrown at two sticks in such a way that each stick has exactly 50% chance to receive the ring. Compute the mean time it takes until at least one of the two sticks receives three rings.

Exercise 5.25 Taking $\mathbb{N} := \{0, 1, 2, \ldots\}$, consider the two-dimensional random walk $(Z_k)_{k\geqslant 0} = (X_k, Y_k)_{k\geqslant 0}$ on $\mathbb{N} \times \mathbb{N}$ with the transition probabilities

$$\begin{aligned} &\mathbb{P}(X_{k+1} = x + 1,\ Y_{k+1} = y \mid X_k = x,\ Y_k = y) \\ &= \mathbb{P}(X_{k+1} = x,\ Y_{k+1} = y + 1 \mid X_k = x,\ Y_k = y) \\ &= \frac{1}{2}, \qquad k \geqslant 0, \end{aligned}$$

and let

$$A = [2, \infty) \times [2, \infty) = \big\{(x, y) \in \mathbb{N} \times \mathbb{N} \ : \ x \geqslant 2,\ y \geqslant 2\big\}.$$

4					
3					
2					
1					
0					
	0	1	2	3	4

Table 5.2: Domain A with $N = 2$ (in blue).

Let also

$$T_A := \inf\{n \geqslant 0 \ : \ X_n \geqslant 2 \text{ and } Y_n \geqslant 2\}$$

denote the hitting time of the set A by the random walk $(Z_k)_{k\geqslant 0}$, and consider the mean hitting times

$$\mu_A(x, y) := \mathbb{E}[T_A \mid X_0 = x,\ Y_0 = y], \qquad x, y \in \mathbb{N}.$$

a) Give the value of $\mu_A(x, y)$ when $x \geqslant 2$ and $y \geqslant 2$.
b) Show that $\mu_A(x, y)$ solves the equation

$$\mu_A(x, y) = 1 + \frac{1}{2}\mu_A(x + 1, y) + \frac{1}{2}\mu_A(x, y + 1), \quad x, y \in \mathbb{N}. \qquad (5.4.11)$$

c) Show that $\mu_A(1, 2) = \mu_A(2, 1) = 2$ and $\mu_A(0, 2) = \mu_A(2, 0) = 4$.

d) In each round of a ring toss game, a ring is thrown at two sticks in such a way that each stick has exactly 50% chance to receive the ring. Compute the mean time it takes until both sticks receive at least two rings.

Exercise 5.26 Let $(X_n)_{n\geqslant 0}$ be a Markov chain with state space $\mathbb{S}$ and transition probability matrix $(P_{ij})_{i,j\in\mathbb{S}}$. Our goal is to compute the expected value of the infinite discounted series

$$h(i) := \mathbb{E}\left[\sum_{n\geqslant 0}\beta^n c(X_n) \,\Big|\, X_0 = i\right], \qquad i \in \mathbb{S},$$

where $\beta \in (0,1)$ is the discount coefficient and $c(\cdot)$ is a utility function, starting from state (i).

a) Show, by a first step analysis argument, that $h(i)$ satisfies the equation

$$h(i) = c(i) + \beta \sum_{j\in\mathbb{S}} P_{ij} h(j)$$

for every state $(i) \in \mathbb{S}$.

b) Consider the Markov chain on the state space $\mathbb{S} = \{0,1,2\}$ with transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left[\begin{array}{ccc} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{array}\right] \end{array},$$

and the utility function $c : \mathbb{S} \to \mathbb{Z}$ defined by

$$c(0) = \$5, \quad c(1) = -\$2, \quad c(2) = 0.$$

Compute the accumulated utility $h(k)$ after starting from states $k = 0, 1, 2$, by taking $\beta := 1$.

Exercise 5.27 Consider a sequence $(X_n)_{n\geqslant 1}$ of independent Bernoulli random variables with the distribution

$$\mathbb{P}(X_n = a) = p, \qquad \mathbb{P}(X_n = b) = q, \qquad n \geqslant 1,$$

where $p \in (0,1]$ and $q := 1 - p$. Let $T^{(m)}$ denote the time of the first appearance of m consecutive a's in $(X_n)_{n\geqslant 1}$, with *e.g.* $T^{(4)} = 8$ in the following sequence:

$$(\underset{1}{\underset{\uparrow}{b}}, \underset{2}{\underset{\uparrow}{a}}, \underset{3}{\underset{\uparrow}{a}}, \underset{4}{\underset{\uparrow}{b}}, \overbrace{\underset{5}{\underset{\uparrow}{a}}, \underset{6}{\underset{\uparrow}{a}}, \underset{7}{\underset{\uparrow}{a}}, \underset{8}{\underset{\uparrow}{a}}}^{4 \text{ times}} b, a, a, b, \ldots).$$

a) By first step analysis, find an equation satisfied by $\mathbb{E}[T^{(m)}]$.
b) Compute the mean time $\mathbb{E}[T^{(m)}]$ until we encounter an m-winning streak, for all $m \geqslant 1$.

Hint. We have

$$\sum_{k=1}^{m} kp^{k-1} = \frac{\partial}{\partial p}\sum_{k=0}^{m} p^k = \frac{\partial}{\partial p}\left(\frac{1-p^{m+1}}{1-p}\right) = \frac{1-(m+1)p^m + mp^{m+1}}{(1-p)^2}.$$

Exercise 5.28 Consider a *Markov Decision Process* (MDP)* on a state space $\mathbb{S}$, with set of actions $\mathbb{A}$ and family $(P^a)_{a\in\mathbb{A}}$ of transition probability matrices

$$P : \mathbb{S}\times\mathbb{S}\times\mathbb{A} \to [0,1],$$
$$(k,l,a) \longmapsto P^a_{k,l}.$$

Using first step analysis, derive the *Bellman equation*

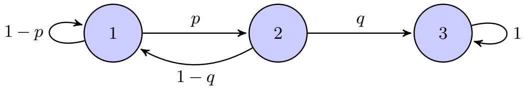
$$V^*(k) = R(k) + \gamma \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P^a_{k,l} V^*(l), \qquad k \in \mathbb{S},$$

satisfied by the optimal value function

$$V^*(k) = \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0}\gamma^n R(X_n) \,\middle|\, X_0 = k\right],$$

over all *policies* $\pi : \mathbb{S} \to \mathbb{A}$ giving the action chosen at every given state in $\mathbb{S}$, where $\gamma \in (0,1)$ is a discount factor and $R : \mathbb{S} \to \mathbb{R}$ is a reward function.

Exercise 5.29 (Vinay and Kok (2019)). The double-heralding protocol for entanglement generation in quantum cryptography involves two rounds of photon transfer, the failure of either of which will cause the process to be restarted.



The protocol is modeled using a Markov chain $(X_n)_{n\geqslant 0}$ described by the above graph, in which $p \in (0,1)$ is the probability of passing the first round, and $q \in (0,1)$ is the probability of passing the second round, conditional on passing the first. Let

* See Problem 5.39 for details.

$$T_3 = \inf \left\{ n \geqslant 0 \ : \ X_n = 3 \right\}$$

denote the first hitting time of state ③ by the chain $(X_n)_{n \geqslant 0}$.

a) Using first step analysis, find the mean time to completion of double-heralding after starting from state ⓘ, $i = 1, 2$.
b) Find the Probability Generating Function (PGF)

$$G_i(s) = \mathbb{E}\left[s^{T_3} \,\middle|\, X_0 = i \right], \qquad -1 \leqslant s \leqslant 1,$$

of T_3 after starting from state ⓘ, $i = 1, 2$.

Hint. Start by deriving a system of equations satisfied by $G_i(s)$ using first step analysis.

c) Find the probability distribution $\mathbb{P}(T_3 = k \mid X_0 = 1)$ of the completion time after starting from state ①.

Hint. Use the power series expansion

$$\frac{\sqrt{(1-p)^2 + 4(1-q)p}}{1 - (1-p)s - p(1-q)s^2} = \sum_{n=0}^{\infty} \frac{s^n}{z_+^{n+1}} - \sum_{n=0}^{\infty} \frac{s^n}{z_-^{n+1}},$$

where

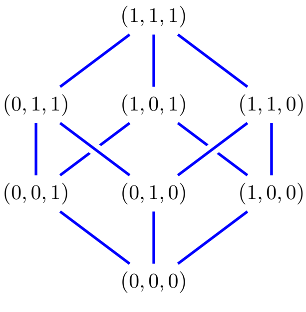
$$z_\pm := \frac{p - 1 \pm \sqrt{(1-p)^2 + 4(1-q)p}}{2(1-q)p}.$$

d) Using $G_1(s)$, recover the mean time to completion of double-heralding after starting from state ①, as found in Question (a).

Exercise 5.30 We consider an ant moving randomly on the vertices of the 3-dimensional cube $\mathcal{C}_3$ represented as

$$\mathcal{C}_3 = \left\{ (e_1, e_2, e_3) \ : \ e_1, e_2, e_3 \in \{0, 1\} \right\},$$

by choosing a new edge with probability 1/3 at every time step.



Using first step analysis, compute the mean time $h(r)$, $r = 0, 1, 2, 3$, until the ant reaches the vertex $(0, 0, 0)$ after starting from a vertex in the set $\mathcal{S}_r$ of vertices which are at distance $r = 0, 1, 2, 3$ from $(0, 0, 0)$, with $\mathcal{S}_0 = \{(0, 0, 0)\}$ and $\mathcal{S}_3 = \{(1, 1, 1)\}$.

Problem 5.31. We consider an ant moving randomly on the vertices of the d-dimensional (hyper)cube $\mathcal{C}_d$ represented as

$$\mathcal{C}_d = \big\{(e_1, \ldots, e_d) \ : \ e_1, \ldots e_d \in \{0, 1\}\big\},$$

by choosing a new edge with probability $1/d$ at every time step. We aim at computing the mean time $h(r)$ until the ant reaches the vertex $(0, \ldots, 0)$ after starting from a vertex in the set $\mathcal{S}_r$ of vertices which are at distance $r \in \{0, \ldots, d\}$ of $(1, \ldots, 1)$, with $\mathcal{S}_0 = \{(1, \ldots, 1)\}$ and $\mathcal{S}_d = \{(0, \ldots, 0)\}$.

a) Give the value of $h(d)$.
b) Find a relation between $h(0)$ and $h(1)$.
c) Using first step analysis, find a relationship between $h(r)$, $h(r-1)$ and $h(r+1)$ for $r = 1, \ldots, d-1$.
d) Letting $f(r) := h(r+1) - h(r)$, $r = 0, \ldots, d-1$, find a recurrence relation between $f(r)$ and $f(r-1)$ for $r = 1, \ldots, d-1$.
e) Find the value of $f(0)$ and solve the equation of Question (d) for $f(r)$, $r = 1, \ldots, d$.

   Hint. The solution of the equation

$$rf(r-1) = d + (d-r)f(r), \qquad r = 1, 2, \ldots, d,$$

   with $f(0) = -1$ is given by

$$f(r) = -\frac{1}{\binom{d-1}{r}} \sum_{k=0}^{r} \binom{d}{k}, \qquad r = 0, 1, \ldots, d.$$

f) Using a telescoping identity, find the value of $h(r)$ for $r = 0, \ldots, d$.
g) Give the values of $h(0)$, $h(1)$ and $h(2)$.
h) Find the values of $h(r)$ for $r = 0, 1, \ldots, d$ in the following cases:

   i) $d = 1$,
   ii) $d = 2$,
   iii) $d = 3$.

Problem 5.32 Let $(X_n)_{n\geqslant 0}$ be a Markov chain on $\{0, 1, \ldots, N\}$, $N \geqslant 1$, with transition matrix $P = \big[\ P_{i,j} \ big\big]_{0\leqslant i,j\leqslant N}$.

a) Consider the hitting times

$$T_0 = \inf\{n \geqslant 0 \ : \ X_n = 0\}, \qquad T_N = \inf\{n \geqslant 0 \ : \ X_n = N\},$$

and

$$g(k) = \mathbb{P}(T_0 < T_N \mid X_0 = k), \qquad k = 0, 1, \ldots, N.$$

What are the values of $g(0)$ and $g(N)$?

b) Show, using first step analysis, that the function g satisfies the relation

$$g(k) = \sum_{l=0}^{N} P_{k,l} g(l), \qquad k = 1, \ldots, N-1. \tag{5.4.12}$$

c) In this question and the following ones we consider the Wright-Fisher stochastic model in population genetics, in which the state X_n denotes the number of individuals in the population at time n, and

$$P_{k,l} = \mathbb{P}(X_{n+1} = l \mid X_n = k) = \binom{N}{l} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l},$$

$k, l = 0, 1, \ldots, N$. Write down the transition matrix P when $\underline{N = 3}$.

d) Show, from Question (b), that given that the solution to (5.4.12) is unique, we have

$$\mathbb{P}(T_0 < T_N \mid X_0 = k) = \frac{N-k}{N}, \qquad k = 0, 1, \ldots, N.$$

e) Let

$$T_{0,N} = \inf\{n \geqslant 0 \ : \ X_n = 0 \text{ or } X_n = N\},$$

and

$$h(k) = \mathbb{E}[T_{0,N} \mid X_0 = k], \qquad k = 0, 1, \ldots, N.$$

What are the values of $h(0)$ and $h(N)$?

f) Show, using first step analysis, that the function h satisfies the relation

$$h(k) = 1 + \sum_{l=0}^{N} P_{k,l} h(l), \qquad k = 1, 2, \ldots, N-1.$$

g) Assuming that $\underline{N = 3}$, compute

$$h(k) = \mathbb{E}[T_{0,3} \mid X_0 = k], \qquad k = 0, 1, 2, 3.$$

Problem 5.33 Pattern recognition (Exercise 4.7 continued). Consider a sequence $(X_n)_{n \geqslant 0}$ of *i.i.d.* Bernoulli random variables taking values in a two-letter alphabet $\{a, b\}$, with

$$\mathbb{P}(X_n = a) = p \quad \text{and} \quad \mathbb{P}(X_n = b) = q = 1 - p, \quad n \geqslant 0,$$

with $0 < p < 1$, and the process $(Z_n)_{n \geqslant 1}$ defined by

$$Z_n := (X_{n-1}, X_n), \qquad n \geqslant 1.$$

a) Argue that $(Z_n)_{n\geqslant 1}$ is a Markov chain with four possible states (or words) $\{aa, ab, ba, bb\}$, and write down its 4×4 transition matrix.

b) Let

$$\tau_{ab} = \inf\{n \geqslant 1 \ : \ Z_n = (a, b)\}$$

denote the first time of appearance of the pattern "ab" in the sequence $(X_0, X_1, X_2, \ldots)$. Give the value of

$$G_{ab}(s) := \mathbb{E}\big[s^{\tau_{ab}} \,\big|\, Z_1 = (a, b)\big], \qquad -1 < s < 1.$$

c) Consider the probability generating functions

$$G_{aa}(s) := \mathbb{E}\big[s^{\tau_{ab}} \,\big|\, Z_1 = (a, a)\big], \quad \text{and} \quad G_{ba}(s) := \mathbb{E}\big[s^{\tau_{ab}} \,\big|\, Z_1 = (b, a)\big],$$

$-1 < s < 1$. Using first step analysis, complete the system of equations

$$\begin{cases} G_{aa}(s) = psG_{aa}(s) + qsG_{ab}(s), \\ \\ G_{ba}(s) = \quad ? \quad + \quad ? \end{cases} \tag{5.4.13}$$

d) Compute $G_{aa}(s)$ and $G_{ba}(s)$ by solving the system (5.4.13).

e) Using probability generating functions, compute the averages

$$\mathbb{E}[\tau_{ab} \mid Z_1 = (a, a)] \quad \text{and} \quad \mathbb{E}[\tau_{ab} \mid Z_1 = (b, a)].$$

f) Find the average time it takes until we encounter the pattern "ab" in the sequence $(X_0, X_1, X_2, \ldots)$ started with $X_0 = a$.

Problem 5.34 Consider a gambling process $(X_n)_{n\geqslant 0}$ on the state space $\{0, 1, \ldots, N\}$, with transition probabilities

$$\mathbb{P}(X_{n+1} = k + 1 \mid X_n = k) = p, \qquad \mathbb{P}(X_{n+1} = k - 1 \mid X_n = k) = q,$$

$k = 1, 2, \ldots, N - 1$, with $p + q = 1$. Let

$$\tau := \inf\{n \geqslant 0 \ : \ X_n = 0 \text{ or } X_n = N\}$$

denote the time until the process hits either state ⓪ or state N, and consider the expectation

$$h(k) := \mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k\right],$$

of the random sum $\displaystyle\sum_{0\leqslant i<\tau} X_i$ of all chain values visited *before* the process hits 0 or N after starting from $k = 0, 1, 2, \ldots, N$.

a) Give the values of $h(0)$ and $h(N)$.*
b) Show, by first step analysis, that $h(k)$ satisfies the equations

$$h(k) = k + ph(k+1) + qh(k-1), \quad k = 1, 2, \ldots, N-1. \tag{5.4.14}$$

From now on we take $\underline{p = q = 1/2}$.
c) Find a particular solution of Equation (5.4.14).
d) Knowing that the solution of the associated *homogeneous equation*

$$f(k) = \frac{1}{2} f(k+1) + \frac{1}{2} f(k-1), \qquad k = 1, 2, \ldots, N-1,$$

takes the form

$$f(k) = C_1 + C_2 k,$$

show that the expected value $h(k)$ solution of (5.4.14) is given by

$$h(k) = k \frac{N^2 - k^2}{3}, \qquad k = 0, 1, \ldots, N.$$

e) Compute $h(1)$ when $\underline{N = 2}$ and explain why this result makes pathwise sense.
f) Suppose that you start a business with initial monthly income of \$4K. Every month the income you receive from that business may *increase* or *decrease* by \$1K with equal probabilities $(1/2, 1/2)$. You decide to stop that business as soon as your monthly income hits the levels 0 or \$70K, whichever comes first.

   i) Compute the expected duration of your business (in number of months).†
   ii) Compute your expected accumulated wealth until the month before you stop your business.
   iii) Compute your expected accumulated wealth under the assumption that your monthly income remains constant equal to \$4K over the same mean duration as in Question (f-*i*) above.
   iv) Any comment?

Problem 5.35 (Chen (2004), Propositions 2.14-2.15). Given $(X_n)_{n \geqslant 0}$ a Markov chain with transition probability matrix $P = (P_{i,j})_{i,j \in \mathbb{S}}$ on a state space $\mathbb{S}$ and $v = (v_k)_{k \in \mathbb{S}}$ a nonnegative vector, we say that $u^* = (u_i^*)_{i \in \mathbb{S}}$ is the minimal non-negative solution to the equation

---

* We apply the convention $\displaystyle\sum_{i=0}^{-1} = \sum_{0 \leqslant i < 0} = 0$.

† Recall that $\mathbb{E}[T_{0,N} \mid X_0 = k] = k(N-k)$, $k = 0, 1, \ldots, N$, see (2.3.19).

$$u_i = v_i + \sum_{k\in\mathbb{S}} P_{i,k} u_k, \qquad i \in \mathbb{S}. \tag{5.4.15}$$

if u^* satisfies (5.4.15) and any other solution u of (5.4.15) satisfies $u_i \geqslant u_i^*$, $i \in \mathbb{S}$.

For $i, j \in \mathbb{S}$, let

$$f_{i,j}^{(n)} = \mathbb{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i) = \mathbb{P}(T_j = n \mid X_0 = i),$$

$n \geqslant 0$, where T_j denotes the first hitting time of state $\textcircled{j}$ by $(X_n)_{n\geqslant 0}$.

a) Give the value of $f_{i,j}^{(1)}$ from the transition probability matrix P.

b) Using first step analysis, show that for all $j \in \mathbb{S}$, $\big(f_{i,j}^{(n)}\big)_{i\in\mathbb{S}}$ satisfies the equation

$$f_{i,j}^{(n+1)} = \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} f_{k,j}^{(n)}, \qquad i, j \in \mathbb{S},\ n \geqslant 0. \tag{5.4.16}$$

c) Let

$$f_{i,j} := \mathbb{P}(T_j < \infty \mid X_0 = i) = \sum_{n\geqslant 1} f_{i,j}^{(n)}, \qquad i, j \in \mathbb{S}.$$

Show that

$$f_{i,j} = P_{i,j} + \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} f_{k,j}, \qquad i, j \in \mathbb{S}. \tag{5.4.17}$$

d) Show that for all $j \in \mathbb{S}$, $(f_{i,j})_{i\in\mathbb{S}}$ is the unique minimal solution to Equation (5.4.17).

Hint: Letting $\tilde{f}$ denote another solution of (5.4.17), show, using (5.4.16) and induction on $n \geqslant 1$, that

$$\tilde{f}_{i,j} \geqslant \sum_{l=1}^{n} f_{i,j}^{(l)}, \quad i, j \in \mathbb{S}, \quad n \geqslant 1.$$

e) Let $g_{i,j}^{(1)} := f_{i,j}^{(1)}$ and

$$g_{i,j}^{(n+1)} := f_{i,j}^{(n+1)} + n \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} f_{k,j}^{(n)}, \quad i, j \in \mathbb{S}, \quad n \geqslant 1.$$

Using (5.4.16), show by induction on n that $g_{i,j}^{(n)} = n f_{i,j}^{(n)}$, $i, j \in \mathbb{S}$, $n \geqslant 1$.

f) Let

$$h_{i,j} := \mathbb{E}[T_j < \infty \mid X_0 = i] = \sum_{n\geqslant 1} n \mathbb{P}(T_j = n \mid X_0 = i) = \sum_{n\geqslant 1} g_{i,j}^{(n)}, \quad i, j \in \mathbb{S}.$$

Show that

$$h_{i,j} = f_{i,j} + \sum_{\substack{k\in \mathbb{S}\\ k\neq j}} P_{i,k} h_{k,j}, \qquad i,j \in \mathbb{S}, \tag{5.4.18}$$

where

$$f_{i,j} := \mathbb{P}(T_j < \infty \mid X_0 = i) = \sum_{n\geqslant 1} f_{i,j}^{(n)}, \qquad i,j \in \mathbb{S}.$$

g) Show that for all $j \in \mathbb{S}$, $(h_{i,j})_{i\in\mathbb{S}}$ is the unique minimal solution to Equation (5.4.18).

Hint: Letting $\tilde{h}$ denote another solution of (5.4.18), show, using (5.4.16) and induction on $n \geqslant 1$, that

$$\tilde{h}_{i,j} \geqslant \sum_{l=1}^{n} g_{i,j}^{(l)}, \qquad i,j \in \mathbb{S}, \quad n \geqslant 1.$$

Problem 5.36 Probabilistic automata (Gusev (2014)). Given an alphabet $\Sigma = \{a, b\}$ made of two letters we denote by Σ^* the set of all words over $\{a, b\}$, *i.e.* Σ^* is made of all finite sequences of symbols in $\{a, b\}$.

A *language* $\mathcal{L}$ over $\{a, b\}$ is a collection of (finite) words in Σ^*.

In computer science, an *automaton* given by a function

$$f : \{a, b\} \times \{0, 1, \ldots, n\} \longrightarrow \{0, 1, \ldots, n\}$$

is reading words of the form $a_1 a_2 \cdots a_m \in \mathcal{L}$ by producing a sequence $x_1, x_2, \ldots, x_m$ of integers *via* the recursion

$$x_1 := f(a_1, x_0),\ x_2 := f(a_2, x_1),\ x_3 := f(a_3, x_2), \ldots, x_m := f(a_m, x_{m-1}).$$

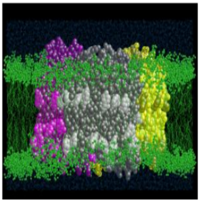
A word $a_1 a_2 \cdots a_m \in \mathcal{L}$, $m \geqslant 1$, is said to *synchronize* the automaton f *to state* ⓝ if we have $x_m =$ⓝ. Here, ⓝ is regarded as a *sink state*, also called *accepting state*.

One says that the automaton f *recognizes* the language $\mathcal{L}$ if every word $a_1 a_2 \cdots a_m \in \mathcal{L}$ synchronizes the automaton f *to state* ⓝ, *i.e.* satisfies $x_m =$ ⓝ, starting from *any* initial state (x_0).

Synchronizing automata are connected to algebra and combinatorics, and they have applications in many areas including robotics, coding theory, network security management, chip design, industrial automation, biocomputing, etc.

a) Let $n = 5$ and consider the automaton given by the function





f	0	1	2	3	4	5
a	1	2	3	3	1	5
b	0	0	0	4	5	5

Draw a graph for this automaton and find the smallest integers $l, m \geqslant 1$ such that f can recognize the word $a^l b^m$ starting from *any* state of the automaton.

b) Determine the language recognized by the automaton and give an example of a five-letter word that cannot synchronize the automaton when it starts from state $\textcircled{4}$.

Hint: The set of words recognized by the automaton can be written using the notation $\Sigma^* xxxxx \Sigma^*$ which represents the concatenations of a word in Σ^* followed by a certain word $xxxxx$ (always the same, to be determined), followed by another word in Σ^*.

c) Let $(X_k)_{k\geqslant 1}$ be a sequence of *i.i.d.* random variables taking values in $\{a, b\}$, with $\mathbb{P}(X_k = a) = p \in (0,1)$ and $\mathbb{P}(X_k = b) = q = 1 - p$. Consider the random process $(Y_k)_{k\geqslant 0}$ started at Y_0, with

$$Y_1 = f(X_1, Y_0), \ Y_2 = f(X_2, Y_1), \ldots, Y_k = f(X_k, Y_{k-1}), \ldots$$

Show that the process $(Y_k)_{k\geqslant 0}$ is a Markov chain on the state space $\{0, 1, 2, 3, 4, 5\}$, and write down its graph and its transition matrix.

d) Find the average time it takes until the automaton f becomes synchronized by the random words generated from $(X_k)_{k\geqslant 1}$, starting from the initial state $Y_0 = \textcircled{0}$.

Hint: Call $h_5(k)$ the average time it takes to reach state $\textcircled{5}$ starting from state $k = 0, 1, 2, 3, 4, 5$, and check first that $h_5(4) = p h_5(0)$.

e) (10 points) Let now $n = 4$ and consider the automaton g defined by

g	0	1	2	3	4
a	0	2	2	1	4
b	0	0	3	4	4

Draw a graph for this automaton.

f) Find the unique shortest word that synchronizes the automation to state (0) after starting from all states $1, 2, 3$, and the unique shortest word that synchronizes the automaton to state (4) after starting from all states $1, 2, 3$.

g) Consider the random process $(Z_k)_{k\geqslant 0}$ started at Z_0, with

$$Z_1 = g(X_1, Z_0),\ Z_2 = g(X_2, Z_1), \ldots, Z_k = g(X_k, Z_{k-1}), \ldots$$

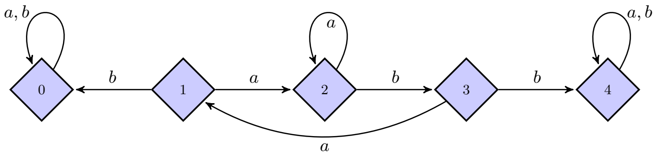
Show that the process $(Z_k)_{k\geqslant 0}$ is a Markov chain on the state space $\{0, 1, 2, 3, 4\}$, draw the graph of the chain and write down its transition matrix.

h) Find the probability that the first synchronized word is "abab" when the automaton is started from state (1). Note that according to Question (f), synchronization may here occur through state (0) or through state (4).

Exercise 5.37 Consider the probabilistic automaton g defined by

g	0	1	2	3	4
a	0	2	2	1	4
b	0	0	3	4	4

This automaton has two sink states (0) and (4), and its graph is given as follows:



a) Find the shortest word that synchronizes this automaton to state (4) after starting from any of the states $1, 2, 3$.

b) Consider the $\{a, b\}$-valued two-state Markov chain $(X_n)_{n\geqslant 0}$ with *transition probability matrix*

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

and the Markov chain on $(Z_k)_{k\geqslant 0}$ the state space $\{0, 1, 2, 3, 4\}$ started at Z_0, with

$$Z_1 = g(X_1, Z_0),\ Z_2 = g(X_2, Z_1), \ldots,\ Z_k = g(X_k, Z_{k-1}), \ldots$$

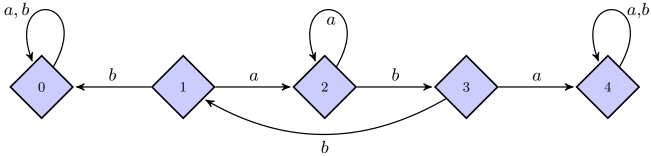
Draw the graph of the chain $(Z_k)_{k\geqslant 0}$ and write down its transition probability matrix.

c) Find the probability that the first synchronized word is "aabb" when the automaton is started from state ①.

Exercise 5.38 Consider the probabilistic automaton g defined by

g	0	1	2	3	4
a	0	2	2	4	4
b	0	0	3	1	4

This automaton has two sink states ⓪ and ④, and its graph is given as follows:



a) Find the unique shortest word that synchronizes this automaton to state ④ after starting from any of the states $1, 2, 3$.

b) We assume that letters are generated from an $\{a, b\}$-valued two-state Markov chain $(X_n)_{n\geqslant 0}$ with the *transition probability matrix*

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

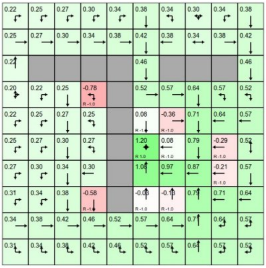
Find the probability that the first synchronized word is "aba" when the automaton is started from state ①.

Problem 5.39 *Markov Decision Processes* (MDPs) have applications in game theory, robotics, automated control, operations research, information theory, multi-agent systems, swarm intelligence, genetic algorithms, etc, through their use in *reinforcement learning*, and *Q-learning*, see *here* for a GridWorld-based algorithmic simulation.

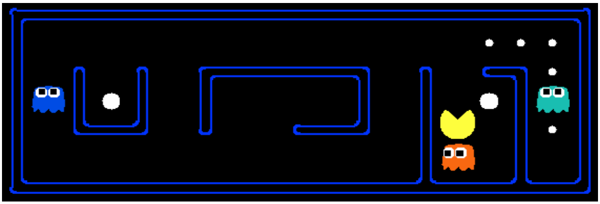
A *Markov Decision Process* (MDP) with state space $\mathbb{S}$ consists in:

- a *finite* set $\mathbb{A}$ of possible *actions*,
- a family $\big(P^{(a)}\big)_{a\in\mathbb{A}}$ of *transition probability matrices* $\big(P^{(a)}_{i,j}\big)_{i,j\in\mathbb{S}}$, and
- a *policy* $\pi : \mathbb{S} \to \mathbb{A}$ giving the action $\pi(k) \in \mathbb{A}$ taken at every given state in $k \in \mathbb{S}$.

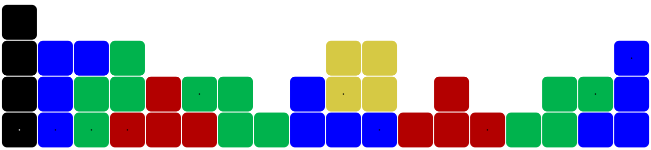
When a MDP is in state $X_n = k$ at time n we look up the action $\pi(k) \in \mathbb{A}$ given by the policy π, and we generate the new value X_{n+1} using the transition probabilities $P^{(\pi(k))}_{k,\cdot} = \big(P^{(\pi(k))}_{k,l}\big)_{l\in\mathbb{S}}$.



The Pacman game is a natural application of Markov Decision Processes, see ***here***.



The Tetris game can also be modeled as a Markov decision process whose (random) state consists in a given board configuration together with one of seven tile shapes. The actions are naturally chosen among the 40 placement choices for the falling tile, and a new tile shape is chosen with uniform probability 1/7 at each time step.



The following 10 questions are *interdependent*, and should be treated in sequence.

a) Consider a Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S}$ and transition matrix $P = (P_{i,j})_{i,j\in\mathbb{S}}$. Derive the first step analysis equation for the value function

$$V(k) := \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \,\middle|\, X_0 = k\right], \qquad k \in \mathbb{S}, \tag{5.4.19}$$

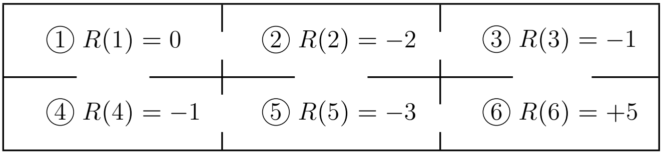
defined as the total accumulated reward obtained after starting from state ⓚ, where $R : \mathbb{S} \to \mathbb{R}$ is a reward function.*

b) In **Questions (b1)**, **(b2)** and **(b3)** below we consider the deterministic MDP on the state space $\mathbb{S} = \{1, 2, 3, 4, 5, 6, 7\}$ with actions $\mathbb{A} = \{\downarrow, \rightarrow\}$ and transition probability matrices

$$P^{(\downarrow)} := \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad P^{(\rightarrow)} := \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and the reward function $R : \mathbb{S} \to \mathbb{R}$ given by

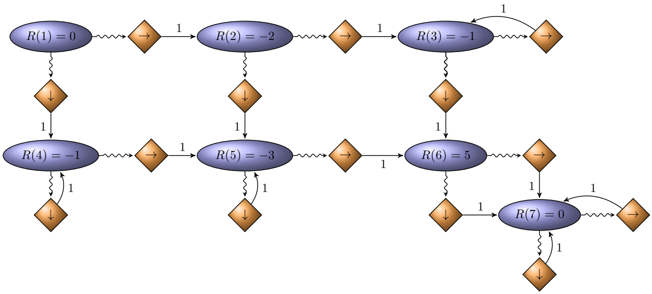
$$R(1) = 0,\ R(2) = -2,\ R(3) = -1,\ R(4) = -1,\ R(5) = -3,\ R(6) = 5. \tag{5.4.20}$$



This MDP can be represented by the following graph, in which the ⇝ arrows represent the policy choices, while the straight arrows denote Markov transitions.

---

* We always assume that $R(\cdot)$ and $(X_n)_{n\geqslant 0}$ are such that the series in (5.4.19) are convergent.



b1) Compute the *optimal action-value* function*

$$Q^*(k,\downarrow) := \underset{\pi\,:\,\pi(k)=\downarrow}{\mathrm{Max}} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right] \tag{5.4.21}$$

and

$$Q^*(k,\rightarrow) := \underset{\pi\,:\,\pi(k)=\rightarrow}{\mathrm{Max}} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right], \tag{5.4.22}$$

starting from state $X_0 = k \in \mathbb{S}$, in the following order: $Q^*(7,\downarrow)$, $Q^*(7,\rightarrow)$, $Q^*(6,\downarrow)$, $Q^*(6,\rightarrow)$, $Q^*(3,\downarrow)$, $Q^*(3,\rightarrow)$, $Q^*(5,\rightarrow)$, $Q^*(5,\downarrow)$, $Q^*(2,\downarrow)$, $Q^*(2,\rightarrow)$, $Q^*(4,\rightarrow)$, $Q^*(4,\downarrow)$, $Q^*(1,\downarrow)$, $Q^*(1,\rightarrow)$.

b2) Find the *optimal value* function

$$V^*(k) := \underset{\pi}{\mathrm{Max}}\, \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right],$$

at all states $k = 1, 2, \ldots, 7$.

Remark: The R package MDPtoolbox can be used to check your results, however, explanations are required.

b3) Find the optimal policy $\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5))$ of actions leading to the optimal gain starting from any state.†

Remark: The R package MDPtoolbox can be used to check your results, however, explanations are required.

* In the maxima (5.4.21) the action is taken equal to "↓", resp. "→" at the first step only.

† The values of $\pi^*(6)$ and $\pi^*(7)$ are not considered here because they do not affect the total reward.

c) **Questions (c1), (c2), (c3)** below are for a general MDP.

c1) Using first step analysis, show that for a general MDP the *optimal action-value* function*

$$Q^*(k,a) := \underset{\pi\,:\,\pi(k)=a}{\mathrm{Max}}\ \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n)\ \middle|\ X_0 = k\right], \qquad k \in \mathbb{S}, \quad a \in \mathbb{A}, \tag{5.4.23}$$

can be written using the transition probability matrix $P^{(a)}$ and the optimal value function $V^*(\cdot)$.†

c2) How can the optimal policy $\pi^*(k)$ at state $k \in \mathbb{S}$ be computed from the optimal action-value function $Q^*(k,a)$?

c3) By applying first step analysis to the *optimal value* function

$$V^*(k) := \underset{\pi}{\mathrm{Max}}\ \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n)\ \middle|\ X_0 = k\right], \qquad k \in \mathbb{S}, \tag{5.4.24}$$

derive the *Bellman equation* satisfied by the value function $V(k)$ using the *optimal action-value* function $Q^*(k,a)$.

d) In **Questions (d1), (d2)** and **(d3)** below we let $p \in [0,1]$ and consider the stochastic MDP on the state space $\mathbb{S} = \{1,2,3,4,5,6,7\}$, with actions $\mathbb{A} = \{\downarrow, \rightarrow\}$ and transition probability matrices

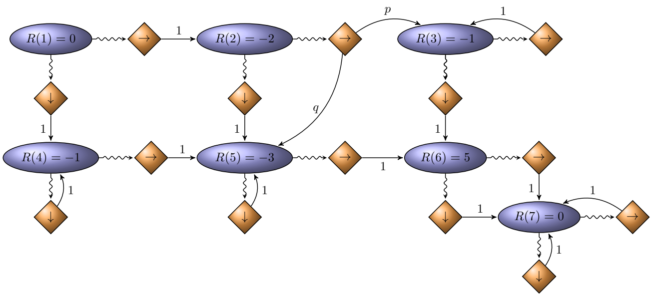
$$P^{(\downarrow)} := \begin{bmatrix} 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&1&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1 \end{bmatrix} \qquad P^{(\rightarrow)} := \begin{bmatrix} 0&1&0&0&0&0&0\\ 0&0&p&0&q&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&1&0\\ 0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1 \end{bmatrix},$$

and the reward function (5.4.20). This MDP can be represented by the following graph, in which the ⇝ arrows represent policy choices, while the straight arrows denote Markov transitions.

---

* In the maxima (5.4.23) the action is taken equal to a at the first step only. After moving to a new state we maximize the future reward according to the best policy choice.

† We always assume that $R(\cdot)$ and $(X_n)_{n\geqslant 0}$ are such that the series in (5.4.23) and (5.4.24) is convergent.



d1) Using the argument of Question (c1), compute the *optimal action-value* function*

$$Q^*(k,\downarrow) := \underset{\pi\,:\,\pi(k)=\downarrow}{\mathrm{Max}} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right] \tag{5.4.25}$$

and

$$Q^*(k,\rightarrow) := \underset{\pi\,:\,\pi(k)=\rightarrow}{\mathrm{Max}} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right], \tag{5.4.26}$$

starting from state $X_0 = k \in \mathbb{S}$, in the following order: $Q^*(7,\downarrow)$, $Q^*(7,\rightarrow)$, $Q^*(6,\downarrow)$, $Q^*(6,\rightarrow)$, $Q^*(3,\downarrow)$, $Q^*(3,\rightarrow)$, $Q^*(5,\rightarrow)$, $Q^*(5,\downarrow)$, $Q^*(2,\downarrow)$, $Q^*(2,\rightarrow)$, $Q^*(4,\rightarrow)$, $Q^*(4,\downarrow)$, $Q^*(1,\downarrow)$, $Q^*(1,\rightarrow)$.

Remark: Some values of $Q^*(k,\downarrow)$, $Q^*(k,\rightarrow)$ may now depend on p.

d2) Using the result of Question (d1), find the *optimal value* function

$$V^*(k) := \underset{\pi}{\mathrm{Max}}\, \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right],$$

at all states $k = 1, 2, \ldots, 7$, depending on the value of $p \in [0, 1]$.

Remark: The R package MDPtoolbox can be used to check your results, however, explanations are required.

d3) Find the optimal policy $\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5))$ of actions leading to the optimal gain starting from any state, depending on the value of $p \in [0, 1]$.†

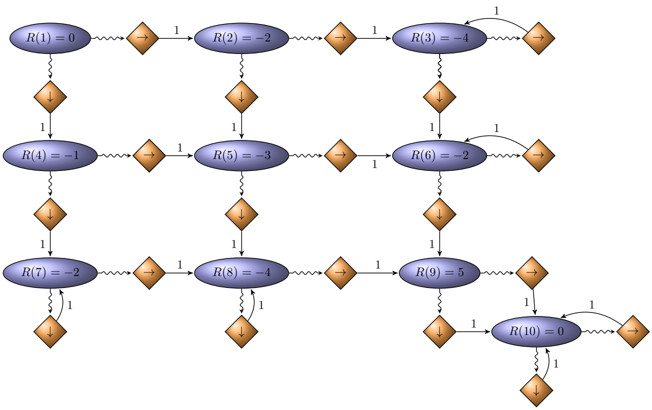
* In the maxima (5.4.25) the action is taken equal to "↓", resp. "→" at the first step only.

† The values of $\pi^*(6)$ and $\pi^*(7)$ are not considered because they do not affect the total reward.

Remark: The R package MDPtoolbox can be used to check your results, however, explanations are required.

```
install.packages("MDPtoolbox")
library(MDPtoolbox)
P <- array(0,c(7,7,2))
P[,,1] <- matrix(c(.,...,.,.,...,.), nrow=7, ncol=7, byrow=TRUE)
P[,,2] <- matrix(c(.,...,.,.,...,.), nrow=7, ncol=7, byrow=TRUE)
R <- array(0,c(7,2))
R[,1] <- matrix(c(.,...,.), nrow=1, ncol=7, byrow=TRUE)
R[,2] <- R[,1]
mdp_check(P, R)
mdp_value_iteration(P,R,discount=1)
```

Exercise 5.40 We consider the deterministic Markov Decision Process (MDP) on the state space $\mathbb{S} = \{1, 2, \ldots, 10\}$ with actions $\mathbb{A} = \{\downarrow, \rightarrow\}$ and reward function $R : \mathbb{S} \to \mathbb{R}$ represented in the following graph.



a) Compute the optimal action-value functional $Q^*(k, a)$, $k = 1, 2, \ldots, 9$, $a \in \{\rightarrow, \downarrow\}$.
b) Compute the optimal value function $V^*(k)$ for $k = 1, 2, \ldots, 9$.
c) Compute the optimal policy $\pi^*(k) \in \{\rightarrow, \downarrow\}$ for $k = 1, 2, \ldots, 9$.

Chapter 6
Classification of States

In this chapter we present the notions of communicating, transient and recurrent states, as well as the concept of irreducibility of a Markov chain. We also examine the notions of positive and null recurrence, periodicity, and aperiodicity of such chains. Those topics will be important when analysing the long-run behavior of Markov chains in the next chapter.

6.1 Communicating States

Definition 6.1. *A state* $\textcircled{j} \in \mathbb{S}$ *is to be* accessible *from another state* $\textcircled{i} \in \mathbb{S}$*, and we write* $\textcircled{i} \longmapsto \textcircled{j}$*, if there exists a* finite *integer* $n \geqslant 0$ *such that*

$$[P^n]_{i,j} = \mathbb{P}(X_n = j \mid X_0 = i) > 0.$$

In other words, it is possible to travel from $\textcircled{i}$ to $\textcircled{j}$ with non-zero probability in a certain (random) number of steps. We also say that state ⓘ leads to state $\textcircled{j}$, and when $i \neq j$ we have

$$\mathbb{P}(T_j^r < \infty \mid X_0 = i) \geqslant \mathbb{P}(T_j^r \leqslant n \mid X_0 = i) \geqslant \mathbb{P}(X_n = j \mid X_0 = i) > 0.$$

Remark 6.2. *Since $P^0 = \mathrm{I}$ and $[P^0]_{i,j} = \mathbb{P}(X_0 = j \mid X_0 = i) = \mathbb{1}_{\{i=j\}}$ the definition of accessibility states implicitly that any state (i) is always accessible from itself (in zero time steps) even if $P_{i,i} = 0$.*

In case $(\mathrm{i}) \longmapsto (j)$ and $(j) \longmapsto (\mathrm{i})$ we say that (i) and (j) *communicate*[*] and we write $(\mathrm{i}) \longleftrightarrow (j)$.

The binary relation "$\longleftrightarrow$" is a called an equivalence relation as it satisfies the following properties:[†]

a) Reflexivity:

   For all $i \in \mathbb{S}$ we have $(\mathrm{i}) \longleftrightarrow (\mathrm{i})$.

b) Symmetry:

   For all $i, j \in \mathbb{S}$ we have that $(\mathrm{i}) \longleftrightarrow (j)$ is equivalent to $(j) \longleftrightarrow (\mathrm{i})$.

c) Transitivity:

   For all $i, j, k \in \mathbb{S}$ such that $(\mathrm{i}) \longleftrightarrow (j)$ and $(j) \longleftrightarrow (\mathrm{k})$, we have $(\mathrm{i}) \longleftrightarrow (\mathrm{k})$.

Proof. It is clear that the relation $\longleftrightarrow$ is reflexive and symmetric. The proof of transitivity can be stated as follows. If $(\mathrm{i}) \longmapsto (j)$ and $(j) \longmapsto (\mathrm{k})$, there exists $a \geqslant 1$ and $b \geqslant 1$ such that

$$[P^a]_{i,j} > 0, \qquad [P^b]_{j,k} > 0.$$

Next, by the Markov property and (4.1.2), for all $n \geqslant a + b$ we have

$$\begin{aligned}
&\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{l,m \in \mathbb{S}} \mathbb{P}(X_n = k,\ X_{n-b} = l,\ X_a = m \mid X_0 = i) \\
&= \sum_{l,m \in \mathbb{S}} \mathbb{P}(X_n = k \mid X_{n-b} = l)\mathbb{P}(X_{n-b} = l \mid X_a = m)\mathbb{P}(X_a = m \mid X_0 = i) \\
&\geqslant \mathbb{P}(X_n = k \mid X_{n-b} = j)\mathbb{P}(X_{n-b} = j \mid X_a = j)\mathbb{P}(X_a = j \mid X_0 = i) \\
&= [P^a]_{i,j}[P^{n-a-b}]_{j,j}[P^b]_{j,k} \qquad (6.1.1) \\
&\geqslant 0.
\end{aligned}$$

The conclusion follows by taking $n = a + b$, in which case we have

$$\mathbb{P}(X_n = k \mid X_0 = i) \geqslant [P^a]_{i,j}[P^b]_{j,k} > 0.$$

□

[*] In graph theory, one says that (i) and (i) are *strongly connected*.

[†] Please refer to MH1300 - Foundations of Mathematics for more information on equivalence classes.

The equivalence relation '$\longleftrightarrow$" induces a *partition* of $\mathbb{S}$ into disjoint classes $A_1, A_2, \ldots, A_m$ such that $\mathbb{S} = A_1 \cup \cdots \cup A_m$, and

a) we have $\textcircled{i} \longleftrightarrow \textcircled{j}$ for all $i, j \in A_q$, and

b) we have $\textcircled{i} \not\longleftrightarrow \textcircled{j}$ whenever $i \in A_p$ and $j \in A_q$ with $p \neq q$.

The sets $A_1, A_2, \ldots, A_m$ are called the *communicating classes* of the chain.

Definition 6.3. *A Markov chain whose state space is made of a unique communicating class is said to be* irreducible*, otherwise the chain is said to be* reducible*.*

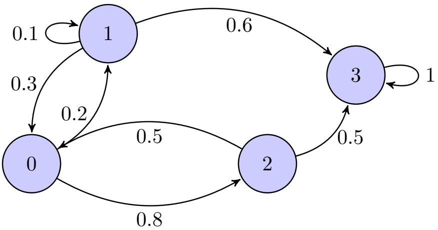
In particular, a Markov chain is irreducible if its transition matrix P is *regular*, *i.e.* if there exists $n \geqslant 1$ such that all entries of the matrix power P^n are strictly positive.

The R package "markovchain" can be used to test the irreducibility of a given chain.

```
install.packages("markovchain")
library (markovchain)
statesNames <- c("0", "1")
mcA <- new("markovchain",transitionMatrix = matrix(c(0.7,0.3,0.1,0.9),
byrow = TRUE, nrow = 2,dimnames = list(statesNames, statesNames)))
is.irreducible(mcA)
```

Clearly, all states in $\mathbb{S}$ communicate when $(X_n)_{n\geqslant 0}$ is irreducible. In case the *i-th* column of a transition matrix P vanishes, *i.e.* $P_{k,i} = 0$, $k \in \mathbb{S}$, then state $\textcircled{i}$ cannot be reached from any other state and $\textcircled{i}$ becomes a communicating class on its own, as is the case of state $\textcircled{1}$ in Exercise 4.10 for $n \geqslant 2$, or in Exercise 7.13. The same is true of absorbing states. However, having a returning loop with probability strictly lower than one is not sufficient to turn a given state into a communicating class on its own. Clearly, the existence of at least one absorbing state $\textcircled{i}$ with $P_{i,i} = 1$ makes a chain reducible.

Exercise: Find the communicating classes of the Markov chain with transition matrix (5.4.10) for the equivalence relation "$\longleftrightarrow$".



$$P = \begin{bmatrix} 0 & 0.2 & 0.8 & 0 \\ 0.3 & 0.1 & 0 & 0.6 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The above state space $\mathbb{S} = \{0, 1, 2, 3\}$ is partitioned into two communicating classes which are $\{0, 1, 2\}$ and $\{3\}$.

6.2 Recurrent States

Definition 6.4. *A state* $\textcircled{i} \in \mathbb{S}$ *is said to be* recurrent *if, starting from state* $\textcircled{i}$*, the chain* $(X_n)_{n\geqslant 0}$ *will return to state* $\textcircled{i}$ *within a finite (random) time, with probability* 1*, i.e.,*

$$p_{ii} := \mathbb{P}(T_i^r < \infty \mid X_0 = i) = \mathbb{P}(X_n = i \text{ for some } n \geqslant 1 \mid X_0 = i) = 1. \tag{6.2.1}$$

The next Proposition 6.5 uses the mean number of returns R_i to state $\textcircled{i}$ defined in (5.4.2), and relies on the geometric distribution (5.4.3) of R_i given that $X_0 = i$.

Proposition 6.5. *For any state* $\textcircled{i} \in \mathbb{S}$*, the following statements are equivalent:*

i) the state $\textcircled{i} \in \mathbb{S}$ *is recurrent, i.e.* $p_{ii} = 1$,
ii) the number of returns to $\textcircled{i} \in \mathbb{S}$ *is* a.s.* *infinite, i.e.*

$$\mathbb{P}(R_i = \infty \mid X_0 = i) = 1, \;\; \textit{i.e.} \;\; \mathbb{P}(R_i < \infty \mid X_0 = i) = 0, \tag{6.2.2}$$

iii) the mean number of returns to $\textcircled{i} \in \mathbb{S}$ *is* infinite*, i.e.*

$$\mathbb{E}[R_i \mid X_0 = i] = \infty, \tag{6.2.3}$$

iv) we have

$$\sum_{n\geqslant 1} f_{i,i}^{(n)} = 1, \tag{6.2.4}$$

where $f_{i,i}^{(n)} := \mathbb{P}(T_i^r = n \mid X_0 = i)$, $n \geqslant 1$*, is the distribution of* T_i^r*.*

Proof. Question (i) follows by the definition (6.2.1) of recurrent states.

ii) Relation (6.2.2) is equivalent to (6.2.1) by (5.4.4) and (5.4.5).
iii) Relation (6.2.3) is equivalent to (6.2.1) by (5.4.6).
iv) Relation (6.2.4) is equivalent to (6.2.1) by (5.4.1).

□

* "Almost surely", *i.e.* with probability one.

For example, state ⓪ is recurrent for the random walk of Chapter 3 when $p = q = 1/2$, while it is not recurrent if $p \neq q$ as by (3.4.14) we have

$$p_{0,0} = \mathbb{P}(T_0 < \infty) = 2\min(p,q). \tag{6.2.5}$$

As a consequence of (6.2.3), we have the following result.

Corollary 6.6. *A state $i \in \mathbb{S}$ is recurrent if and only if*

$$\sum_{n\geqslant 1}[P^n]_{i,i} = \infty,$$

i.e. the above series diverges*.*

Proof. For all $i, j \in \mathbb{S}$, by (5.4.9) we have

$$\begin{aligned}\mathbb{E}[R_j \mid X_0 = i] &= \mathbb{E}\left[\sum_{n\geqslant 1}\mathbb{1}_{\{X_n=j\}} \;\middle|\; X_0 = i\right] = \sum_{n\geqslant 1}\mathbb{E}[\mathbb{1}_{\{X_n=j\}} \mid X_0 = i] \\ &= \sum_{n\geqslant 1}\mathbb{P}(X_n = j \mid X_0 = i) = \sum_{n\geqslant 1}[P^n]_{i,j}, \end{aligned} \tag{6.2.6}$$

as in (5.4.9). To conclude, we let $j = i$ and apply (6.2.3). □

Corollary 6.6 admits the following consequence, which shows that any state communicating with a recurrent state is itself recurrent. In other words, recurrence is a *class property*, as all states in a given communicating class are recurrent as soon as one of them is recurrent.

Corollary 6.7. *(Class property). Let ⓙ $\in \mathbb{S}$ be a recurrent state. Then any state ⓘ $\in \mathbb{S}$ that communicates with state ⓙ is also recurrent.*

Proof. By definition, since ⓘ $\longmapsto$ ⓙ and ⓙ $\longmapsto$ ⓘ, there exists $a \geqslant 1$ and $b \geqslant 1$ such that

$$[P^a]_{i,j} > 0 \qquad\qquad [P^b]_{j,i} > 0,$$

and from (6.1.1) applied with $k = i$ we find

$$\begin{aligned}\sum_{n\geqslant a+b}[P^n]_{i,i} &= \sum_{n\geqslant a+b}\mathbb{P}(X_n = i \mid X_0 = i) \\ &\geqslant [P^a]_{i,j}[P^b]_{j,i}\sum_{n\geqslant a+b}[P^{n-a-b}]_{j,j} \\ &= [P^a]_{i,j}[P^b]_{j,i}\sum_{n\geqslant 0}[P^n]_{j,j} \\ &= \infty,\end{aligned}$$

which shows that state $\textcircled{i}$ is recurrent from Corollary 6.6 and the assumption that state $\textcircled{j}$ is recurrent. □

A communicating class $A \subset \mathbb{S}$ is therefore recurrent if any of its states is recurrent.

6.3 Transient States

A state $\textcircled{i} \in \mathbb{S}$ is said to be *transient* when it is not recurrent, *i.e.*, by (6.2.1),

$$p_{ii} = \mathbb{P}(T_i^r < \infty \mid X_0 = i) = \mathbb{P}(X_n = i \text{ for some } n \geqslant 1 \mid X_0 = i) < 1, \tag{6.3.1}$$

or

$$\mathbb{P}(T_i^r = \infty \mid X_0 = i) > 0.$$

Proposition 6.8. *For any state $\textcircled{i} \in \mathbb{S}$, the following statements are equivalent:*

i) the state $\textcircled{i} \in \mathbb{S}$ is transient, i.e. $p_{ii} < 1$,
ii) the number of returns to $\textcircled{i} \in \mathbb{S}$ is a.s.* *finite, i.e.*

$$\mathbb{P}(R_i = \infty \mid X_0 = i) = 0, \text{ i.e. } \mathbb{P}(R_i < \infty \mid X_0 = i) = 1, \tag{6.3.2}$$

iii) the mean number of returns to $\textcircled{i} \in \mathbb{S}$ is finite, *i.e.*

$$\mathbb{E}[R_i \mid X_0 = i] < \infty, \tag{6.3.3}$$

iv) we have

$$\sum_{n \geqslant 1} f_{i,i}^{(n)} < 1. \tag{6.3.4}$$

where $f_{i,i}^{(n)} := \mathbb{P}(T_i^r = n \mid X_0 = i)$, $n \geqslant 1$, is the distribution of T_i^r.

Proof. This is a direct consequence of Proposition 6.5 and the definition (6.3.1) of transience. Regarding point (ii) and the Condition (6.3.2) we also note that the state $\textcircled{i} \in \mathbb{S}$ is transient if and only if

$$\mathbb{P}(R_i = \infty \mid X_0 = i) < 1,$$

which, by (5.4.5) is equivalent to $\mathbb{P}(R_i = \infty \mid X_0 = i) = 0$. □

In other words, a state $\textcircled{i} \in \mathbb{S}$ is *transient* if and only if

$$\mathbb{P}(R_i < \infty \mid X_0 = i) > 0,$$

* "Almost surely", *i.e.* with probability one.

which by (5.4.4) is equivalent to

$$\mathbb{P}(R_i < \infty \mid X_0 = i) = 1,$$

i.e. the number of returns to state $i \in \mathbb{S}$ is finite with a non-zero probability which is necessarily equal to one. As a consequence of Corollary 6.6 we also have the following result.

Corollary 6.9. *A state $i \in \mathbb{S}$ is transient if and only if*

$$\sum_{n \geqslant 1} [P^n]_{i,i} < \infty,$$

i.e. the above series converges*.*

Similarly to Corollary 6.7, Corollary 6.9 admits the following consequence, which shows that any state communicating with a transient state is itself transient. Therefore, transience is also a *class property*, as all states in a given communicating class are transient as soon as one of them is transient.

Corollary 6.10. *(Class property). Let $\textcircled{j} \in \mathbb{S}$ be a transient state. Then any state $\textcircled{i} \in \mathbb{S}$ that communicates with state $\textcircled{j}$ is also transient.*

Proof. If a state $\textcircled{i} \in \mathbb{S}$ communicates with a transient state $\textcircled{j}$ then $\textcircled{i}$ is also transient (otherwise the state $\textcircled{j}$ would be recurrent by Corollary 6.7). □

A communicating class $A \subset \mathbb{S}$ is therefore transient if any of its states is transient. By Corollary 6.9 and the relation

$$\sum_{n \geqslant 0} [P^n]_{i,j} = \left[(I - P)^{-1}\right]_{i,j}, \qquad i, j \in \mathbb{S}, \tag{6.3.5}$$

see (5.4.9), we find that a chain with finite state space cannot be transient as the matrix $I - P$ is invertible in this case, cf. Problem 4.14. Indeed, 0 is clearly an eigenvalue of $I - P$ with eigenvector $[1, 1, \ldots, 1]$, therefore $I - P$ is not invertible, and a finite state space chain always admits at least one recurrent state, as noted in Theorem 6.11 below.

Clearly, any absorbing state is recurrent, and any state that leads to an absorbing state is transient.

Example

For the two-state Markov chain of Section 4.5, Relations (4.5.4)-(4.5.5) show that

$$\sum_{n \geqslant 1} [P^n]_{0,0} = \sum_{n \geqslant 1} \frac{b + a\lambda_2^n}{a + b} = \begin{cases} \infty, & \text{if } b > 0, \\ \displaystyle\sum_{n \geqslant 1} (1 - a)^n < \infty, & \text{if } b = 0 \text{ and } a > 0, \end{cases}$$

hence state ⓪ is transient if $b = 0$ and $a > 0$, and recurrent otherwise. Similarly, we have

$$\sum_{n\geqslant 1}[P^n]_{1,1} = \sum_{n\geqslant 1}\frac{a+b\lambda_2^n}{a+b} = \begin{cases} \infty, & \text{if } a > 0, \\ \displaystyle\sum_{n\geqslant 1}(1-b)^n < \infty, & \text{if } a = 0 \text{ and } b > 0, \end{cases}$$

hence state ① is transient if $a = 0$ and $b > 0$, and recurrent otherwise.

The above results can be recovered by a simple first step analysis for $g_i(j) = \mathbb{P}(T_i^r < \infty \mid X_0 = j)$, $i, j \in \{0, 1\}$, *i.e.*

$$\begin{cases} g_0(0) = a g_0(1) + 1 - a \\ g_0(1) = b + (1-b) g_0(1) \\ g_1(0) = (1-a) g_1(0) + a \\ g_1(1) = b g_1(0) + 1 - b, \end{cases}$$

which shows that $g_0(0) = 1$ if $b > 0$ and $g_1(1) = 1$ if $a > 0$.

We close this section with the following result for Markov chains with finite state space.

Theorem 6.11. *Let $(X_n)_{n\geqslant 0}$ be a Markov chain with* finite *state space* $\mathbb{S}$. *Then $(X_n)_{n\geqslant 0}$ has at least one recurrent state.*

Proof. Recall that from (5.4.6) we have

$$\mathbb{E}[R_j \mid X_0 = i] = p_{ij}(1-p_{jj})\sum_{n\geqslant 1} n(p_{jj})^{n-1} = \frac{p_{ij}}{1-p_{jj}},$$

for any states ⓘ, ⓙ $\in \mathbb{S}$. Assuming that the state ⓙ $\in \mathbb{S}$ is transient we have $p_{jj} < 1$ by (6.3.1), hence by (6.2.6) we have

$$\mathbb{E}[R_j \mid X_0 = i] = \sum_{n\geqslant 1}[P^n]_{i,j} < \infty,$$

which is finite by (6.3.3), implying* that

$$\lim_{n\to\infty}[P^n]_{i,j} = 0$$

* For any sequence $(a_n)_{n\geqslant 0}$ of *nonnegative* real numbers, $\displaystyle\sum_{n\geqslant 0} a_n < \infty$ implies $\lim_{n\to\infty} a_n = 0$.

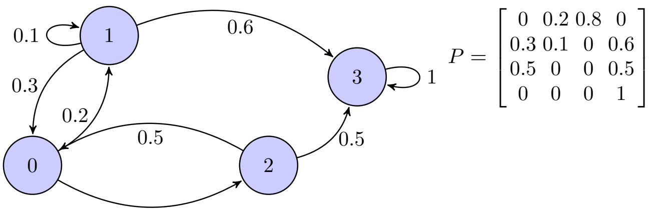
for all transient states $j \in \mathbb{S}$. In case all states in $\mathbb{S}$ were transient, since $\mathbb{S}$ is finite, by the *law of total probability* (4.2.2) we would have

$$0 = \sum_{j\in\mathbb{S}} \lim_{n\to\infty} [P^n]_{i,j} = \lim_{n\to\infty} \sum_{j\in\mathbb{S}} [P^n]_{i,j} = \lim_{n\to\infty} 1 = 1,$$

which is a contradiction. Hence not all states can be transient, and there exists at least one recurrent state. □

Exercises:

i) Find which states are transient and recurrent in the chain (5.4.10) represented as follows:



Answer: State ③ is clearly recurrent since we have $T_3^r = 1$ with probability one when $X_0 = 3$. State ② is transient because

$$1 - p_{2,2} = \mathbb{P}(T_2^r = \infty \mid X_0 = 2) = \frac{4}{7} \geqslant \mathbb{P}(X_1 = 3 \mid X_0 = 2) = 0.5 > 0, \tag{6.3.6}$$

and state ① is transient because

$$\mathbb{P}\big(T_1^r = \infty \mid X_0 = 1\big) = 0.8 \geqslant \mathbb{P}(X_1 = 3 \mid X_0 = 1) = 0.6, \tag{6.3.7}$$

see the Exercise 5.5 for the computations of

$$p_{1,1} = \mathbb{P}(T_1^r < \infty \mid X_0 = 1) = 0.8$$

and

$$p_{2,2} = \mathbb{P}(T_2^r < \infty \mid X_0 = 2) = \frac{3}{7}.$$

By Corollary 6.7, the states ⓪ and ① are transient because they communicate with state ②.

ii) Which are the recurrent states in the simple random walk $(S_n)_{n\geqslant 0}$ of Chapter 3 on $\mathbb{S} = \mathbb{Z}$?

Answer: First, we note that this random walk is irreducible as all states communicate when $p \in (0,1)$. The simple random walk $(S_n)_{n\geqslant 0}$ on $\mathbb{S} = \mathbb{Z}$ has the transition matrix

$$P_{i,i+1} = p, \qquad P_{i,i-1} = q = 1-p, \qquad i \in \mathbb{Z}.$$

We have

$$[P^n]_{i,i} = \mathbb{P}(S_n = i \mid S_0 = i) = \mathbb{P}(S_n = 0 \mid S_0 = 0),$$

with

$$\mathbb{P}(S_{2n} = 0) = \binom{2n}{n} p^n q^n \qquad \text{and} \qquad \mathbb{P}(S_{2n+1} = 0) = 0, \qquad n \in \mathbb{N}.$$

Hence

$$\begin{aligned}\sum_{n\geqslant 0} [P^n]_{0,0} &= \sum_{n\geqslant 0} \mathbb{P}(S_n = 0 \mid S_0 = 0) = \sum_{n\geqslant 0} \mathbb{P}(S_{2n} = 0 \mid S_0 = 0) \\ &= \sum_{n\geqslant 0} \binom{2n}{n} p^n q^n = H(1) = \frac{1}{\sqrt{1-4pq}},\end{aligned}$$

and

$$\mathbb{E}[R_0 \mid S_0 = 0] = \sum_{n\geqslant 1} \mathbb{P}(S_n = 0 \mid S_0 = 0) = \frac{1}{\sqrt{1-4pq}} - 1,$$

where $H(s)$ is defined in (3.4.8), cf. also Problem 3.18-(f).

Consequently, by Corollary 6.6, all states ⓘ $\in \mathbb{Z}$ are recurrent when $p = q = 1/2$, whereas by Corollary 6.9 they are all transient when $p \neq q$, cf. Corollary 6.7.

Alternatively we could reach the same conclusion by directly using (3.4.14) and (6.2.1) which state that

$$\mathbb{P}(T_i^r < \infty \mid X_0 = i) = 2\min(p,q).$$

6.4 Positive *vs.* Null Recurrence

The expected time of return (or mean recurrence time) to a state ⓘ $\in \mathbb{S}$ is given by

$$\begin{aligned}\mu_i(i) :&= \mathbb{E}[T_i^r \mid X_0 = i] \\ &= \sum_{n\geqslant 1} n \mathbb{P}\big(T_i^r = n \,\big|\, X_0 = i\big)\end{aligned}$$

$$= \sum_{n\geqslant 1} n f_{i,i}^{(n)}.$$

Recall that a state (i) is recurrent when $\mathbb{P}(T_i^r < \infty \mid X_0 = i) = 1$, *i.e.* when the random return time T_i^r is almost surely *finite* starting from state (i). However, the recurrence property yields no information on the finiteness of its expected value $\mu_i(i) = \mathbb{E}[T_i^r \mid X_0 = i]$, cf. the example (1.6.5).

Definition 6.12. *A* recurrent *state* $i \in \mathbb{S}$ *is said to be:*

a) positive recurrent *if the mean return time to* (i) *is* finite*, i.e.*

$$\mu_i(i) = \mathbb{E}[T_i^r \mid X_0 = i] < \infty,$$

b) null recurrent *if the mean return time to* (i) *is* infinite*, i.e.*

$$\mu_i(i) = \mathbb{E}[T_i^r \mid X_0 = i] = \infty.$$

Exercise: Which states are positive/null recurrent in the simple random walk $(S_n)_{n\geqslant 0}$ of Chapter 3 on $\mathbb{S} = \mathbb{Z}$?

From (3.4.20) and (3.4.17) we know that $\mathbb{E}[T_i^r \mid S_0 = i] = \infty$ for all values of $p \in (0, 1)$, hence all states of the random walk on $\mathbb{Z}$ are null recurrent when $p = 1/2$, while all states are transient when $p \neq 1/2$ due to (3.4.14).

The following Theorem 6.13 shows in particular that a Markov chain with finite state space cannot have any null recurrent state, cf. *e.g.* Corollary 2.3 in Kijima (1997), and also Corollary 3.7 in Asmussen (2003).

Theorem 6.13. *Assume that the state space* $\mathbb{S}$ *of a Markov chain* $(X_n)_{n\geqslant 0}$ *is* finite*. Then all recurrent states in* $\mathbb{S}$ *are also positive recurrent.*

As a consequence of Definition 6.3, Corollary 6.7, and Theorems 6.11 and 6.13 we have the following corollary.

Corollary 6.14. *Let* $(X_n)_{n\geqslant 0}$ *be an irreducible Markov chain with* finite *state space* $\mathbb{S}$*. Then* all *states of the chain* $(X_n)_{n\geqslant 0}$ *are positive recurrent.*

6.5 Periodicity and Aperiodicity

Given a state $i \in \mathbb{S}$, consider the sequence

$$\{n \geqslant 1 \ : \ [P^n]_{i,i} > 0\}$$

of integers which represent the possible travel times from state (i) to itself.

Definition 6.15. *The* period *of the state* $i \in \mathbb{S}$ *is the greatest common divisor*[*] *of the sequence*

$$\{n \geqslant 1 \ : \ [P^n]_{i,i} > 0\}.$$

A state $i \in \mathbb{S}$ having period 1 is said to be *aperiodic*. This is the case in particular when $P_{ii} > 0$, *i.e.* when $\textcircled{i}$ admits a returning loop with nonzero probability.

In particular, any absorbing state is both aperiodic and recurrent. A *recurrent* state $i \in \mathbb{S}$ is said to be *ergodic* if it is both *positive recurrent* and *aperiodic*.

If $[P^n]_{i,i} = 0$ for all $n \geqslant 1$ then the set $\{n \geqslant 1 \ : \ [P^n]_{i,i} > 0\}$ is empty and by convention the period of state $\textcircled{i}$ is defined to be 0. In this case, state $\textcircled{i}$ is also transient.

Note also that if

$$\{n \geqslant 1 \ : \ [P^n]_{i,i} > 0\}$$

contains two distinct numbers that are relatively prime to each other (*i.e.* their greatest common divisor is 1) then state $\textcircled{i}$ aperiodic.

Proposition 6.16 shows that periodicity is a *class property*, as all states in a given communicating class have same periodicity.

Proposition 6.16. *(Class property). All states that belong to a same communicating class have the same period.*

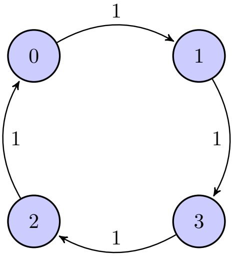
Proof. (See *e.g. here*). Assume that state $\textcircled{i}$ has period d_i, that $\textcircled{j}$ communicates with $\textcircled{i}$, and let $n \in \{m \geqslant 1 \ : \ [P^m]_{j,j} > 0\}$. Since $\textcircled{i}$ and $\textcircled{j}$ communicate, there exists $k, l \geqslant 1$ such that $[P^k]_{i,j} > 0$ and $[P^l]_{j,i} > 0$, hence by (6.1.1) we have $[P^{k+l}]_{i,i} > 0$ hence $k + l$ is a multiple of d_i. Similarly by (6.1.1) we also have $[P^{n+k+l}]_{i,i} > 0$, hence $n + k + l$ and n are multiples of d_i, which implies $d_j \geqslant d_i$. Exchanging the roles of $\textcircled{i}$ and $\textcircled{j}$ we obtain similarly that $d_i \geqslant d_j$. $\square$

A Markov chain is said to be *aperiodic* when *all* of its states are aperiodic. Note that any state that communicates with an aperiodic state becomes itself aperiodic. In particular, if a communicating class contains an aperiodic state then the whole class becomes aperiodic.

Examples

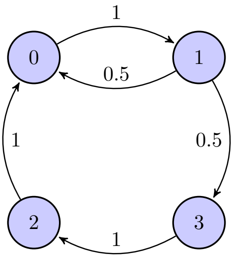
i) The chain

[*] Please refer to MH1300 - Foundations of Mathematics or MH1301 Discrete Mathematics for more information on greatest common divisors.



clearly has periodicity equal to 4.

ii) Consider the following chain:

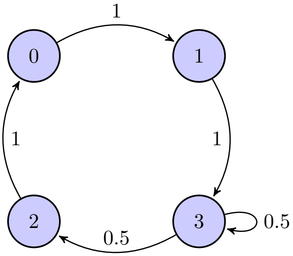


Here we have

$$\begin{aligned}
\{n \geqslant 1 \ : \ [P^n]_{0,0} > 0\} &= \{2, 4, 6, 8, 10, \ldots\}, \\
\{n \geqslant 1 \ : \ [P^n]_{1,1} > 0\} &= \{2, 4, 6, 8, 10, \ldots\}, \\
\{n \geqslant 1 \ : \ [P^n]_{2,2} > 0\} &= \{4, 6, 8, 10, 12, \ldots\}, \\
\{n \geqslant 1 \ : \ [P^n]_{3,3} > 0\} &= \{4, 6, 8, 10, 12, \ldots\},
\end{aligned}$$

hence *all* states have period 2, and this is also consequence of Proposition 6.16.

iii) Consider the following chain:



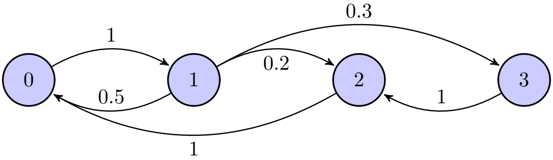
(6.5.1)

Here we have

$$\begin{aligned}\{n \geqslant 1 \ : \ [P^n]_{0,0} > 0\} &= \{4,5,6,7,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{1,1} > 0\} &= \{4,5,6,7,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{2,2} > 0\} &= \{4,5,6,7,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{3,3} > 0\} &= \{4,5,6,7,\ldots\},\end{aligned}$$

hence *all* states have period 1, see also Proposition 6.16.

iv) Next, consider the modification of (6.5.1):



Here the chain is aperiodic since we have

$$\begin{aligned}\{n \geqslant 1 \ : \ [P^n]_{0,0} > 0\} &= \{2,3,4,5,6,7,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{1,1} > 0\} &= \{2,3,4,5,6,7,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{2,2} > 0\} &= \{3,4,5,6,7,8,\ldots\},\\ \{n \geqslant 1 \ : \ [P^n]_{3,3} > 0\} &= \{4,6,7,8,9,10,\ldots\},\end{aligned}$$

hence *all* states have period 1.

Exercises:

i) What is the periodicity of the simple random walk $(S_n)_{n\geqslant 0}$ of Chapter 3 on $\mathbb{S} = \mathbb{Z}$?

Answer: By (3.3.3) We have

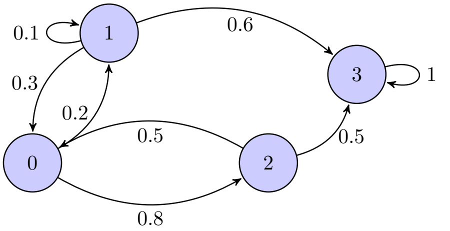
$$[P^{2n}]_{i,i} = \binom{2n}{n} p^n q^n > 0 \quad \text{and} \quad [P^{2n+1}]_{i,i} = 0, \qquad n \geqslant 0,$$

hence

$$\{n \geqslant 1 \ : \ [P^n]_{i,i} > 0\} = \{2, 4, 6, 8, \ldots\},$$

and the chain has period 2.

ii) Find the periodicity of the chain (5.4.10).



$$\begin{bmatrix} 0 & 0.2 & 0.8 & 0 \\ 0.3 & 0.1 & 0 & 0.6 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Answer: States $\textcircled{0}$, $\textcircled{1}$, $\textcircled{2}$ and $\textcircled{3}$ have period 1, hence the chain is aperiodic.

iii) The chain of Figure 4.3 is aperiodic since it is irreducible and state $\textcircled{3}$ has a returning loop.

Exercises

Exercise 6.1 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\{0, 1, 2, 3\}$, with transition matrix

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

a) Draw the graph of this chain and find its communicating classes. Is this Markov chain reducible? Why?
b) Find the periods of states $\textcircled{0}$, $\textcircled{1}$, $\textcircled{2}$, and $\textcircled{3}$.
c) Compute $\mathbb{P}(T_0 < \infty \mid X_0 = 0)$, $\mathbb{P}(T_0 = \infty \mid X_0 = 0)$, and $\mathbb{P}(R_0 < \infty \mid X_0 = 0)$.
d) Which state(s) is (are) absorbing, recurrent, and transient?

Exercise 6.2 Consider the Markov chain on $\{0, 1, 2\}$ with transition matrix

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

a) Is the chain irreducible? Give its communicating classes.
b) Which states are absorbing, transient, recurrent, positive recurrent?
c) Find the period of every state.

Exercise 6.3 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\{0,1,2,3,4\}$, with transition matrix

$$\begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) Draw the graph of this chain.
b) Find the periods of states ⓪, ①, ②, and ③.
c) Which state(s) is (are) absorbing, recurrent, and transient?
d) Is the Markov chain reducible? Why?

Exercise 6.4 Consider the Markov chain $(X_n)_{n\geqslant 0}$ with transition matrix

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 5} = \begin{bmatrix} 1/2 & 0 & 1/4 & 0 & 0 & 1/4 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1/6 & 1/2 & 1/6 & 0 & 0 & 1/6 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) Is the chain $(X_n)_{n\geqslant 0}$ reducible? If yes, find its communicating classes.
b) Determine the transient and recurrent states of $(X_n)_{n\geqslant 0}$.
c) Find the period of each state.

Exercise 6.5 Consider the Markov chain $(X_n)_{n\geqslant 0}$ with transition matrix

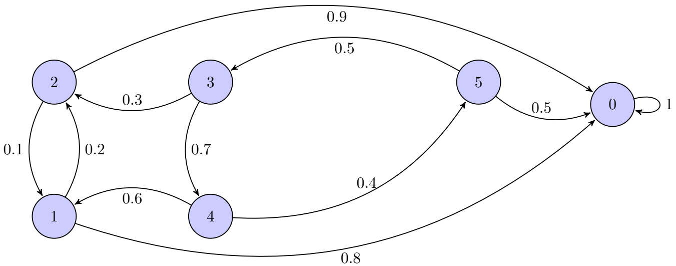
$$\begin{bmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{bmatrix}.$$

a) Is the chain $(X_n)_{n\geqslant 0}$ irreducible? If not, give its communicating classes.
b) Find the period of each state. Which states are absorbing, transient, recurrent, positive recurrent?

Exercise 6.6 In the following chain, find:

a) the communicating class(es),

b) the transient state(s),
c) the recurrent state(s),
d) the positive recurrent state(s),
e) the period of every state.



Exercise 6.7 Consider two boxes containing a total of N balls. At each unit of time one ball is chosen randomly among N and moved to the other box.

a) Write down the transition matrix of the Markov chain $(X_n)_{n\geqslant 0}$ with state space $\mathbb{S} = \{0, 1, 2, \ldots, N\}$, representing the number of balls in the first box.
b) Determine the periodicity, transience and recurrence of the Markov chain.

Exercise 6.8

a) Is the Markov chain of Exercise 4.10-(a) recurrent? positive recurrent?
b) Find the periodicity of every state.
c) Same questions for the success runs Markov chain of Exercise 4.10-(b).

Problem 6.9 Let $\alpha > 0$ and consider the Markov chain with state space $\mathbb{S} = \mathbb{N}$ and transition matrix given by

$$P_{i,i-1} = \frac{1}{\alpha + 1}, \quad P_{i,i+1} = \frac{\alpha}{\alpha + 1}, \qquad i \geqslant 1.$$

and a reflecting barrier at 0, such that $P_{0,1} = 1$. Compute the mean return times $\mathbb{E}[T_k^r \mid X_0 = k]$ for $k \in \mathbb{N}$, and show that the chain is positive recurrent if and only if $\alpha < 1$.

Exercise 6.10 (Antal and Redner (2005), § 5). Consider a cookie-excited random walk $(S_n)_{n\geqslant 0}$ on the half line $\mathbb{Z}_+$, with probabilities $(p, q) = (1/2, 1/2)$ of moving up and down without cookies, and probabilities $(\tilde{p}, \tilde{q})$ of moving up and down on cookie locations, with $\tilde{p} > \tilde{q}$. We assume that $(S_n)_{n\geqslant 0}$ starts at $S_0 = 0$ with no cookie at state $⓪$, and that every cookie location at states $ⓘ$, $i \geqslant 1$, contains initially a same number $k \geqslant 1$ of cookies. Only one single cookie can be eaten at each step.

a) Give the number of cookies initially contained in the region $\{1, 2, \ldots, L\}$, $L \geqslant 1$.
b) Give the minimum number of time steps needed to consume all cookies by traveling within $\{1, 2, \ldots, L\}$.
c) Assuming a positive average drift $\tilde{p} - \tilde{q} > 0$ on cookie locations at every time step, give the average number of time steps needed to travel from state ① to state Ⓛ, assuming that all states contain cookies.
d) Find a condition on $\tilde{p}$ and k ensuring the consumption of all cookies while traveling from from ① to Ⓛ.
e) Find a sufficient condition based on $\tilde{p}$ and k for the transience of this cookie random walk.

Chapter 7
Long-Run Behavior of Markov Chains

This chapter is concerned with the large time behavior of Markov chains, including the computation of their limiting and stationary distributions. Here the notions of recurrence, transience, and classification of states introduced in the previous chapter play a major role.

7.1 Limiting Distributions

Definition 7.1. *A Markov chain* $(X_n)_{n\geqslant 0}$ *is said to admit a* limiting probability distribution *if the following conditions are satisfied:*

i) the limits

$$\lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) \tag{7.1.1}$$

exist for all $i, j \in \mathbb{S}$, *and*

ii) they form a probability distribution *on* $\mathbb{S}$, *i.e.*

$$\sum_{j\in\mathbb{S}} \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = 1, \tag{7.1.2}$$

for all $i \in \mathbb{S}$.

Note that Condition (7.1.2) is always satisfied if the limits (7.1.1) exist and the state space $\mathbb{S}$ is finite.

As remarked in (4.5.6) and (4.5.7) above, the two-state Markov chain has a limiting distribution given by

$$[\pi_0, \pi_1] = \left[\frac{b}{a+b}, \frac{a}{a+b}\right], \tag{7.1.3}$$

provided that $(a,b) \neq (0,0)$ and $(a,b) \neq (1,1)$, while the corresponding mean return times are given from (5.3.3) by

$$(\mu_0(0), \mu_1(1)) = \left(1 + \frac{a}{b}, 1 + \frac{b}{a}\right),$$

i.e. the limiting probabilities are given by the mean return time inverses, as

$$[\pi_0, \pi_1] = \left[\frac{b}{a+b}, \frac{a}{a+b}\right] = \left[\frac{1}{\mu_0(0)}, \frac{1}{\mu_1(1)}\right] = \left[\frac{\mu_1(0)}{\mu_0(1)+\mu_1(0)}, \frac{\mu_0(1)}{\mu_0(1)+\mu_1(0)}\right].$$

This fact is not a simple coincidence, and it is actually a consequence of the following more general result, which shows that the longer it takes on average to return to a state, the smaller the probability is to find the chain in that state. Recall that a chain $(X_n)_{n\geqslant 0}$ is said to be recurrent, resp. aperiodic, if all its states are recurrent, resp. aperiodic.

Theorem 7.2. *(Karlin and Taylor (1998), Theorem IV.4.1). Consider a Markov chain $(X_n)_{n\geqslant 0}$ satisfying the following 3 conditions:*

i) irreducibility,
ii) recurrence, and
iii) aperiodicity.

Then, the chain $(X_n)_{n\geqslant 0}$ admits the limiting distribution

$$\lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = \frac{1}{\mu_j(j)}, \qquad i, j \in \mathbb{S}, \tag{7.1.4}$$

independently of the initial state $i \in \mathbb{S}$, where

$$\mu_j(j) = \mathbb{E}[T_j^r \mid X_0 = j] \in [1, \infty]$$

is the mean return time to state $\textcircled{j} \in \mathbb{S}$.

In Theorem 7.2, Condition (i), resp. Condition (ii), is satisfied from Proposition 6.16, resp. from Proposition 6.7, provided that at least one state is aperiodic, resp. recurrent, since the chain is irreducible.

The conditions stated in Theorem 7.2 are sufficient, but they are not all necessary. For example, a Markov chain may admit a limiting distribution when the recurrence and irreducibility Conditions (i) and (iii) above are not satisfied, cf. for example Exercise 7.16-(b) below.

Note that the limiting probability (7.1.4) is independent of the initial state $\textcircled{i}$, and it vanishes whenever the state $\textcircled{i}$ is transient or null recurrent, cf. Proposition 7.4 below. In the case of the two-state Markov chain this result in consistent with (4.5.6), (4.5.7), and (7.1.3). However it does not apply to *e.g.* the simple random walk of Chapter 3 which is *not* recurrent when $p \neq q$ from (6.2.5), and has period 2.

For an aperiodic chain with finite state space $\mathbb{S}$, we can show that the limit $\lim_{n\to\infty} \mathbb{P}(X_n = i \mid X_0 = j)$ exists for all $i, j \in \mathbb{S}$ by breaking the chain into communicating classes, however it may depend on the initial state $\textcircled{j}$. This however does not apply to the random walk of Chapter 3 which is not aperiodic and has an infinite state space, although it can be turned into an aperiodic chain by allowing a draw as in Exercise 2.1.

The following sufficient condition for existence of a limiting distribution is a consequence of Theorem IV.1.1 in Karlin and Taylor (1998).

Proposition 7.3. *Consider a Markov chain* $(X_n)_{n\geqslant 0}$ *with* finite *state space* $\mathbb{S} = \{0, 1, \ldots, N\}$, *whose transition matrix* P *is* regular, *i.e. there exists* $n \geqslant 1$ *such that all entries of the matrix power* P^n *are non-zero. Then* $(X_n)_{n\geqslant 0}$ *admits a limiting probability distribution* $\pi = (\pi_i)_{i=0,1,\ldots,N}$ *given by*

$$\pi_j = \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i), \qquad 0 \leqslant i, j \leqslant N, \tag{7.1.5}$$

A chain with finite state space is regular if it is aperiodic and irreducible, cf. Proposition 1.7 in Levin et al. (2009).

We close this section with the following proposition, whose proof uses an argument similar to that of Theorem 6.11.

Proposition 7.4. *Let* $(X_n)_{n\geqslant 0}$ *be a Markov chain with a transient state* $\textcircled{j} \in \mathbb{S}$. *Then we have*

$$\lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = 0,$$

for all $\textcircled{i} \in \mathbb{S}$.

Proof. Since $\textcircled{j}$ is a transient state, the probability p_{jj} of return to $\textcircled{j}$ in finite time satisfies $p_{jj} < 1$ by definition, hence by Relation (5.4.6) page 178, the expected number of returns to $\textcircled{j}$ starting from state $\textcircled{i}$ is finite:*

$$\begin{aligned}\mathbb{E}[R_j \mid X_0 = i] &= \mathbb{E}\left[\sum_{n\geqslant 1} \mathbb{1}_{\{X_n = j\}} \;\Big|\; X_0 = i\right] \\ &= \sum_{n\geqslant 1} \mathbb{E}\big[\mathbb{1}_{\{X_n = j\}} \mid X_0 = i\big]\end{aligned}$$

* The exchange of infinite sums and expectation is valid in particular for nonnegative series.

$$= \sum_{n \geqslant 1} \mathbb{P}(X_n = j \mid X_0 = i)$$
$$= \frac{p_{ij}}{1 - p_{jj}} < \infty.$$

The convergence of the above series implies the convergence to 0 of its general term, *i.e.*

$$\lim_{n \to \infty} \mathbb{P}(X_n = j \mid X_0 = i) = 0$$

for all ⓘ $\in \mathbb{S}$, which is the expected conclusion. □

7.2 Stationary Distributions

Definition 7.5. *A* probability distribution *on* $\mathbb{S}$ *is any a family* $\pi = (\pi_i)_{i \in \mathbb{S}}$ *in* $[0,1]$ *such that*

$$\sum_{i \in \mathbb{S}} \pi_i = 1.$$

Next, we state the definition of stationary distribution.

Definition 7.6. *A* probability distribution $\pi = (\pi_i)_{i \in \mathbb{S}}$ *on* $\mathbb{S}$ *is said to be* stationary *if, starting* X_0 *at time* 0 *with the distribution* $(\pi_i)_{i \in \mathbb{S}}$*, it turns out that the distribution of* X_1 *is still* $(\pi_i)_{i \in \mathbb{S}}$ *at time* 1*.*

In other words, $(\pi_i)_{i \in \mathbb{S}}$ is stationary for the Markov chain with transition matrix P if, letting

$$\mathbb{P}(X_0 = i) := \pi_i, \qquad i \in \mathbb{S},$$

at time 0, implies

$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_0 = i) = \pi_i, \quad i \in \mathbb{S},$$

at time 1. This also means that

$$\pi_j = \mathbb{P}(X_1 = j) = \sum_{i \in \mathbb{S}} \mathbb{P}(X_1 = j \mid X_0 = i) \mathbb{P}(X_0 = i) = \sum_{i \in \mathbb{S}} \pi_i P_{i,j}, \quad j \in \mathbb{S},$$

i.e. the distribution π is stationary if and only if the vector π is *invariant* (or stationary) by the matrix P, that means

$$\pi = \pi P. \tag{7.2.1}$$

Note that in contrast with (5.1.4), the multiplication by P in (7.2.1) is on the *right*-hand side and *not* on the left-hand side. The relation (7.2.1) can be rewritten as the *global balance condition*

$$\sum_{i\in\mathbb{S}}\pi_i P_{i,k} = \pi_k = \pi_k \sum_{j\in\mathbb{S}} P_{k,j} = \sum_{j\in\mathbb{S}} \pi_k P_{k,j}, \tag{7.2.2}$$

which is illustrated in Figure 7.1.

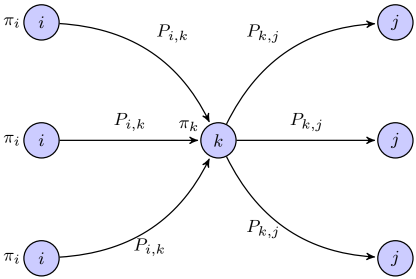


Fig. 7.1: Global balance condition (discrete time).

We also note that the stationarity and limiting properties of distributions are quite different concepts. If the chain is started in the stationary distribution then it will remain in that distribution at *any* subsequent time step (which is stronger than saying that the chain will reach that distribution after an *infinite* number of time steps). On the other hand, in order to reach the limiting distribution the chain can be started from any given initial distribution or even from any fixed given state, and it will converge to the limiting distribution if it exists. Nevertheless, the limiting and stationary distribution may coincide in some situations as in Theorem 7.8 below.

More generally, assuming that X_n has the invariant (or stationary) distribution π at time n, *i.e.* $\mathbb{P}(X_n = i) = \pi_i$, $i \in \mathbb{S}$, we have

$$\begin{aligned}\mathbb{P}(X_{n+1} = j) &= \sum_{i\in\mathbb{S}} \mathbb{P}(X_{n+1} = j \mid X_n = i)\mathbb{P}(X_n = i) \\ &= \sum_{i\in\mathbb{S}} P_{i,j}\mathbb{P}(X_n = i) = \sum_{i\in\mathbb{S}} P_{i,j}\pi_i \\ &= [\pi P]_j = \pi_j, \qquad j \in \mathbb{S},\end{aligned}$$

since the transition matrix of $(X_n)_{n\geqslant 0}$ is *time homogeneous*, hence

$$\mathbb{P}(X_n = j) = \pi_j, \quad j \in \mathbb{S}, \quad \Longrightarrow \quad \mathbb{P}(X_{n+1} = j) = \pi_j, \quad j \in \mathbb{S}.$$

By induction on $n \geqslant 0$, this yields

$$\mathbb{P}(X_n = j) = \pi_j, \qquad j \in \mathbb{S}, \quad n \geqslant 1,$$

i.e. the chain $(X_n)_{n\geqslant 0}$ remains in the same distribution π at all times $n \geqslant 1$, provided that it has been started with the stationary distribution π at time $n = 0$.

Proposition 7.7. *Assume that* $\mathbb{S} = \{0, 1, \ldots, N\}$ *is finite and that the limits*

$$\pi_j^{(i)} := \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = \lim_{n\to\infty} [P^n]_{i,j}$$

exist for all $j \in \mathbb{S}$ *and are independent of the initial state* $i \in \mathbb{S}$*, i.e. we have*

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_0^{(0)} & \pi_1^{(0)} & \cdots & \pi_N^{(0)} \\ \pi_0^{(1)} & \pi_1^{(1)} & \cdots & \pi_N^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_0^{(N)} & \pi_1^{(N)} & \cdots & \pi_N^{(N)} \end{bmatrix}.$$

Then for every $i = 0, 1, \ldots, N$*, the vector* $\pi^{(i)} := (\pi_j^{(i)})_{j\in\{0,1,\ldots,N\}}$ *is a stationary distribution and we have*

$$\pi^{(i)} = \pi^{(i)} P, \tag{7.2.3}$$

i.e. $\pi^{(i)}$ *is invariant (or stationary) by* P*,* $i = 0, 1, \ldots, N$*.*

Proof. We have

$$\begin{aligned} \pi_j^{(i)} : &= \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) \\ &= \lim_{n\to\infty} \mathbb{P}(X_{n+1} = j \mid X_0 = i) \\ &= \lim_{n\to\infty} \sum_{l\in\mathbb{S}} \mathbb{P}(X_{n+1} = j \mid X_n = l)\mathbb{P}(X_n = l \mid X_0 = i) \\ &= \lim_{n\to\infty} \sum_{l\in\mathbb{S}} P_{l,j}\mathbb{P}(X_n = l \mid X_0 = i) \\ &= \sum_{l\in\mathbb{S}} P_{l,j} \lim_{n\to\infty} \mathbb{P}(X_n = l \mid X_0 = i) \\ &= \sum_{l\in\mathbb{S}} \pi_l^{(i)} P_{l,j}, \qquad i, j \in \mathbb{S}, \end{aligned}$$

where we exchanged limit and summation because the state space $\mathbb{S}$ is assumed to be finite, which shows that

$$\pi^{(i)} = \pi^{(i)} P,$$

i.e. (7.2.3) holds and $\pi^{(i)}$ is a stationary distribution, $i = 0, 1, \ldots, N$. □

Proposition 7.7 can be applied in particular when the limiting distribution $\pi_j := \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i)$ does not depend on the initial state $\textcircled{i}$, *i.e.*

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \cdots & \pi_N \\ \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \cdots & \pi_N \end{bmatrix}.$$

For example, the limiting distribution (7.1.3) of the two-state Markov chain is also an invariant distribution, *i.e.* it satisfies (7.2.1). In particular, we have the following result which does not require the finiteness of the state space $\mathbb{S}$.

Theorem 7.8. *(Karlin and Taylor (1998), Theorem IV.4.2). Assume that the Markov chain $(X_n)_{n\geqslant 0}$ satisfies the following 3 conditions:*

i) irreducibility,
ii) positive recurrence, and
iii) aperiodicity.

Then the chain $(X_n)_{n\geqslant 0}$ admits a limiting distribution

$$\pi_j := \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = \lim_{n\to\infty} [P^n]_{i,j} = \frac{1}{\mu_j(j)}, \qquad i, j \in \mathbb{S},$$

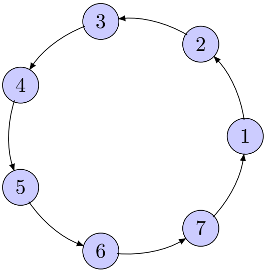
independently of the initial state $i \in \mathbb{S}$, which also forms a stationary distribution $(\pi_j)_{j\in\mathbb{S}} = (1/\mu_j(j))_{j\in\mathbb{S}}$, uniquely determined by the equation

$$\pi = \pi P.$$

In Theorem 7.8 above, Condition (ii), is satisfied from Proposition 6.16, provided that at least one state is aperiodic, since the chain is irreducible. We also note that the limiting distribution π does not depend on the initial distribution of X_0.

See Exercise 7.23 for an application of Theorem 7.8 on an infinite state space.

In the following trivial example of a finite circular chain, Theorems 7.2 and 7.8 cannot be applied since the chain is not aperiodic, and it clearly does not admit a limiting distribution. However, Theorem 7.10 below applies here, and the chain admits a stationary distribution: one can easily check that $\mu_k(k) = n$ and $\pi_k = 1/n = 1/\mu_k(k)$, $k = 1, 2, \ldots, n$, with $n = 7$.



$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{7.2.4}$$

In view of Theorem 6.13, we have the following corollary of Theorem 7.8:

Corollary 7.9. *Consider an irreducible aperiodic Markov chain with finite state space* $\mathbb{S}$*. Then, the limiting probabilities*

$$\pi_i := \lim_{n\to\infty} \mathbb{P}(X_n = i \mid X_0 = j) = \frac{1}{\mu_i(i)}, \qquad i, j \in \mathbb{S},$$

exist and form a stationary distribution which is uniquely determined by the equation

$$\pi = \pi P.$$

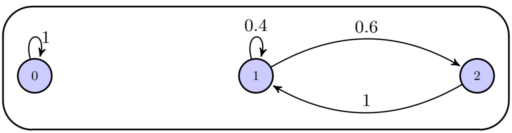
Corollary 7.9 can also be applied separately to derive a stationary distribution on each closed component of a reducible chain.

The convergence of the two-state chain to its stationary distribution has been illustrated in Figure 4.6. Before proceeding further we make some comments on the assumptions of Theorems 7.2 and 7.8.

Remarks

- Irreducibility.

The irreducibility assumption on the chain in Theorems 7.2 and 7.8 is truly required in general, as a reducible chain may have a limiting distribution that depends on the initial state as in the following trivial example on the state space $\{0, 1, 2\}$:



in which the chain is aperiodic and positive recurrent, but not irreducible. Note that the sub-chain $\{1, 2\}$ admits $[\pi_1, \pi_2] = [1/1.6, 0.6/1.6]$ as stationary and limiting distribution, however any vector of the form

$(1-\alpha, \alpha\pi_1, \alpha\pi_2)$ is also a stationary distribution on $\mathbb{S} = \{0, 1, 2\}$ for any $\alpha \in [0, 1]$, showing the non uniqueness of the stationary distribution.

More generally, in case the chain is not irreducible we can split it into subchains and consider the subproblems separately. For example, when the state space $\mathbb{S}$ is a finite set it admits at least one communicating class $A \subset \mathbb{S}$ that leads to no other class, and it admits a stationary distribution π_A by Corollary 7.11 since it is irreducible, hence a chain with finite state space $\mathbb{S}$ admits at least one stationary distribution of the form $(0, 0, \ldots, 0, \pi_A)$.

Similarly, the constant two-state Markov chain with transition matrix $P = I$ is reducible, it admits an infinity of stationary distributions, and a limiting distribution which is dependent on the initial state.

- Aperiodicity.

The conclusions of Theorems 7.2, 7.8 and Corollary 7.9 ensure the existence of the limiting distribution by requiring the aperiodicity of the Markov chain. Indeed, the limiting distribution may not exist when the chain is not aperiodic. For example, the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is not aperiodic (both states have period 2) and it has no limiting distribution because*

$$\mathbb{P}(X_{2n} = 1 \mid X_0) = 1 \quad \text{and} \quad \mathbb{P}(X_{2n+1} = 1 \mid X_0) = 0, \qquad n \in \mathbb{N}. \tag{7.2.5}$$

The chain does have an invariant (or stationary) distribution π solution of $\pi = \pi P$, and given by

$$\pi = [\pi_0, \pi_1] = \left[\frac{1}{2}, \frac{1}{2}\right].$$

- Positive recurrence.

Theorem 7.8, Theorem 7.10 below, and Corollary 7.9 do not apply to the unrestricted random walk $(S_n)_{n \geqslant 0}$ of Chapter 3, because this chain is not positive recurrent, cf. Relations (3.4.17) and (3.4.20), and admits no stationary distribution.

If a stationary distribution $\pi = (\pi_i)_{i \in \mathbb{Z}}$ existed it would satisfy the equation $\pi = \pi P$ which, according to (4.3.1), would read

$$\pi_i = p\pi_{i-1} + q\pi_{i+1}, \qquad i \in \mathbb{Z},$$

i.e.

* This two-state chain is a particular case of the circular chain (7.2.4) for $n = 2$.

$$(p+q)\pi_i = p\pi_{i-1} + q\pi_{i+1}, \qquad i \in \mathbb{Z},$$

or

$$\pi_{i+1} - \pi_i = \frac{p}{q}(\pi_i - \pi_{i-1}), \qquad i \in \mathbb{Z}.$$

As in the direct solution method of page 63, this implies

$$\pi_{i+1} - \pi_i = \left(\frac{p}{q}\right)^i (\pi_1 - \pi_0), \qquad i \in \mathbb{N},$$

so that by a telescoping summation argument we have

$$\begin{aligned}\pi_k &= \pi_0 + \sum_{i=0}^{k-1} (\pi_{i+1} - \pi_i) \\ &= \pi_0 + (\pi_1 - \pi_0) \sum_{i=0}^{k-1} \left(\frac{p}{q}\right)^i \\ &= \pi_0 + (\pi_1 - \pi_0) \frac{1-(p/q)^k}{1-p/q}, \qquad k \in \mathbb{N},\end{aligned}$$

which cannot satisfy the condition $\sum_{k\in\mathbb{Z}} \pi_k = 1$, with $p \neq q$.

When $p = q = 1/2$ we similarly obtain

$$\pi_k = \pi_0 + \sum_{i=0}^{k-1} (\pi_{i+1} - \pi_i) = \pi_0 + k(\pi_1 - \pi_0), \qquad k \in \mathbb{Z},$$

and in this case as well, the sequence $(\pi_k)_{k\geqslant 0}$ cannot satisfy the condition $\sum_{k\in\mathbb{Z}} \pi_k = 1$. Since there is no uniform distribution on an infinite space, we conclude that the chain does not admit a stationary distribution. Hence the stationary distribution of a Markov chain may not exist at all.

In addition, it follows from (3.3.3) and the Stirling approximation formula that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(S_{2n} = 2k \mid S_0 = 0) &= \lim_{n\to\infty} \frac{(2n)!}{(n+k)!(n-k)!} p^{n+k} q^{n-k} \\ &\leqslant \lim_{n\to\infty} \frac{(2n)!}{2^{2n} n!^2} \\ &= \lim_{n\to\infty} \frac{1}{\sqrt{\pi n}} = 0, \qquad k \in \mathbb{N},\end{aligned}$$

as in Problem 3.18-(f), so that the limiting distribution does not exist as well. Here, Theorem 7.2 cannot be applied because the chain is not aperiodic (it has period 2), however aperiodicity and irreducibility are not sufficient in general when the state space is infinite, cf. *e.g.* the model of Exercise 2.1.

The following theorem gives sufficient conditions for the existence of a stationary distribution, without requiring aperiodicity or finiteness of the state space. As noted in (7.2.5) above, the limiting distribution may not exist in this case. See also Problem 6.9 and Exercise 7.23 for an example of a null recurrent chain which does not admit a stationary distribution.

Theorem 7.10. *(Bosq and Nguyen (1996), Theorem 4.1). Consider a Markov chain $(X_n)_{n\geqslant 0}$ satisfying the following two conditions:*

i) irreducibility, and
ii) positive recurrence.

Then, the probabilities

$$\pi_i = \frac{1}{\mu_i(i)}, \qquad i \in \mathbb{S},$$

form a stationary distribution which is uniquely determined by the equation $\pi = \pi P$.

Note that the conditions stated in Theorem 7.10 are sufficient, but they are not all necessary. For example, Condition (ii) is not necessary as the trivial constant chain, whose transition matrix $P = \mathrm{I}$ is reducible, does admit a stationary distribution.

Note that the positive recurrence assumption in Theorem 7.2 is required in general on infinite state spaces. For example, the process in Exercise 7.23 is positive recurrent for $\alpha < 1$ only, whereas no stationary distribution exists when $\alpha \geqslant 1$.

As a consequence of Corollary 6.14 we have the following corollary of Theorem 7.10, which does not require aperiodicity for the stationary distribution to exist.

Corollary 7.11. *Let $(X_n)_{n\geqslant 0}$ be an irreducible Markov chain with* finite *state space $\mathbb{S}$. Then, the probabilities*

$$\pi_k = \frac{1}{\mu_k(k)}, \qquad k \in \mathbb{S},$$

form a stationary distribution which is uniquely determined by the equation

$$\pi = \pi P.$$

Similarly to Proposition 7.3, any Markov chain $(X_n)_{n\geqslant 0}$ with *finite* state space and whose transition matrix P is *regular* admits a stationary distribution, as it is irreducible.

According to Corollary 7.11, the limiting distribution and stationary distribution both exist (and coincide) when the chain is irreducible aperiodic with finite state space $\mathbb{S}$, and in this case we have $\pi_k > 0$ for all $k \in \mathbb{S}$ by Corollaries 6.14 and 7.11. When the chain is irreducible it is usually easier to compute the stationary distribution, which will give us the limiting distribution.

Under the assumptions of Theorem 7.8, if the stationary and limiting distributions both exist then they are equal and in this case we only need to compute one of them. However, in some situations only the stationary distribution might exist. According to Corollary 7.11 above the stationary distribution always exists when the chain is irreducible with finite state space, nevertheless the limiting distribution may not exist if the chain is not aperiodic, consider for example the two-state switching chain with $a = b = 1$.

Finding a limiting distribution

In summary:

- We usually attempt first to compute the stationary distribution whenever possible, and this also gives the limiting distribution when it exists. For this, we first check whether the chain is positive recurrent, aperiodic and irreducible, in which case the limiting distribution can be found by solving $\pi = \pi P$ according to Theorem 7.8.

- In case the above properties are not satisfied we need to compute the limiting distribution by taking the limit $\lim_{n\to\infty} P^n$ of the powers P^n of the transition matrix, if possible by decomposing the state space in communicating classes as in *e.g.* Exercise 7.12. This can turn out to be much more complicated and done only in special cases. If the chain has period $d \geqslant 2$ we may need to investigate the limits $\lim_{n\to\infty} P^{nd}$ instead, see *e.g.* Exercise 7.4 and (7.2.6)-(7.2.7) below.

To further summarize, we note that by Theorem 7.2 we have

a) irreducible + recurrent + aperiodic $\Longrightarrow$ existence of a limiting distribution,

   by Theorem 7.8 we get

b) irreducible + positive recurrent + aperiodic $\Longrightarrow$ existence of a limiting distribution which is also stationary,

   and by Theorem 7.10 we get

c) irreducible + positive recurrent $\Longrightarrow$ existence of a stationary distribution.

In addition, the limiting or stationary distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ satisfies

$$\pi_i = \frac{1}{\mu_i(i)}, \qquad i \in \mathbb{S},$$

in all above cases (a), (b) and (c).

The ergodic theorem

The *Ergodic Theorem*, cf. *e.g.* Theorem 1.10.2 of Norris (1998) states the following.

Theorem 7.12. *Assume that the chain* $(X_n)_{n\geqslant 0}$ *is* irreducible. *Then the sample average time spent at state* ⓘ *converges almost surely to* $1/\mu_i(i)$, *i.e.*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=i\}} = \frac{1}{\mu_i(i)}, \qquad i \in \mathbb{S}.$$

In case $(X_n)_{n\geqslant 0}$ is also positive recurrent, Theorem 7.10 shows that we also have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=i\}} = \pi_i, \qquad i \in \mathbb{S},$$

where $(\pi_i)_{i\in\mathbb{S}}$ is the stationary distribution of $(X_n)_{n\geqslant 0}$. We refer to Figure 4.6 for an illustration of convergence in the setting of the Ergodic Theorem 7.12.

Example. Consider the maze random walk (5.3.7) with transition matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The equation $\pi = \pi P$ yields

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_2 \\ \pi_2 = \pi_1 + \frac{1}{2}\pi_3 \\ \pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_6 \\ \pi_4 = \frac{1}{2}\pi_7 \\ \pi_5 = \frac{1}{2}\pi_6 + \frac{1}{3}\pi_8 \\ \pi_6 = \frac{1}{2}\pi_3 + \frac{1}{2}\pi_5 \\ \pi_7 = \pi_4 + \frac{1}{3}\pi_8 \\ \pi_8 = \frac{1}{2}\pi_5 + \frac{1}{2}\pi_7 + \pi_9 \\ \pi_9 = \frac{1}{3}\pi_8, \end{cases} \qquad \text{hence} \qquad \begin{cases} \pi_1 = \frac{1}{2}\pi_2 \\ \pi_2 = \pi_3 \\ \pi_3 = \pi_6 \\ \pi_4 = \frac{1}{2}\pi_3 \\ \frac{1}{2}\pi_7 = \frac{1}{3}\pi_8 \\ \pi_6 = \pi_5 \\ \pi_6 = \pi_7 \\ \pi_9 = \frac{1}{3}\pi_8, \end{cases}$$

and

$$\begin{aligned} 1 &= \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6 + \pi_7 + \pi_8 + \pi_9 \\ &= \pi_1 + 2\pi_1 + 2\pi_1 + \pi_1 + 2\pi_1 + 2\pi_1 + 2\pi_1 + 3\pi_1 + \pi_1 \\ &= 16\pi_1, \end{aligned}$$

hence

$$\pi_1 = \frac{1}{16}, \ \pi_2 = \frac{2}{16}, \ \pi_3 = \frac{2}{16}, \ \pi_4 = \frac{1}{16},$$
$$\pi_5 = \frac{2}{16}, \ \pi_6 = \frac{2}{16}, \ \pi_7 = \frac{2}{16}, \ \pi_8 = \frac{3}{16}, \ \pi_9 = \frac{1}{16},$$

cf. Figures 7.3 and 7.2 below, and we check that since $\mu_1(1) = 16$ by (5.3.8), we indeed have

$$\pi_1 = \frac{1}{\mu_1(1)} = \frac{1}{16},$$

according to Corollary 7.11.

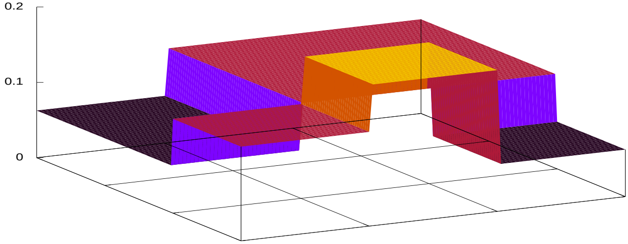


Fig. 7.2: Stationary distribution on the maze (5.3.7).

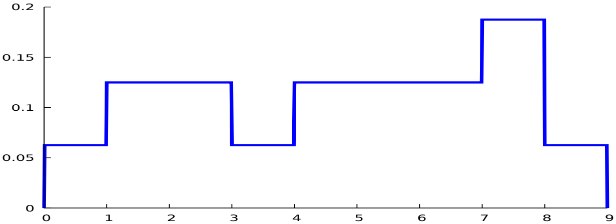
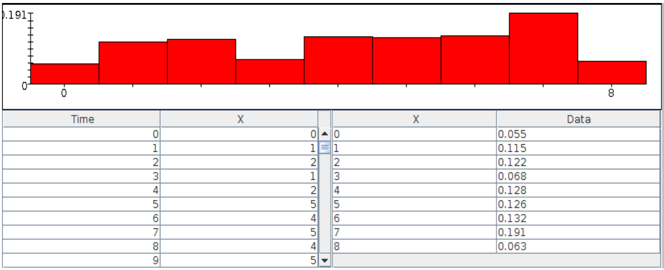


Fig. 7.3: Stationary distribution by state numbering.

The stationary probability distribution of Figures 7.3 and (7.2) can be compared to the proportions of time spent at each state simulated in Figure 7.4 using this Markov chain experiment.



Time	X
0	0
1	1
2	2
3	1
4	2
5	5
6	4
7	5
8	4
9	5

X	Data
0	0.055
1	0.115
2	0.122
3	0.068
4	0.128
5	0.126
6	0.132
7	0.191
8	0.063

Fig. 7.4: Simulated stationary distribution.

Note that this chain has period 2 and the matrix powers $(P^n)_{n\geqslant 0}$ do not converge as n tends to infinity, *i.e.* it does not admit a limiting distribution. In fact, using the following Matlab/Octave commands:

```
P = [0,1,0,0,0,0,0,0,0;
1/2,0,1/2,0,0,0,0,0,0;
0,1/2,0,0,0,1/2,0,0,0;
0,0,0,0,0,0,1,0,0;
0,0,0,0,0,1/2,0,1/2,0;
0,0,1/2,0,1/2,0,0,0,0;
0,0,0,1/2,0,0,0,1/2,0;
0,0,0,0,1/3,0,1/3,0,1/3;
0,0,0,0,0,0,0,1,0]
mpower(P,1000)
mpower(P,1001)
```

we see that

$$\lim_{n\to\infty} P^{2n} = \begin{bmatrix} 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \end{bmatrix}, \qquad (7.2.6)$$

and

$$\lim_{n\to\infty} P^{2n+1} = \begin{bmatrix} 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \\ 1/8 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/8 \\ 0 & 1/4 & 0 & 1/8 & 0 & 1/4 & 0 & 3/8 & 0 \end{bmatrix}, \qquad (7.2.7)$$

which shows that, although $(P^n)_{n\geqslant 1}$ admits two converging subsequences, $\lim_{n\to\infty} P^n$ does not exist, therefore this Markov chain does not admit a limiting distribution.

7.3 Markov Chain Monte Carlo (MCMC)

The goal of the Markov Chain Monte Carlo (MCMC) method, or Metropolis algorithm, is to generate random samples according to a target distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ via a Markov chain that admits π as limiting and stationary distribution. It can be applied in particular in the setting of large state spaces $\mathbb{S}$, cf. *e.g.* Problem 7.36-(a).

A Markov chain $(X_k)_{k\geqslant 0}$ with transition matrix P on a state space $\mathbb{S}$ is said to satisfy the *detailed balance* (or *reversibility*) condition with respect to the probability distribution $\pi = (\pi_i)_{i\in S}$ if

$$\pi_i P_{i,j} = \pi_j P_{j,i}, \qquad i, j \in \mathbb{S}, \tag{7.3.1}$$

see Figure 7.5.

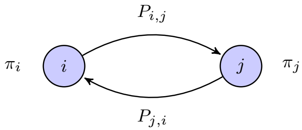


Fig. 7.5: Detailed balance condition (discrete time).

Lemma 7.13. *The detailed balance condition* (7.3.1) *implies the global balance condition* (7.2.2).

Proof. By summation over $i \in \mathbb{S}$ in (7.3.1), we have

$$\sum_{i\in\mathbb{S}} \pi_i P_{i,j} = \sum_{i\in\mathbb{S}} \pi_j P_{j,i} = \pi_j \sum_{i\in\mathbb{S}} P_{j,i} = \pi_j, \quad j \in \mathbb{S},$$

which shows that $\pi P = \pi$, *i.e.* π is a stationary distribution for P, cf. *e.g.* Problem 7.32-(c). □

If the transition matrix P satisfies the detailed balance condition with respect to π then the probability distribution of X_n will naturally converge to the stationary distribution π in the long run, *e.g.* under the hypotheses of Theorem 7.8, *i.e.* when the chain $(X_k)_{k\geqslant 0}$ is positive recurrent, aperiodic, and irreducible.

Metropolis-Hastings algorithm

In general, however, the detailed balance (or reversibility) condition (7.3.1) may not be satisfied by π and P. In this case, starting from a proposal matrix P, one can construct a modified transition matrix $\widetilde{P}$ that will satisfy the detailed balance condition with respect to π. This modified transition matrix $\widetilde{P}$ is defined by

$$\begin{aligned}\widetilde{P}_{i,j} &:= P_{i,j} \times \min\left(1, \frac{\pi_j P_{j,i}}{\pi_i P_{i,j}}\right) \\ &= \min\left(P_{i,j}, \frac{\pi_j P_{j,i}}{\pi_i}\right) \\ &= \begin{cases} P_{j,i}\dfrac{\pi_j}{\pi_i} & \text{if } \ \pi_j P_{j,i} \leqslant \pi_i P_{i,j}, \\ \\ P_{i,j} & \text{if } \ \pi_j P_{j,i} \geqslant \pi_i P_{i,j}, \end{cases} \end{aligned} \tag{7.3.2}$$

for $i \neq j$. We note that

$$\sum_{\substack{j \in \mathbb{S} \\ j \neq i}} \widetilde{P}_{i,j} \leqslant \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} P_{i,j} \leqslant 1,$$

and for $i \in \mathbb{S}$ we let

$$\begin{aligned}\widetilde{P}_{i,i} &:= 1 - \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} \widetilde{P}_{i,j} \\ &= P_{i,i} + \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} P_{i,j}\left(1 - \min\left(1, \frac{\pi_j P_{j,i}}{\pi_i P_{i,j}}\right)\right) \\ &= P_{i,i} + \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} P_{i,j}\left(1 - \frac{\pi_j P_{j,i}}{\pi_i P_{i,j}}\right)^+ \\ &= P_{i,i} + \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} \left(P_{i,j} - \frac{\pi_j P_{j,i}}{\pi_i}\right)^+ .\end{aligned}$$

Clearly, we have $\widetilde{P} = P$ when the detailed balance (or reversibility) condition (7.3.1) is satisfied by P. In the general case, we can check that for $i \neq j$, we have

$$\pi_i \widetilde{P}_{i,j} = \left\{\begin{array}{ll} P_{j,i}\pi_j = \pi_j \widetilde{P}_{j,i} & \text{if } \ \pi_j P_{j,i} \leqslant \pi_i P_{i,j}, \\ \\ \pi_i P_{i,j} = \pi_j \widetilde{P}_{j,i} & \text{if } \ \pi_j P_{j,i} \geqslant \pi_i P_{i,j}, \end{array}\right\} = \pi_j \widetilde{P}_{j,i},$$

hence $\widetilde{P}$ satisfies the detailed balance condition with respect to π (the condition is obviously satisfied when $i = j$). Therefore, the random simulation of $(\widetilde{X}_n)_{n \geqslant 0}$ according to the transition matrix $\widetilde{P}$ will provide samples of the distribution π in the long run as n tends to infinity, provided that the chain $(\widetilde{X}_n)_{n \geqslant 0}$ is positive recurrent, aperiodic, and irreducible.

In standard MCMC sampling, the transition matrix is usually symmetric with $P_{i,j} = P_{j,i} > 0$, $i, j \in \mathbb{S}$, typically $P_{i,j} = \varphi(i-j)$ with φ a Gaussian type density kernel, and the modified transition matrix $\widetilde{P}$ simplifies to

$$\widetilde{P}_{i,j} := P_{i,j} \times \min\left(1, \frac{\pi_j}{\pi_i}\right) = \min\left(P_{i,j}, \frac{\pi_j}{\pi_i} P_{i,j}\right) = \begin{cases} P_{i,j}\dfrac{\pi_j}{\pi_i} & \text{if } \ \pi_j < \pi_i, \\ \\ P_{i,j} & \text{if } \ \pi_j \geqslant \pi_i, \end{cases}$$

for $i \neq j$, with

$$\widetilde{P}_{i,i} := 1 - \sum_{\substack{j \in \mathbb{S} \\ j \neq i}} \widetilde{P}_{i,j} = P_{i,i} + \sum_{\substack{j \in \mathbb{S}, j \neq i \\ \pi_j < \pi_i}} P_{i,j}\left(1 - \frac{\pi_j}{\pi_i}\right), \qquad i \in \mathbb{S}.$$

In other words, starting from state $\textcircled{i}$, a proposal $\textcircled{j}$ is generated with probability $P_{i,j}$. This proposal is then accepted if $\pi_j \geqslant \pi_i$, otherwise if $\pi_j < \pi_i$, the proposal is accepted with probability π_j/π_i, and one remains at state $\textcircled{i}$ with probability $1 - \pi_j/\pi_i$.

Generating posterior samples using MCMC

In this case, $\pi = (\pi_i)_{i \in \mathbb{S}}$ represents the prior distribution of a model parameter. Given $\mathcal{O}$ a set of observations sampled according to a distribution $(p_k)_{k \in \mathcal{O}}$, we are given a likelihood function $l(k|i)$ which represents the probability of observing $k \in \mathcal{O}$ when the system parameter is $\textcircled{i}$.

The posterior probability distribution $p(i|k)$ of being in the state $\textcircled{i}$ given that we observed $k \in \mathcal{O}$ is obtained by the Bayes formula as

$$p(i|k) = l(k|i)\frac{\pi_i}{p_k}, \qquad i \in \mathbb{S}, \ k \in \mathcal{O}.$$

Computing the posterior distribution $p(i|k)$ and generating the corresponding random samples may require to estimate the distribution p_k, $k \in \mathcal{O}$.

The Markov Chain Monte Carlo method provides an efficient way to generate random samples according to the posterior distribution $p(i|k)$. For this, the ratio π_j/π_i in (7.3.2) is replaced with

$$\frac{l(k|j)\pi_j}{l(k|i)\pi_i} = \frac{p(j|k)p_k}{p(i|k)p_k} = \frac{p(j|k)}{p(i|k)}, \qquad i, j \in \mathbb{S}, \ k \in \mathcal{O}, \tag{7.3.3}$$

which uses the information given by the observation $\textcircled{k}$.

Relation (7.3.3) shows that the proposal $\textcircled{j}$ generated with probability $P_{i,j}$ is accepted if $p(j|k) \geqslant p(i|k)$, *i.e.* if its posterior probability $p(j|k)$ given the observation k is higher than the posterior probability $p(i|k)$ of the initial state $\textcircled{i}$. Otherwise, if $p(j|k) < p(i|k)$ the proposal $\textcircled{j}$ is accepted only with the probability given by (7.3.3).

Improved versions of the MCMC algorithms include the Hamiltonian Monte Carlo method and the No U-Turn Sampler (NUTS).

In Table 7.1 we summarize the definitions introduced in this chapter and in Chapter 6.

Table 7.1: Summary of Markov chain properties.

Property	Definition
Absorbing (state)	$P_{i,i} = 1$
Recurrent (state)	$\mathbb{P}(T_i^r < \infty \mid X_0 = i) = 1$
Transient (state)	$\mathbb{P}(T_i^r < \infty \mid X_0 = i) < 1$
Positive recurrent (state)	Recurrent *and* $\mathbb{E}[T_i^r \mid X_0 = i] < \infty$
Null recurrent (state)	Recurrent *and* $\mathbb{E}[T_i^r \mid X_0 = i] = \infty$
Aperiodic (state or chain)	Period(s) $= 1$
Ergodic (state or chain)	Positive recurrent *and* aperiodic
Irreducible (chain)	*All* states communicate
Regular (chain)	All coefficients of P^n are > 0 for some $n \geqslant 1$
Stationary distribution π	Obtained from solving $\pi = \pi P$

Exercises

Exercise 7.1 Find the stationary distribution $[\pi_0, \pi_1]$ of the two-state Markov chain on $\mathbb{S} = \{0, 1\}$ with transition probability matrix

$$P = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix} \end{matrix}.$$

Exercise 7.2 We consider the Markov chain of Exercise 4.10-(a).

a) Is this chain irreducible, aperiodic, recurrent, positive recurrent?
b) Does this Markov chain admit a stationary distribution?
c) Does this Markov chain admit a limiting distribution?

Exercise 7.3 We consider the success runs Markov chain of Exercise 4.10-(b).

a) Is the success runs chain irreducible, aperiodic, recurrent, positive recurrent?
b) Does this Markov chain admit a stationary distribution?
c) Does this Markov chain admit a limiting distribution?

Exercise 7.4 We consider the Ehrenfest chain (4.3.2)-(4.3.3).

a) Is the Ehrenfest chain irreducible, aperiodic, recurrent, positive recurrent?
b) Does this Markov chain admit a stationary distribution?
   Hint: Try a binomial distribution.
c) Does this Markov chain admit a limiting distribution?

Exercise 7.5 Consider the Bernoulli-Laplace chain $(X_n)_{n\geqslant 0}$ of Exercise 4.9 with state space $\mathbb{S} = \{0, 1, 2, \ldots, N\}$ and transition matrix

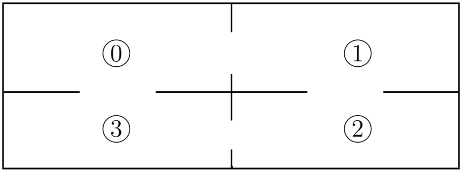
$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 1/N^2 & 2(N-1)/N^2 & (N-1)^2/N^2 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 2^2/N^2 & 4(N-2)/N^2 & (N-2)^2/N^2 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 3^2/N^2 & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \vdots & 0 & 3^2/N^2 & 0 & 0 \\ 0 & 0 & \cdots & 0 & (N-2)^2/N^2 & 4(N-2)/N^2 & 2^2/N^2 & 0 \\ 0 & 0 & \cdots & 0 & 0 & (N-1)^2/N^2 & 2(N-1)/N^2 & 1/N^2 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

i.e.

$$P_{k,k-1} = \frac{k^2}{N^2}, \quad P_{k,k} = \frac{2k(N-k)}{N^2}, \quad P_{k,k+1} = \frac{(N-k)^2}{N^2}, \quad k = 1, 2, \ldots, N-1.$$

a) Is the Bernoulli-Laplace chain irreducible, aperiodic, recurrent, positive recurrent?
b) Does this Markov chain admit a stationary distribution?
c) Does this Markov chain admit a limiting distribution?

Exercise 7.6 Consider a robot evolving in the following circular maze, moving from one room to a different room at each time step, according to a Markov chain with equal transition probabilities.



Let $X_n \in \{0,1,2,3\}$ denote the state of the robot at time $n \in \mathbb{N}$.

a) Write down the transition matrix P of the chain.
b) By first step analysis, compute the mean **return** times $\mu_0(k)$ from state $k = 0,1,2,3$ to state ⓪.

   Hint: One may use the symmetry of the problem to simplify the calculations.
c) Guess an invariant (or stationary) probability distribution $[\pi_0, \pi_1, \pi_2, \pi_3]$ for the chain, and show that it does satisfy the condition $\pi = \pi P$.

Exercise 7.7 A signal processor is analysing a sequence of signals that can be either distorted or non-distorted. It turns out that on average, 1 out of 4 signals following a distorted signal are distorted, while 3 out of 4 signals are non-distorted following a non-distorted signal.

a) Let $X_n \in \{D, N\}$ denote the state of the *n-th* signal being analysed by the processor. Show that the process $(X_n)_{n\geqslant 1}$ can be modeled as a Markov chain and determine its transition matrix.
b) Compute the stationary distribution of $(X_n)_{n\geqslant 1}$.
c) In the long run, what fraction of analysed signals are distorted?
d) Given that the last observed signal was distorted, how long does it take on average until the next non-distorted signal?
e) Given that the last observed signal was non-distorted, how long does it take on average until the next distorted signal?

Exercise 7.8 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\{0,1,2,3\}$ with transition probability matrix P given by

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.2 & 0 & 0.8 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0.4 & 0.6 & 0 & 0 \end{bmatrix}.$$

a) Draw the graph of this chain. Is the chain reducible?
b) Find the recurrent, transient, and absorbing state(s) of this chain.
c) Compute the fraction of time spent at state ⓪ in the long run.

d) On the average, how long does it take to reach state ⓪ after starting from state ②?

Exercise 7.9 Consider the transition probability matrix

$$P = \left[P_{i,j} \right]_{0 \leqslant i,j \leqslant 3} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.1 & 0.4 & 0.2 & 0.3 \\ 0.2 & 0.2 & 0.5 & 0.1 \\ 0.3 & 0.3 & 0.4 & 0 \end{bmatrix}.$$

a) Compute the limiting distribution $[\pi_0, \pi_1, \pi_2, \pi_3]$ of this Markov chain.
b) Compute the average time $\mu_0(1)$ it takes to the chain to travel from state ① to state ⓪.

   Hint: The data of the first row of the matrix P should play no role in the computation of $\mu_0(k)$, $k = 0, 1, 2, 3$.
c) Prove by direct computation that the relation $\pi_0 = 1/\mu_0(0)$ holds, where $\mu_0(0)$ represents the mean return time to state ⓪ for this chain.

Exercise 7.10 Consider the Markov chain with transition probability matrix

$$\left[P_{i,j} \right]_{0 \leqslant i,j \leqslant 3} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}.$$

a) Show that the chain is periodic* and compute its period.
b) Determine the stationary distribution of this chain.

Exercise 7.11 The lifetime of a given component of a machine is a discrete random variable T with distribution

$$\mathbb{P}(T = 1) = 0.1, \quad \mathbb{P}(T = 2) = 0.2, \quad \mathbb{P}(T = 3) = 0.3, \quad \mathbb{P}(T = 4) = 0.4.$$

The component is immediately replaced with a new component upon failure, and the machine starts functioning with a new component. Compute the long run probability of finding the machine about to fail at the next time step.

Exercise 7.12 Suppose that a Markov chain has the one-step transition probability matrix P on the state space $\{A, B, C, D, E\}$ given by

* A chain is periodic when all states have the same period

$$P = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0.2 & 0 & 0.4 & 0 & 0.4 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Find $\lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 = C)$.

Exercise 7.13 Consider the Markov chain $(X_n)_{n\geqslant 0}$ on the state space $\{0, 1, 2, 3, 4\}$ with transition probability matrix P given by

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/7 & 6/7 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

a) Draw the graph of this chain.
b) Identify the communicating class(es).
c) Find the recurrent, transient, and absorbing state(s) of this chain.
d) Find $\lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 4)$.

Exercise 7.14 Three out of 4 trucks passing under a bridge are followed by a car, while only 1 out of every 5 cars passing under that same bridge is followed by a truck. Let $X_n \in \{C, T\}$ denote the nature of the *n-th* vehicle passing under the bridge, $n \geqslant 1$.

a) Show that the process $(X_n)_{n\geqslant 1}$ can be modeled as a Markov chain, and write down its transition matrix.
b) Compute the stationary distribution of $(X_n)_{n\geqslant 1}$.
c) In the long run, what fraction of vehicles passing under the bridge are trucks?
d) Given that the last vehicle seen was a truck, how long does it take on average until the next truck is seen under that same bridge?

Exercise 7.15 Consider a discrete-time Markov chain $(X_n)_{n\geqslant 0}$ on $\mathbb{S} = \{1, 2, \ldots, N\}$, whose transition matrix $P = (P_{i,j})_{1\leqslant i,j\leqslant N}$ is assumed to be *symmetric*, *i.e.* $P_{i,j} = P_{j,i}$, $1 \leqslant i, j \leqslant N$,

a) Find an invariant (or stationary) distribution of the chain.

   Hint: The equation $\pi P = \pi$ admits an easy solution.
b) Assume further that $P_{i,i} = 0$, $1 \leqslant i \leqslant N$, and that $P_{i,j} > 0$ for all $1 \leqslant i < j \leqslant N$. Find the period of every state.

Exercise 7.16 (Gusev (2014), Problem 5.36-(d) continued).

a) Is the chain $(Y_k)_{k\geqslant 0}$ reducible? Find its communicating classes.
b) Find the limiting distribution, and the possible stationary distributions of the chain $(Y_k)_{k\geqslant 0}$.

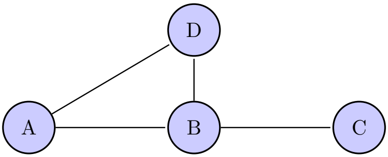
Exercise 7.17 Consider the Markov chain with transition matrix

$$\begin{bmatrix} q & p & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ q & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & p \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where $p, q \in (0,1)$ satisfy $p+q=1$.

a) Compute the stationary distribution $[\pi_0, \pi_1, \pi_2, \pi_3, \pi_4]$ of this chain.
b) Compute the limiting distribution of the chain.

Exercise 7.18 Four players A, B, C, D are connected by the following network, and play by exchanging a token.



At each step of the game, the player who holds the token chooses another player he is connected to, and sends the token to that player.

a) Assuming that the player choices are made at random and are equally distributed, model the states of the token as a Markov chain $(X_n)_{n\geqslant 1}$ on $\{A, B, C, D\}$ and give its transition matrix.
b) Compute the stationary distribution $[\pi_A, \pi_B, \pi_C, \pi_D]$ of $(X_n)_{n\geqslant 1}$.

   Hint: To simplify the resolution, start by arguing that we have $\pi_A = \pi_D$.
c) Compute the mean return times $\mu_D(i)$, $i \in \{A, B, C, D\}$. On average, how long does player D have to wait to recover the token ?
d) In the long run, what is the probability that player D holds the token?

Exercise 7.19 Consider a Markov chain with transition probability matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix},$$

with $a+b+c=1$.

a) Compute the matrix power P^n for all $n \geqslant 2$.
b) Does this Markov chain admit a limiting distribution? If yes, compute that distribution.
c) Does this Markov chain admit a stationary distribution? Compute this distribution if it exists.

Exercise 7.20 Consider a game server that can become offline with probability p and can remain online with probability $q = 1 - p$ on any given day. Assume that the random time N it takes to fix the server has the geometric distribution

$$\mathbb{P}(N = k) = \beta(1-\beta)^{k-1}, \qquad k \geqslant 1,$$

with parameter $\beta \in (0,1)$. We let $X_n = 1$ when the server is online on day n, and $X_n = 0$ when it is offline.

a) Show that the process $(X_n)_{n\geqslant 0}$ can be modeled as a discrete-time Markov chain, and write down its transition matrix.
b) Compute the probability that the server is online in the long run, in terms of the parameters β and p.

Exercise 7.21 Let $(X_n)_{n\geqslant 0}$ be an irreducible aperiodic Markov chain on the finite state space $\mathbb{S} = \{1, 2, \ldots, N\}$.

a) Show that there *exists* a state $i \in \{1, 2, \ldots, N\}$ such that the mean return time $\mu_i(i)$ from state i to itself is lower or equal to N, *i.e.* $\mu_i(i) \leqslant N$.
b) Show that there *exists* a state $i \in \{1, 2, \ldots, N\}$ such that the mean return time $\mu_i(i)$ from state i to itself is higher or equal to N, *i.e.* $\mu_i(i) \geqslant N$.

Exercise 7.22 Consider a Markov chain on the state space $\{1, 2, \ldots, N\}$. For any $i \in \{2, \ldots, N-1\}$, the chain has probability $p \in (0,1)$ to switch from state (i) to state (i+1), and probability $q = 1 - p$ to switch from (i) to (i−1). When the chain reaches state (i) = (1) it rebounds to state (2) with probability p or stays at state (1) with probability q. Similarly, after reaching state (N) it rebounds to state (N−1) with probability q, or remains at N with probability p.

a) Write down the transition probability matrix of this chain.
b) Is the chain reducible?
c) Determine the absorbing, transient, recurrent, and positive recurrent states of this chain.
d) Compute the stationary distribution of this chain.
e) Compute the limiting distribution of this chain.

Exercise 7.23 (Problem 6.9 continued). Let $\alpha > 0$ and consider the Markov chain with state space $\mathbb{S} = \mathbb{N}$ and transition matrix given by

$$P_{i,i-1} = \frac{1}{\alpha+1}, \quad P_{i,i+1} = \frac{\alpha}{\alpha+1}, \qquad i \geqslant 1.$$

and a reflecting barrier at 0, such that $P_{0,1} = 1$.

a) Show that when $\alpha \in (0,1)$ this chain admits a stationary distribution of the form

$$\pi_k = \alpha^{k-1}(1-\alpha^2)/2, \qquad k \geqslant 1,$$

where the value of π_0 has to be determined.

b) Does this Markov chain admit a stationary distribution when $\alpha \geqslant 1$?

c) Show that this Markov chain is positive recurrent when $\alpha < 1$.

Exercise 7.24 Consider two discrete-time stochastic processes $(X_n)_{n\geqslant 0}$ and $(Y_n)_{n\geqslant 0}$ on a state space $\mathbb{S}$, such that

$$X_n = Y_n, \qquad n \geqslant \tau,$$

where τ is a random time called the *coupling time* of $(X_n)_{n\geqslant 0}$ and $(Y_n)_{n\geqslant 0}$.

a) Show that for all $x \in \mathbb{S}$ and $n \in \mathbb{N}$ we have

$$\mathbb{P}(X_n = x) \leqslant \mathbb{P}(Y_n = x) + \mathbb{P}(\tau > n) \qquad x \in \mathbb{S}, \quad n \geqslant 0.$$

Hint: Use the law of total probability as

$$\mathbb{P}(A) = \mathbb{P}(A \cap \{\tau \leqslant n\}) + \mathbb{P}(A \cap \{\tau > n\}).$$

b) Show that for all $n \in \mathbb{N}$ we have

$$\operatorname{Sup}_{x\in\mathbb{S}} |\mathbb{P}(X_n = x) - \mathbb{P}(Y_n = x)| \leqslant \mathbb{P}(\tau > n), \qquad n \geqslant 0.$$

Exercise 7.25 (Aldous and Diaconis (1986), Jonasson (2009)). Let $(X_n)_{n\geqslant 1}$ denote a Markov chain on a finite state space $\mathbb{S}$, and let $\tau \geqslant 0$ denote a random time such that the distribution π of X_n given $\{\tau \leqslant n\}$ does not depend on $n \geqslant 0$, *i.e.*

$$\mathbb{P}(X_n \in A \mid \tau \leqslant n) = \pi(A), \qquad A \subset \mathbb{S}, \quad n \geqslant 0.$$

a) Show that

$$\mathbb{P}(X_n \in A) = \pi(A) + (\mathbb{P}(X_n \in A \mid \tau > n) - \pi(A))\mathbb{P}(\tau > n),$$

$\subset \mathbb{S}$, $n \geqslant 0$.

Hint. Split $\mathbb{P}(X_n \in A)$ as

$$\mathbb{P}(X_n \in A) = \mathbb{P}(X_n \in A \text{ and } \tau \leqslant n) + \mathbb{P}(X_n \in A \text{ and } \tau > n).$$

b) Show the total variation distance bound

$$\|\mathbb{P}(X_n \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} := \operatorname*{Sup}_{A \subset \mathbb{S}} |\mathbb{P}(X_n \in A) - \pi(A)| \leqslant \mathbb{P}(\tau > n),$$

between π and the distribution of X_n, $n \geqslant 0$.

Hint. Use the inequalities

$$-1 \leqslant a - 1 \leqslant a - b \leqslant 1 - b \leqslant 1, \qquad a, b \in [0, 1].$$

c) Give an example of a random time such that the distribution π of X_n given $\{\tau \leqslant n\}$ does not depend on $n \geqslant 0$.

Exercise 7.26 Let $M = (M_{i,j})_{1 \leqslant i,j \leqslant n}$ denote a *column-stochastic* matrix, *i.e.* M is such that

$$\sum_{i=1}^{n} M_{i,j} = 1, \qquad j = 1, 2, \ldots, n,$$

and assume that M has strictly positive entries, *i.e.*

$$M_{i,j} > 0, \qquad i, j = 1, 2, \ldots, n.$$

We let $\|x\|_1 = \sum_{k=1}^{n} |x_k|$ denote the ℓ^1 norm of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Prove the following statements using only Markov chain reasoning.

a) Show that M admits 1 as (right) eigenvalue and that the corresponding eigenspace has dimension 1.
b) Show that there exists a unique vector $y \in \mathbb{R}^n$ with positive components such that $My = y$ with $\|y\|_1 = 1$, which can be computed as $y = \lim_{k \to \infty} M^k x_0$ for any initial guess x_0 with positive components such that $\|x_0\|_1 = 1$.

Exercise 7.27 Consider an irreducible positive recurrent Markov chain $(X_n)_{n \geqslant 0}$ with unique stationary distribution π on a state space $\mathbb{S}$, and let

$$\tau_x := \inf\{n \geqslant 1 \ : \ X_n = x\}$$

denote the first return time to state $x \in \mathbb{S}$.

a) Let

$$R_n^x := \sum_{k=1}^{n} \mathbb{1}_{\{X_k = x\}}$$

denote the number of returns to state $x \in \mathbb{S}$ from time 1 to time n. Show that the stationary distribution $\pi = (\pi_x)_{x \in \mathbb{S}}$ satisfies

$$\pi_x = \lim_{n\to\infty} \frac{\mathbb{E}[R_n^x]}{n}, \qquad x \in \mathbb{S}.$$

Hint. Show that the limit satisfies $\pi = \pi P$.

b) Let

$$N_{x,y} := \sum_{n=1}^{\tau_x} \mathbf{1}_{\{X_n = y\}}$$

denote the number of visits to state y before the first return to state x. Show that we have

$$\pi_y = \frac{\mathbb{E}[N_{x,y} \mid X_0 = x]}{\mathbb{E}[\tau_x \mid X_0 = x]}, \qquad x, y \in \mathbb{S}.$$

Hint. Use the law of large numbers for regenerative processes.

c) Show that $N_{x,y}$ has a geometric distribution, and find its parameter in terms of $\alpha_{x,y} := \mathbb{P}(N_{x,y} \geqslant 1 \mid X_0 = x)$ and $\alpha_{y,x} := \mathbb{P}(N_{y,x} \geqslant 1 \mid X_0 = y)$, $x, y \in \mathbb{S}$.

d) Find a relation between π_x, π_y, $\alpha_{x,y}$, $\alpha_{y,x}$.
Hint. Recall that we have

$$\pi_x = \frac{1}{\mathbb{E}[\tau_x \mid X_0 = x]}, \qquad x \in \mathbb{S},$$

and

$$\sum_{k\geqslant 1} k r^{k-1} = \frac{1}{(1-r)^2},$$

for any $r \in [0, 1)$, see (A.4).

Problem 7.28 Consider a two-state Markov chain $(X_n)_{n\geqslant 0}$ on $\mathbb{S} = \{0, 1\}$, with transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \left[\begin{array}{cc} 1-a & a \\ b & 1-b \end{array}\right] \end{array},$$

where $a, b \in (0, 1)$.

a) Find the lowest eigenvalue λ of P.

b) Find the stationary distribution (π_0, π_1) of the chain $(X_n)_{n\geqslant 0}$.

c) Show by induction on $n \geqslant 0$ that

$$\left[\begin{array}{c} \mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right) \,\middle|\, X_0 = 0\right] \\ \mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right) \,\middle|\, X_0 = 1\right] \end{array}\right] = \left(\left[\begin{array}{cc} 1-a & a e^t \\ \\ b & (1-b)e^t \end{array}\right]\right)^n \left[\begin{array}{c} 1 \\ \\ 1 \end{array}\right],$$

$n \geqslant 0$, $t \in \mathbb{R}$. In the sequel, we assume that $(X_n)_{n\geqslant 0}$ is started in its *stationary distribution*, *i.e.*

$$\mathbb{P}(X_0 = 0) = \pi_0, \qquad \mathbb{P}(X_0 = 1) = \pi_1.$$

d) Show that for all $n \geqslant 1$ we have

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right]$$
$$= [\sqrt{\pi_0}, \sqrt{\pi_1}e^{t/2}]\left(\begin{bmatrix} \lambda+(1-\lambda)\pi_0 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & (\lambda+(1-\lambda)\pi_1)e^t \end{bmatrix}\right)^{n-1}\begin{bmatrix} \sqrt{\pi_0} \\ \sqrt{\pi_1}e^{t/2} \end{bmatrix}.$$

Hint. Diagonalize P as

$$\begin{bmatrix} 1-a & a \\ b & (1-b) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\pi_0}} & 0 \\ 0 & \frac{1}{\sqrt{\pi_1}} \end{bmatrix}\begin{bmatrix} \sqrt{\pi_0} & -\sqrt{\pi_1} \\ \sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}\begin{bmatrix} \sqrt{\pi_0} & \sqrt{\pi_1} \\ -\sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix}\begin{bmatrix} \sqrt{\pi_0} & 0 \\ 0 & \sqrt{\pi_1} \end{bmatrix},$$

and use the fact that

$$\begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & e^t \end{bmatrix}.$$

e) Find the largest eigenvalue $\mu(t)$ of the matrix

$$M(t) := \begin{bmatrix} \lambda+(1-\lambda)\pi_0 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & (\lambda+(1-\lambda)\pi_1)e^t \end{bmatrix}.$$

In the sequel, we assume that $\lambda \geqslant 0$.

f) Show that for all $n \geqslant 0$ and $t \in \mathbb{R}_+$ we have

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] \leqslant (\pi_0 + \pi_1 e^t)(\mu(t))^{n-1} \leqslant (\mu(t))^n.$$

Hint. Use e.g. Proposition 9 in Foucart (2010).

g) Using the Markov inequality, show that

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - \pi_1) \geqslant z\right) \leqslant \mathrm{e}^{-n((\pi_1+z)t-\log\mu(t))}, \qquad z > 0, \quad t > 0.$$

h) Show that for all $n \geqslant 1$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - \pi_1) \geqslant z\right) \leqslant \exp\left(-2\frac{1-\lambda}{1+\lambda}nz^2\right), \qquad z > 0.$$

Hint. Find the value $t(x)$ of $t > 0$ that maximizes $t \mapsto xt - \log \mu(t)$ for x fixed in $(0,1)$, and then show that

$$\frac{xt(x) - \log \mu(t(x))}{(x - \pi_1)^2} \geqslant 2\frac{1-\lambda}{1+\lambda}, \qquad x \in (0,1).$$

Problem 7.29 Let $(X_n)_{n\geqslant 0}$ denote an irreducible aperiodic Markov chain on a finite state space $\mathbb{S}$, with transition matrix $P = (P_{i,j})_{i,j\in\mathbb{S}}$ and stationary distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$. We let

$$R_n^i := \sum_{k=1}^n \mathbb{1}_{\{X_k = i\}}$$

denote the number of returns to state $i \in \mathbb{S}$ from time 1 to time n.

a) Show that the stationary distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ satisfies

$$\pi_i = \lim_{n\to\infty} \frac{\mathbb{E}[R_n^i]}{n}, \qquad i \in \mathbb{S}. \tag{7.3.4}$$

Hint. Show that the limit satisfies $\pi = \pi P$.

b) Define the sequence $(\tau_k)_{k\geqslant 1}$ recursively as

$$\tau_1 := \inf\{l > 1 \ : \ X_l = X_1\},$$

and

$$\tau_k := \inf\{l > \tau_{k-1} \ : \ X_l = X_1\}, \qquad k \geqslant 2.$$

Show, using e.g. Theorem 31 page 15 of Freedman (1983) and the law of large numbers for regenerative processes, see Corollary 14 page 106 of Serfozo (2009), that

$$\pi_i = \frac{\mathbb{E}\left[\sum_{j=1}^{\tau_1 - 1} \mathbb{1}_{\{X_j = i\}}\right]}{\mathbb{E}[\tau_1 - 1]}, \qquad i \in \mathbb{S}.$$

c) Let τ be a stopping time for $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$, $n \geqslant 0$, with $\mathbb{E}[\tau] < \infty$. By writing

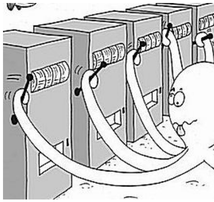
$$T := \inf\{l > \tau \ : \ X_l = X_1\}$$

as $T = \tau_\kappa$ where κ is a stopping time* for $(\mathcal{F}_{\tau_k})_{k\geqslant 1}$, show that

$$\pi_i = \frac{\mathbb{E}\left[\sum_{j=1}^{T-1} \mathbb{1}_{\{X_j = i\}}\right]}{\mathbb{E}[T - 1]}, \qquad i \in \mathbb{S}.$$

Hint. Use e.g. Theorem 2 of Chewi (2017).

* See *e.g.* § 2 of Chewi (2017).



Problem 7.30
Multi-Armed Bandits (MAB) have applications from recommender systems and information retrieval to healthcare and finance, due to its stellar performance combined with attractive properties, such as learning from less feedback, see Bouneffouf et al. (2020). For example, the Uber Data Science team leverages MAB testing to rank restaurants on the main feed of the Uber Eats app. The GrabFood "Recommended for You" widget also uses MABs for recommendation solutions.

We consider an N-arm bandit in which arm $n^\circ i$ is modeled by a two-state Markov chain $(X_n^{(i)})_{n\geqslant 0}$ on $\mathbb{S} := \{0,1\}$, with transition matrix $P^{(i)}$ and stationary distribution $(\pi_0^{(i)}, \pi_1^{(i)})$, $i = 1, \ldots, N$, ordered as $\pi_1^{(1)} \leqslant \cdots \leqslant \pi_1^{(N)}$. Given an $\{1, \ldots, N\}$-valued policy $(\alpha_k)_{k\geqslant 1}$, we let

$$T_n^{(i,\alpha)} := \sum_{k=1}^{n} \mathbb{1}_{\{\alpha_k = i\}}, \qquad i = 1, \ldots, N,$$

denote the number of times the arm i is selected by the policy $(\alpha_k)_{k\geqslant 1}$ until time $n \geqslant 1$. The reward of arm $n^\circ i$ after it has been pulled $n \geqslant 1$ times is $X_n^{(i)}$, and the *regret* $\mathcal{R}_n^\alpha$ at time n of the policy $(\alpha_k)_{k\geqslant 1}$ is given by

$$\mathcal{R}_n^\alpha := n\pi_1^{(N)} - \mathbb{E}\left[\sum_{i=1}^{N} \sum_{k=1}^{T_n^{(i,\alpha)}} X_k^{(i)}\right], \quad n \geqslant 1.$$

a) Bounded regret (Problem 7.28 continued).

i) Show that for any stopping time τ for $\mathcal{F}_n := \sigma(X_0^{(i)}, \ldots, X_n^{(i)})$, $n \geqslant 0$, letting $R_\tau^{(i)} := \sum_{k=1}^{\tau} \mathbb{1}_{\{X_k^{(i)}=1\}}$ denote the number of returns to state ① until time τ by the chain $(X_n^{(i)})_{n\geqslant 1}$, we have

$$\left|\mathbb{E}[R_\tau^{(i)}] - \pi_1^{(i)}\mathbb{E}[\tau]\right| \leqslant \mathbb{E}[T - \tau], \quad i = 1, \ldots, N.$$

Hint. Use the relations

$$R_{T-1}^{(i)} - (T - \tau) \leqslant R_T^{(i)} - (T - \tau) \leqslant R_\tau^{(i)} \leqslant R_{T-1}^{(i)}$$

in the notation of Question (c) of Problem 7.29.

ii) Show that

$$\left| \mathbb{E}\left[\sum_{i=1}^{N} \sum_{k=1}^{T_n^{(i,\alpha)}} X_k^{(i)} - \sum_{i=1}^{N} \pi_1^{(i)} T_n^{(i,\alpha)} \right] \right| \leqslant 2 \sum_{i=1}^{N} \operatorname*{Max}_{l,j \in \mathbb{S}} \mu_l^{(i)}(j), \quad n > N,$$

where $\mu_l^{(i)}(j)$ denotes the first return time of state $j \in \mathbb{S}$ from state $l \in \mathbb{S}$ by the chain $(X_n^{(i)})_{n \geqslant 0}$.

iii) Show that the *regret* $\mathcal{R}_n^\alpha$ of the policy $(\alpha_k)_{k \geqslant 1}$ is bounded as

$$\mathcal{R}_n^\alpha \leqslant \overline{\mathcal{R}}n_n^\alpha + K, \qquad n > N,$$

for some constant $K > 0$ independent of $n \geqslant 1$, where $\overline{\mathcal{R}}n_n^\alpha$ is the *modified* regret defined as

$$\overline{\mathcal{R}}n_n^\alpha := n\pi_1^{(N)} - \mathbb{E}\left[\sum_{i=1}^{N} \pi_1^{(i)} T_n^{(i,\alpha)} \right], \quad n \geqslant 1.$$

b) Learning at the $\log n$ speed. Let

$$\widehat{m}_n^{(i,\alpha)} := \frac{1}{T_n^{(i,\alpha)}} \sum_{k=1}^{T_n^{(i,\alpha)}} X_k^{(i)}$$

denote the sample average reward obtained from arm $n^\circ i$ until time $n \geqslant 1$ under the policy $(\alpha_k)_{k \geqslant 1}$.

Given $L > 0$, we define the policy $(\alpha_n^*)_{n \geqslant 1}$ by $\alpha_n^* := n$ for $n = 1, \ldots, N$, and for $n > N$ we let α_n^* be the index $i \in \{1, \ldots, N\}$ that maximizes the quantity

$$\widehat{m}_{n-1}^{(i,\alpha^*)} + \sqrt{\frac{L \log n}{T_{n-1}^{(i,\alpha^*)}}}.$$

i) Let $1 \leqslant i < N$ and $n \geqslant N$. Show by contradiction that if $\alpha_n^* = i$, then at least one of the following three conditions must hold:

$$\begin{cases} \widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\dfrac{L \log n}{T_{n-1}^{(N,\alpha^*)}}} \leqslant \pi_1^{(N)}, \\ \widehat{m}_{n-1}^{(i,\alpha^*)} > \pi_1^{(i)} + \sqrt{\dfrac{L \log n}{T_{n-1}^{(i,\alpha^*)}}}, \\ T_{n-1}^{(i,\alpha^*)} < \dfrac{4L \log n}{(\pi_1^{(N)} - \pi_1^{(i)})^2}. \end{cases}$$

ii) Show that letting $\widehat{n}_i := 4L(\log n)/(\pi_1^{(N)} - \pi_1^{(i)})^2$, we have

$$\mathbb{E}\big[T_n^{(i,\alpha^*)}\big] \leqslant \widehat{n}_i + \sum_{\widehat{n}_i<k\leqslant n}\left(\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant \pi_1^{(N)}\right)\right.$$
$$\left. + \mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > \pi_1^{(i)} + \sqrt{\frac{L\log k}{T_{k-1}^{(i,\alpha^*)}}}\right)\right),$$

$1 \leqslant i < N$, $n \geqslant N$.

iii) Letting λ_i denote the smallest eigenvalue of $P^{(i)}$, we assume that $\min_{1\leqslant i\leqslant N}\lambda_i \geqslant 0$, let $\lambda := \operatorname{Max}_{1\leqslant i\leqslant N}\lambda_i$, and assume that $L > (1+\lambda)/(1-\lambda)$.

Show that

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant \pi_1^{(N)}\right) \leqslant \frac{1}{k^{2L(1-\lambda)/(1+\lambda)-1}}$$

and

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > \pi_1^{(i)} + \sqrt{\frac{L\log k}{T_{k-1}^{(i,\alpha^*)}}}\right) \leqslant \frac{1}{k^{2L(1-\lambda)/(1+\lambda)-1}},$$

$i = 1,\ldots,N$, $k > N$.

Hint. Apply the result of Question (A)-(8) of Assignment 1.

iv) Show that the modified regret can be bounded for any $L > (1+\lambda)/(1-\lambda)$ by

$$\overline{\mathcal{R}}n_n^{\alpha^*} \leqslant \sum_{i=1}^{N-1}\frac{\pi_1^{(N)}-\pi_1^{(i)}}{L(1-\lambda)/(1+\lambda)-1} + (\log n)\sum_{i=1}^{N-1}\frac{4L}{\pi_1^{(N)}-\pi_1^{(i)}}, \qquad n > N.$$

Hint. Use a comparison argument between series and integrals.

Problem 7.31 (Aldous and Diaconis (1986), Jonasson (2009)). *Random shuffling* is applied to a deck of $N = 52$ cards by inserting the top card back into the deck at a random location $i \in \{1,\ldots,52\}$ chosen uniformly among $N = 52$ possible positions.

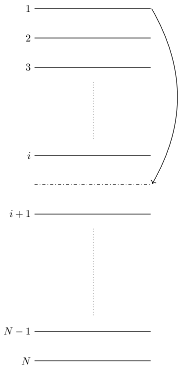


Fig. 7.6: Top to random shuffling.

More formally, consider the Markov chain $(X_n)_{n\geqslant 0}$ on the group

$$\mathbb{S}_N = \{(e_1, \ldots, e_N) \ : \ e_1, \ldots, e_N \in \{1, \ldots, N\}, \ e_i \neq e_j, \ 1 \leqslant i \neq j \leqslant N\}$$

of $N!$ permutations of $(1, \ldots, N)$, built by applying the cycle permutation of indexes

$$(1, \ldots, i) \mapsto (2, \ldots, i, 1)$$

to X_n for some uniformly chosen $i \in \{1, \ldots, N\}$. The transition matrix P of the chain is given by

$$\mathbb{P}(X_{n+1} = (e_1, \ldots, e_N) \mid X_n = (e_1, \ldots, e_N)) := \frac{1}{N},$$

and

$$\mathbb{P}(X_{n+1} = (e_2, \ldots e_i, e_1, e_{i+1}, \ldots, e_N) \mid X_n = (e_1, \ldots, e_N)) := \frac{1}{N},$$

$i = 2, \ldots, N$, with $P_{\sigma,\eta} := 0$ in all other cases, with $\sigma, \eta \in \mathbb{S}_N$.

At time 0 we choose to start with the initial condition $X_0 := (1, \ldots, N)$. We also let $T_0 := 0$, and for $k = 1, \ldots, N-1$ we denote by T_k the first time the original bottom card has moved up to the rank $N-k$ in the deck. Note that at time T_{N-1}, the original bottom card should have moved to the top of the deck.

a) Find the probability distribution

$$\mathbb{P}(T_k - T_{k-1} = m), \quad m \geqslant 1, \quad \text{for} \quad k = 1, \ldots, N-1.$$

Hint. This is a geometric distribution. Find its parameter depending on $k = 1, \ldots, N-1$.

b) Find the mean time $\mathbb{E}[T_k]$ it takes until the original bottom card has moved to the position $N-k$, $k = 1, \ldots, N-1$.

Hint. Use the telescoping identity

$$T_k = (T_k - T_{k-1}) + (T_{k-1} - T_{k-2}) + \cdots + (T_2 - T_1) + (T_1 - T_0).$$

c) Compute $\mathrm{Var}[T_{N-1}]$, and show that $\mathrm{Var}[T_{N-1}] \leqslant CN^2$ for some constant $C > 0$.

Hint. The random variables $T_k - T_{k-1}$, $k = 1, \ldots, N-1$, are independent.

d) Show that for any $a > 0$ we have

$$\lim_{N\to\infty} \mathbb{P}(T_{N-1} > (1+a)N \log N) = 0.$$

Hint. Use Chebyshev's inequality

$$\mathbb{P}(Z - \mathbb{E}[Z] \geqslant x) \leqslant \frac{1}{x^2} \mathrm{Var}[Z], \qquad x > 0,$$

and the bound

$$\sum_{k=1}^{N-1} \frac{1}{k} \leqslant 1 + \log N, \qquad N \geqslant 1.$$

e) What is the distribution of X_n given that $n > T_{N-1}$?

Hint. The answer is intuitive. No proof is required.

f) Based on the answers to Questions (d)-(e), what do you conclude?

Problem 7.32 *Time reversibility* is a fundamental issue in physics as it is akin to the idea of "traveling backward in time". This problem studies the reversibility of Markov chains, and applies it to the computation of stationary and limiting distributions. Given $N \geqslant 1$ and $(X_k)_{k=0,1,\ldots,N}$ a Markov chain with transition matrix P on a state space $\mathbb{S}$, we let

$$Y_k := X_{N-k}, \qquad k = 0, 1, \ldots, N,$$

denote the *time reversal* of $(X_k)_{k=0,1,\ldots,N}$.

a) Assume that X_k has same distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ for every $k = 0, 1, \ldots, N$, *i.e.*

$$\mathbb{P}(X_k = i) = \pi_i, \qquad i \in \mathbb{S}, \quad k = 0, 1, \ldots, N.$$

Show that the process $(Y_k)_{k=0,1,\ldots,N}$ is a (forward) Markov chain (*i.e.* $(X_k)_{k=0,1,\ldots,N}$ has the backward Markov property), and find the transition matrix

$$[\mathbb{P}(Y_{n+1} = j \mid Y_n = i)]_{i,j\in\mathbb{S}}$$

in terms of P and π.

Hint: Use the basic definition of conditional probabilities to compute

$$[\mathbb{P}(Y_{n+1} = j \mid Y_n = i)]_{i,j\in\mathbb{S}},$$

and then show that $(Y_n)_{n=0,1,\ldots,N}$ has the Markov property

$$\mathbb{P}(Y_{n+1} = j \mid Y_n = i_n, \ldots, Y_0 = i_0) = \mathbb{P}(Y_{n+1} = j \mid Y_n = i_n).$$

b) We say that $(X_k)_{k=0,1,\ldots,N}$ is *reversible* for π when $(X_k)_{k=0,1,\ldots,N}$ and $(Y_k)_{k=0,1,\ldots,N}$ have same transition probabilities.

Write down this reversibility condition in terms of P and π. From now on we refer to that condition as the *detailed balance condition*, which can be stated independently of N.

Hint: By "same transition probabilities" we mean

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(Y_{n+1} = j \mid Y_n = i).$$

c) Show that if $(X_k)_{k\geqslant 0}$ is reversible for π, then π is also a stationary distribution for $(X_k)_{k\geqslant 0}$.

Hint: The fact that $\sum_i P_{j,i} = 1$ plays a role here.

d) Show that if an irreducible positive recurrent aperiodic Markov chain is reversible for its stationary distribution π then we have

$$P_{k_1,k_2}P_{k_2,k_3}\cdots P_{k_{n-1},k_n}P_{k_n,k_1} = P_{k_1,k_n}P_{k_n,k_{n-1}}\cdots P_{k_3,k_2}P_{k_2,k_1} \quad (7.3.5)$$

for all sequences $\{k_1, k_2, \ldots, k_n\}$ of states and $n \geqslant 2$.

Hint: This is a standard algebraic manipulation.

e) Show that conversely, if an irreducible positive recurrent aperiodic Markov chain satisfies Condition (7.3.5) for all sequences $\{k_1, k_2, \ldots, k_n\}$ of states, $n \geqslant 2$, then it is reversible for its stationary distribution π.

Hint: This question is more difficult and here you need to apply Theorem 7.8.

f) From now on we assume that $\mathbb{S} = \{0, 1, \ldots, M\}$ and that P is the transition matrix

$$P_{i,i+1} = \frac{1}{2} - \frac{i}{2M}, \qquad P_{i,i} = \frac{1}{2}, \qquad P_{i,i-1} = \frac{i}{2M}, \qquad 1 \leqslant i \leqslant M-1,$$

of the *modified* Ehrenfest chain, with $P_{0,0} = P_{0,1} = P_{M,M-1} = P_{M,M} = 1/2$.

Find a probability distribution π for which the chain $(X_k)_{k\geqslant 0}$ is reversible.

Hint: The reversibility condition will yield a relation that can be used to compute π by induction. Remember to make use of the condition $\sum_i \pi_i = 1$.

g) Confirm that the distribution of Question (f) is invariant (or stationary) by checking explicitly that the equality

$$\pi = \pi P$$

does hold.

h) Show, by quoting the relevant theorem, that π is also the limiting distribution of the *modified* Ehrenfest chain $(X_k)_{k \geqslant 0}$.

i) Show, by the result of Question (h), that

$$\lim_{n \to \infty} \mathbb{E}[X_n \mid X_0 = i] = \frac{M}{2},$$

for all $i = 0, 1, \ldots, M$.

j) Show, using first step analysis and induction on $n \geqslant 0$, that we have

$$\mathbb{E}\left[X_n - \frac{M}{2} \;\Big|\; X_0 = i\right] = \left(i - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^n, \qquad n \geqslant 0,$$

for all $i = 0, 1, \ldots, M$, and that this relation can be used to recover the result of Question (i).

Hint: Letting

$$h_n(i) = \mathbb{E}\left[X_n - \frac{M}{2} \;\Big|\; X_0 = i\right], \qquad n \geqslant 0,$$

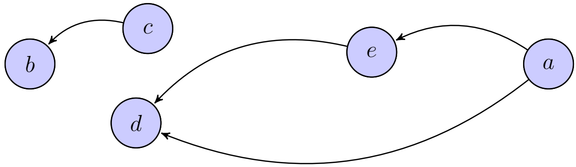
in order to prove the formula by induction one has to

(i) show that the formula holds for $h_0(i)$ when $n = 0$;

(ii) show that assuming that the formula holds for $h_n(i)$, it then holds for $h_{n+1}(i)$.

It can help to start by proving the formula for $h_1(i)$ when $n = 1$ by first step analysis.

Problem 7.33 PageRank™ algorithm. We consider the ranking of five web pages a, b, c, d, e which are linked according to the following graph.



The algorithm works by constructing a self-improving random sequence $(X_n)_{n\geqslant 0}$ which is supposed to "converge" to the best possible search result. Given a search result $X_n = x \in \{a, b, c, d, e\}$, we choose the next search result X_{n+1} with the conditional probability

$$\mathbb{P}(X_{n+1} = y \mid X_n = x) = \frac{1}{n_x}\mathbf{1}_{\{x\to y\}},$$

where n_x denotes the number of outgoing links from x and "$x \to y$" means that x can lead to y in the graph.

a) Model the process $(X_n)_{n\geqslant 0}$ as a Markov chain, and find its transition matrix.

b) Draw the graph of the chain $(X_n)_{n\geqslant 0}$.

   Is the chain $(X_n)_{n\geqslant 0}$ reducible?

c) Does the Markov chain $(X_n)_{n\geqslant 0}$ admit a limiting distribution independent of the initial state?

d) Does the Markov chain $(X_n)_{n\geqslant 0}$ admit a stationary distribution? Find all stationary distribution(s) of the chain $(X_n)_{n\geqslant 0}$.

e) In PageRank™-type algorithms, one typically chooses to perturb the transition matrix P into the new matrix

$$\widetilde{P} := \frac{\varepsilon}{n}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} + (1-\varepsilon)P, \qquad \varepsilon \in (0,1),$$

   with $n = 5$ here, where $1-\varepsilon$ is referred to as the *damping factor*.

   Show that $\widetilde{P}$ is a Markov transition matrix and that the corresponding chain $(\widetilde{X}_n)_{n\geqslant 1}$ is irreducible and aperiodic.

f) Show that $\widetilde{P}$ admits a stationary distribution $\tilde{\pi}$ that satisfies

$$\tilde{\pi} = \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P,$$

   and that all probabilities in $\tilde{\pi}$ are greater than $\varepsilon/5$.

g) Compute the stationary distribution of $\widetilde{P}$.

h) Provide a ranking of the states $\{a, b, c, d, e\}$ based on the stationary distribution $\tilde{\pi}$.
i) Compute the mean return times for $\widetilde{P}$, and show that they are all below $5/\epsilon$.

Problem 7.34 Meta search engines (Schalekamp and van Zuylen (2009)). A **meta search engine** tries to provide a single optimized ranking of search results $\{a, b, c, d, e\}$ based on the outputs of *4* different search engines denoted $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$, a technique known as *rank aggregation*.

Five possible search results a, b, c, d, e have been respectively ranked as

Rank	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$
1	b	c	d	e
2	c	b	e	a
3	d	d	a	d
4	a	e	b	b
5	e	a	c	c

by the Search Engines $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$.

a) A state $y \in \{a, b, c, d, e\}$ is said to be better ranked than another state $x \in \{a, b, c, d, e\}$, and we write $x \preceq y$ and $y \succeq x$ if y ranks higher than x for **at least three** of the four search engines. We also write "$x \npreceq y$" when neither "$x \preceq y$" nor "$x \succeq y$" is satisfied.

   Complete the ranking table

$\preceq$	a	b	c	d	e
a	=				
b		=			
c			=		
d				=	
e					=

   with "$x \preceq y$", "$x \succeq y$" or "$x \npreceq y$" at the positions (x, y), $x, y \in \{a, b, c, d, e\}$. The diagonal entries are not relevant.
b) The meta search engine works by constructing a self-improving random sequence $(X_n)_{n \geqslant 0}$ which is supposed to "converge" to the best possible search result based on the data of the four rankings.

   Given a search result $X_n = x \in \{a, b, c, d, e\}$, we choose the next search result X_{n+1} by assigning probability $1/5$ to each of the search results that are *better ranked* than x. If no search result is better than x, then we keep $X_{n+1} = x$.

What is the state space of the sequence $(X_n)_{n\geqslant 0}$?

Model the process $(X_n)_{n\geqslant 0}$ as a Markov chain, and complete the transition matrix

$$P = \begin{bmatrix} 3/5 & \square & \square & \square & \square \\ 0 & \square & \square & \square & \square \\ \square & 1/5 & \square & \square & \square \\ \square & \square & \square & \square & 0 \\ \square & \square & \square & 1/5 & \square \end{bmatrix}$$

c) Draw the graph of the chain $(X_n)_{n\geqslant 0}$.

Is the chain $(X_n)_{n\geqslant 0}$ reducible?

d) Does the Markov chain $(X_n)_{n\geqslant 0}$ admit a limiting distribution independent of its initial state?

e) Does the Markov chain $(X_n)_{n\geqslant 0}$ admit a stationary distribution? Find all possible stationary distribution(s) of the chain $(X_n)_{n\geqslant 0}$.

f) In PageRank™-type algorithms, one typically chooses to perturb the transition matrix P into the new matrix

$$\widetilde{P} := \frac{\varepsilon}{n} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} + (1-\varepsilon)P,$$

with $n = 5$ here, and $\varepsilon \in (0,1)$. Show that $\widetilde{P}$ is a Markov transition matrix and that the corresponding chain $(\widetilde{X}_n)_{n\geqslant 1}$ is irreducible and aperiodic.

g) Show that $\widetilde{P}$ admits a stationary distribution $\tilde{\pi}$ that satisfies

$$\tilde{\pi} = \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P,$$

and that all probabilities in $\tilde{\pi}$ are greater than $\varepsilon/5$.

h) Compute the stationary distribution of $\widetilde{P}$.

i) Provide a ranking of the states $\{a,b,c,d,e\}$ based on the stationary distribution $\tilde{\pi}$.

j) Compute the mean return times for $\widetilde{P}$, and show that they are all below $5/\epsilon$.

Problem 7.35 (Levin et al. (2009)). Convergence to equilibrium. In this problem we derive quantitative bounds for the convergence of a Markov chain to its stationary distribution π. Let P be the transition matrix of a discrete-time Markov chain $(X_n)_{n\geqslant 0}$ on $\mathbb{S} = \{1,2,\ldots,N\}$. Given two probability distributions $\mu = [\mu_1,\mu_2,\ldots,\mu_N]$ and $\nu = [\nu_1,\mu_2,\ldots,\nu_N]$ on $\{1,2,\ldots,N\}$, the *total variation distance* between μ and ν is defined as

$$\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2}\sum_{k=1}^{N} |\mu_k - \nu_k|.$$

Recall that the vector $\mu P^n = ([\mu P^n]_i)_{i=1,2,\ldots,N}$ denotes the probability distribution of the chain at time $n \in \mathbb{N}$, given it was started with the initial distribution $\mu = [\mu_1, \mu_2, \ldots, \mu_N]$, *i.e.* we have, using matrix product notation,

$$\mathbb{P}(X_n = i) = \sum_{j=1}^{N} \mathbb{P}(X_n = i \mid X_0 = j)\mathbb{P}(X_0 = j) = \sum_{j=1}^{N} \mu_j [P^n]_{j,i} = [\mu P^n]_i,$$

$i = 1, 2, \ldots, N$.

a) Show that for any two probability distributions $\mu = [\mu_1, \mu_2, \ldots, \mu_N]$ and $\nu = [\nu_1, \nu_2, \ldots, \nu_N]$ on $\{1, 2, \ldots, N\}$ we always have $\|\mu - \nu\|_{\mathrm{TV}} \leqslant 1$.

b) Show that for any two probability distributions $\mu = [\mu_1, \mu_2, \ldots, \mu_N]$ and $\nu = [\nu_1, \nu_2, \ldots, \nu_N]$ on $\{1, 2, \ldots, N\}$ and any Markov transition matrix P we have

$$\|\mu P - \nu P\|_{\mathrm{TV}} \leqslant \|\mu - \nu\|_{\mathrm{TV}}.$$

Hint: Use the triangle inequality

$$\left|\sum_{k=1}^{n} x_k\right| \leqslant \sum_{k=1}^{n} |x_k|, \qquad x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

c) Assume that the chain with transition matrix P admits a stationary distribution $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$. Show that for any probability distribution $\mu = [\mu_1, \mu_2, \ldots, \mu_N]$ we have

$$\|\mu P^{n+1} - \pi\|_{\mathrm{TV}} \leqslant \|\mu P^n - \pi\|_{\mathrm{TV}}, \qquad n \geqslant 0.$$

d) Show that the *distance from stationarity*, defined as

$$d(n) := \operatorname*{Max}_{k=1,2,\ldots,N} \|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}}, \qquad n \geqslant 0,$$

satisfies $d(n+1) \leqslant d(n)$, $n \in \mathbb{N}$.

e) Assume that all entries of P are strictly positive. Explain why the chain is aperiodic and irreducible, and why it admits a limiting and stationary distribution.

In what follows we assume that P admits an invariant (or stationary) distribution $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$ such that $\pi P = \pi$, and that

$$P_{i,j} \geqslant \theta \pi_j, \quad \text{for all } i, j = 1, 2, \ldots, N, \tag{7.3.6}$$

for some $0 < \theta < 1$. We also let

$$\Pi := \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_N \end{bmatrix},$$

hence (7.3.6) reads $P \geqslant \theta\Pi$.

f) Show that for all $0 < \theta < 1$ the matrix

$$Q_\theta := \frac{1}{1-\theta}(P - \theta\Pi)$$

is the transition matrix of a Markov chain on $\mathbb{S} = \{1, 2, \ldots, N\}$.

g) Show by induction on $n \in \mathbb{N}$ that we have

$$P^n - \Pi = (1-\theta)^n (Q_\theta^n - \Pi), \qquad n \in \mathbb{N}.$$

h) Show that given any $X_0 = k = 1, 2, \ldots, N$ the total variation distance between the distribution

$$\begin{aligned} [P^n]_{k,\cdot} &= ([P^n]_{k,1}, \ldots, [P^n]_{k,N}) \\ &= [\mathbb{P}(X_n = 1 \mid X_0 = k), \ldots, \mathbb{P}(X_n = N \mid X_0 = k)] \end{aligned}$$

of the chain at time n and the stationary distribution $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$ satisfies

$$\|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} \leqslant (1-\theta)^n, \qquad n \geqslant 1, \quad k = 1, 2, \ldots, N.$$

Conclude that we have $d(n) \leqslant (1-\theta)^n$, $n \geqslant 1$.

i) Show that the *mixing time* of the chain with transition matrix P, defined as

$$t_{\mathrm{mix}} := \min\{n \geqslant 0 \ : \ d(n) \leqslant 1/4\},$$

satisfies

$$t_{\mathrm{mix}} \leqslant \left\lceil \frac{\log 1/4}{\log(1-\theta)} \right\rceil.$$

j) Find the optimal value of θ satisfying the condition $P_{i,j} \geqslant \theta\pi_j$ for all $i, j = 1, 2, \ldots, N$ for the chain of Exercise 4.12 with $N = 3$.

Problem 7.36 (Agapie and Höns (2007)). The voter model is a particular case of the Ising model. This model has applications to spatial statistics, image analysis and segmentation, opinion studies, urban segregation, language change, metal alloys, magnetic materials, liquid/gas coexistence, phase transitions, plasmas, cell membranes in biophysics, ...

In dimension one, the model is built on the state space $\mathbb{S} := \{-1, +1\}^N$ made of elements $z = (z_k)_{1 \leqslant k \leqslant N} \in \mathbb{S}$ whose components $z_k \in \{-1, 1\}$, $k = 1, 2, \ldots N$, are called *spins*.

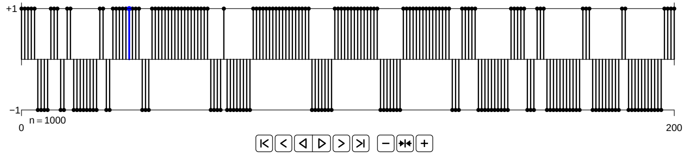


Fig. 7.7: Simulation of the voter model with $N = 199$, $p = 0.98$, and $z_0 = z_{N+1} = +1$.*

We consider a Markov chain $(Z_n)_{n \geqslant 0}$ *on the state space* $\mathbb{S} = \{-1, +1\}^N$, whose transitions from an initial configuration $Z_0 = z = (z_k)_{1 \leqslant k \leqslant N}$ to a new configuration $Z_1 = \tilde{z} = (\tilde{z}_k)_{1 \leqslant k \leqslant N}$ are defined as follows:

First, choose one component z_k in $z = (z_k)_{1 \leqslant k \leqslant N}$ with probability $1/N$, $k = 1, 2, \ldots, N$, and then consider the following cases:

(i) if $(z_{k-1}, z_{k+1}) = (-1, +1)$ or $(z_{k-1}, z_{k+1}) = (+1, -1)$:
$\Rightarrow$ flip the sign of z_k, *i.e.* set $\tilde{z}_k := \pm z_k$ with probability $1/2$,

(ii) if $(z_{k-1}, z_{k+1}) = (+1, +1)$:
$\Rightarrow$ set $\tilde{z}_k := +1$ with probability $p > 0$, and $\tilde{z}_k := -1$ with probability $q > 0$.

(iii) if $(z_{k-1}, z_{k+1}) = (-1, -1)$:
$\Rightarrow$ set $\tilde{z}_k := -1$ with probability $p > 0$, and $\tilde{z}_k := +1$ with probability $q > 0$,

where $p + q = 1$. The probabilities p and q can be respectively viewed as the probabilities of "agreeing", resp. "disagreeing" with two neighbors sharing the same opinion. The boundary conditions z_0 and z_{N+1} are arbitrarily fixed, and the corresponding instructions can be coded in R as follows:

```
if (z[k-1]!=z[k+1]) z[k]=sample(c(-1,1), 1,prob=c(0.5,0.5))
if (z[k-1]==1 && z[k+1]==1) z[k]=sample(c(-1,1), 1, prob=c(q,p))
if (z[k-1]==-1 && z[k+1]==-1) z[k]=sample(c(-1,1), 1, prob=c(p,q))
```

a) Find the cardinality of the state space $\mathbb{S}$.

* Animated figure (works in Acrobat Reader).

b) Replace the question marks (?) below with the corresponding transition probabilities given that $Z_0 = z = (z_k)_{1\leqslant k\leqslant N}$:

   (i) if $(z_{k-1}, z_k, z_{k+1}) = (-1, \pm 1, +1)$ or $(z_{k-1}, z_k, z_{k+1}) = (+1, \pm 1, -1)$,

$$\mathbb{P}(Z_1 = (z_1, \ldots, z_{k-1}, -z_k, z_{k+1}, \ldots, z_N) \mid Z_0 = z) := (?),$$

   $k = 1, 2, \ldots, N$,

   (ii) if $(z_{k-1}, z_k, z_{k+1}) = (+1, +1, +1)$ or $(z_{k-1}, z_k, z_{k+1}) = (-1, -1, -1)$,

$$\mathbb{P}(Z_1 = (z_1, \ldots, z_{k-1}, -z_k, z_{k+1}, \ldots, z_N) \mid Z_0 = z) := (?),$$

   $k = 1, 2, \ldots, N$,

   (iii) if $(z_{k-1}, z_k, z_{k+1}) = (+, -1, +1)$ or $(z_{k-1}, z_k, z_{k+1}) = (-1, +1, -1)$,

$$\mathbb{P}(Z_1 = (z_1, \ldots, z_{k-1}, -z_k, z_{k+1}, \ldots, z_N) \mid Z_0 = z) := (?),$$

   $k = 1, 2, \ldots, N$.

c) Show that, based on the three above cases (i)-(ii)-(iii), the transition probabilities of $(Z_n)_{n\geqslant 0}$ take the general form

$$\mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z) = \frac{1}{N(1 + (p/q)^{z_k(z_{k-1}+z_{k+1})/2})}, \qquad k = 1, 2, \ldots, N, \tag{7.3.7}$$

   where $z = (z_1, \ldots, z_N)$ and

$$\bar{z}^k := (z_1, \ldots, z_{k-1}, -z_k, z_{k+1}, \ldots, z_N) \tag{7.3.8}$$

   denotes the state $z \in \mathbb{S}$ after flipping its kth component z_k.

d) Compute $\mathbb{P}(Z_1 = z \mid Z_0 = z)$ for all $z \in \mathbb{S}$ using the complement rule, Relation (7.3.7), and the law of total probability.

e) Taking $N = 3$ and setting $\underline{z_0 = z_4 = -1}$, *i.e.* $(z_0, z_1, z_2, z_3, z_4)$ takes the form $(-, \pm, \pm, \pm, -)$, complete the missing entries following transition probability matrix P of $(Z_n)_{n\geqslant 0}$ on the state space $\mathbb{S} = \{- - -, - - +, - + -, - - +, + - -, + - +, + + -, + + +\}$:

$$
\begin{array}{c}
\\
---\\
--+\\
-+-\\
-++\\
+--\\
+-+\\
++-\\
+++
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
--- & --+ & -+- & -++ & +-- & +-+ & ++- & +++
\end{array}\\
\left[\begin{array}{cccccccc}
(?) & (?) & q/3 & 0 & q/3 & 0 & 0 & 0\\
(?) & 1/2 & 0 & (?) & 0 & q/3 & 0 & 0\\
(?) & 0 & (1+q)/3 & (?) & 0 & 0 & 1/6 & 0\\
0 & 1/6 & (?) & (?) & 0 & 0 & 0 & 1/6\\
p/3 & 0 & 0 & 0 & 1/2 & (?) & (?) & 0\\
0 & p/3 & 0 & 0 & p/3 & q & 0 & p/3\\
0 & 0 & 1/6 & 0 & (?) & 0 & 1/2 & (?)\\
0 & 0 & 0 & 1/6 & 0 & q/3 & 1/6 & (?)
\end{array}\right]
\end{array}.
$$

Is the chain $(Z_n)_{n\geqslant 0}$ reducible? Why?

f) Does the Markov chain $(Z_n)_{n\geqslant 0}$ admit a limiting distribution? A stationary distribution? Why?

g) Show that if a probability distribution $(\pi_z)_{z\in\mathbb{S}}$ on $\mathbb{S}$ satisfies the relation

$$\frac{\pi_{\bar{z}^k}}{\pi_z} = \frac{\mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z)}{\mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k)}, \qquad k = 1, 2, \ldots, N, \quad z \in \mathbb{S}, \tag{7.3.9}$$

where $\bar{z}^k$ is defined in (7.3.8), then $(\pi_z)_{z\in\mathbb{S}}$ is a *stationary distribution* for the chain $(Z_n)_{n\geqslant 0}$, *i.e.* we have

$$\Big(\mathbb{P}(Z_0 = z) = \pi_z, \ \forall\, z \in \mathbb{S}\Big) \quad \Longrightarrow \quad \Big(\mathbb{P}(Z_1 = z) = \pi_z, \ \forall\, z \in \mathbb{S}\Big).$$

Hint: Start from the law of total probability

$$\begin{aligned}\mathbb{P}(Z_1 = z) &= \mathbb{P}(Z_1 = z \mid Z_0 = z)\mathbb{P}(Z_0 = z)\\ &\quad + \sum_{k=1}^{N} \mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k)\mathbb{P}(Z_0 = \bar{z}^k),\end{aligned}$$

and show, using (7.3.9), that the above equals π_z if $\mathbb{P}(Z_0 = z) = \pi_z$ for all $z \in \mathbb{S}$.

h) Show that

$$\frac{\mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z)}{\mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k)} = \left(\frac{q}{p}\right)^{z_k(z_{k-1}+z_{k+1})/2}, \qquad k = 1, 2, \ldots, N, \quad z \in \mathbb{S}. \tag{7.3.10}$$

Hint: Use Equation (7.3.7) and the relations

$$\frac{1 + (q/p)}{1 + (p/q)} = \frac{q}{p} \quad \text{and} \quad \frac{1 + (p/q)}{1 + (q/p)} = \frac{p}{q}.$$

i) Show that the probability distribution $(\pi_z)_{z\in\mathbb{S}}$ defined as

$$\pi_z := C_\beta \exp\left(\beta \sum_{l=0}^{N} z_l z_{l+1}\right), \qquad z \in \mathbb{S}, \tag{7.3.11}$$

is the stationary and limiting distribution of $(Z_n)_{n\geqslant 0}$, where C_β is a normalization constant* and β is to be given in terms of p and q.

Hint: Using (7.3.10), show that $(\pi_z)_{z\in\mathbb{S}}$ defined in (7.3.11) satisfies (7.3.9).

j) Taking $N = 3$ and $\underline{z_0 = z_4 = +1}$, *i.e.* $(z_0, z_1, z_2, z_3, z_4)$ takes the form $(+, \pm, \pm, \pm, +)$, compute the limiting distribution of each of the 8 configurations in

$$\mathbb{S} = \{---, --+, -+-, --+, +--, +-+, ++-, +++\},$$

and find the value of C_β.

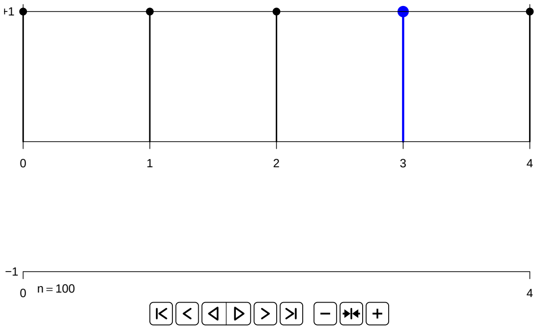


Fig. 7.8: Simulation of the voter model with $N = 3$, $p = \sqrt{0.75}$, and $z_0 = z_4 = +1$.†

Problem 7.37 (Lezaud (1998)). Consider an *irreducible*, *reversible*‡, Markov chain $(X_n)_{n\geqslant 0}$ with transition matrix $P = (P_{i,j})_{1\leqslant i,j\leqslant d}$ and admitting a stationary distribution π on the finite state space $\mathbb{S} = \{1, 2, \ldots, d\}$. For any function $f : \mathbb{R} \to \mathbb{R}$, we let D_f denote the diagonal matrix

* The normalization constant C_β does not have to be computed here.

† Animated figure (works in Acrobat Reader).

‡ *i.e.* $\pi_i P_{i,j} = \pi_j P_{j,i}$, $i, j = 1, \ldots, d$.

$$D_f = \begin{bmatrix} f(1) & 0 & 0 & 0 \cdots & 0 \\ 0 & f(2) & 0 & 0 \cdots & 0 \\ 0 & 0 & f(3) & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \ddots & \vdots \\ 0 & 0 & 0 & 0 \cdots & f(d) \end{bmatrix}.$$

We use the scalar product $\langle \cdot , \cdot \rangle$ and norm $\| \cdot \|$ on $\mathbb{R}^d$ defined as

$$\langle u, v \rangle := \sum_{l=1}^d u(l) v(l) \pi_l, \quad \|u\|^2 := \sum_{l=1}^d |u(l)|^2 \pi_l, \qquad u, v \in \mathbb{R}^d,$$

with the Cauchy-Schwarz inequality

$$|\langle u, v \rangle| \leqslant \|u\| \cdot \|v\|, \qquad u, v \in \mathbb{R}^d.$$

Recall that the norm $\| \cdot \|$ also defines a matrix norm on $\mathbb{R}^{d \times d}$ as

$$\|M\| = \operatorname*{Sup}_{\substack{u \in \mathbb{R}^d \\ u \neq 0}} \frac{\|Mu\|}{\|u\|} = \operatorname*{Sup}_{\|u\|=1} \|Mu\|, \qquad M \in \mathbb{R}^{d \times d}.$$

In what follows, we assume that $(X_n)_{n \geqslant 0}$ is started with π as initial distribution, and $f : \{1, \ldots, d\} \to \mathbb{R}$ denotes any function such that $\|f\|_\infty \leqslant 1$ and $\mathbb{E}[f(X_n)] = 0$, $n \geqslant 0$.

a) Show that 1 is an eigenvalue of single multiplicity for P, and give its eigenvector.
Hint. Use the irreducibility of $(X_n)_{n \geqslant 0}$ and the Perron-Frobenius theorem.

b) Write down the matrix Π of the orthogonal* projection operator on the eigenvector of P with eigenvalue 1.

c) Show by induction on $n \geqslant 0$ that for any state $k \in \{1, \ldots, d\}$ and $\alpha \in \mathbb{R}$, we have

$$\mathbb{E}\left[\exp\left(\alpha \sum_{l=1}^n f(X_l)\right) \,\Big|\, X_0 = k\right] = \sum_{l=1}^d \left[\left(P e^{\alpha D_f}\right)^n\right]_{k,l}, \qquad n \geqslant 0.$$

Remark. This extends Question (b) of Problem 7.28.

d) Show that for any $\alpha \geqslant 0$ and $\gamma \geqslant 0$ we have

$$\mathbb{P}\left(\sum_{l=1}^n f(X_l) \geqslant n\gamma \,\Big|\, X_0 = k\right) \leqslant e^{-\alpha\gamma n} \sum_{l=1}^d \left[\left(P e^{\alpha D_f}\right)^n\right]_{k,l}, \qquad n \geqslant 0.$$

---

* Orthogonality is with respect to the scalar product $\langle \cdot , \cdot \rangle$.

Hint. Use the Chernoff argument.

e) Letting $\lambda_0(\alpha)$ denote the largest eigenvalue of $P\mathrm{e}^{\alpha D_f}$, show that for all $\alpha \geqslant 0$ we have

$$\sum_{k,l=1}^{d} \pi_k \big[\big(P\mathrm{e}^{\alpha D_f} \big)^n \big]_{k,l} \leqslant \mathrm{e}^{\alpha} (\lambda_0(\alpha))^n, \qquad n \geqslant 0. \tag{7.3.12}$$

Hints. (i) Write the left hand side of (7.3.12) as a scalar product and use the Cauchy-Schwarz inequality. (ii) Note that

$$P\mathrm{e}^{\alpha D_f} = \mathrm{e}^{-\alpha D_f/2} \mathrm{e}^{\alpha D_f/2} P \mathrm{e}^{\alpha D_f/2} \mathrm{e}^{\alpha D_f/2}$$

is similar to a self-adjoint operator. (iii) Apply e.g. Proposition 9 in Foucart (2010).

f) Show that for any $\alpha \geqslant 0$ and $\gamma \geqslant 0$ we have

$$\mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} f(X_l) \geqslant \gamma \right) \leqslant \mathrm{e}^{\alpha - n(\alpha\gamma - \log \lambda_0(\alpha))}, \qquad n \geqslant 0.$$

g) Show that for any matrix M we have the relation

$$\mathrm{tr}(\Pi P D_f^n M D_f^m) = \mathrm{tr}(\Pi D_f^n M D_f^m) = \langle f^n, M f^m \rangle, \qquad n, m \geqslant 0. \tag{7.3.13}$$

h) Show that $\lambda_0(\alpha)$ can be expanded as the power series

$$\lambda_0(\alpha) = 1 + \sum_{n \geqslant 1} c_n \alpha^n$$

in the parameter α, with $c_1 = 0$ and

$$c_n = \sum_{p=1}^{n} \frac{(-1)^{p-1}}{p} \sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \langle f^{\nu_1}, S^{k'_1} P (D_f)^{\nu_2} \cdots S^{k'_{p-2}} P (D_f)^{\nu_{p-1}} S^{k'_{p-1}} P f^{\nu_p} \rangle,$$

$n \geqslant 2$.

Hints. (i) Apply Relations II-(2.1) and II-(2.31) in Kato (1995) to the expansion

$$P\mathrm{e}^{\alpha D_f} = \sum_{n \geqslant 0} \alpha^n P \frac{(D_f)^n}{n!}$$

using the reduced resolvent $S := (P - I)^{-1}(I - \Pi)$, see II-(2.10)-(2.12) and p. 74 line -1. (ii) Use the fact that at least one of $k_1, \ldots, k_p$ must be zero in II-(2.31) of Kato (1995), and denote the non-zero indexes by $k'_1, \ldots, k'_{p-1}$. (iii) Use (7.3.13). (iv) Use $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

i) Compute $\displaystyle\sum_{\substack{k_1+\cdots+k_p=p-1\\k_1\geqslant 0,\ldots,k_p\geqslant 0}} \mathbf{1}$. *Hint.* We have $\displaystyle\sum_{\substack{\nu_1+\cdots+\nu_p=n\\\nu_1\geqslant 1,\ldots,\nu_p\geqslant 1}} \mathbf{1} = \binom{n-1}{p-1}$.

j) Show that $c_n \leqslant (5/(1-\lambda_1))^{n-1}/5$, $n \geqslant 2$, where λ_1 is the second largest eigenvalue of P.
Hints. (i) Use the inequalities $n! \geqslant 2^{n-1}$ and $4^n \geqslant \binom{2n}{n}\sqrt{\pi n}$, $n \geqslant 1$, Proposition 9 in Foucart (2010), and the Cauchy-Schwarz inequality. (ii) Show that $\|I - \Pi\| \leqslant 1$. (iii) Note that $P - I$ is invertible on $\mathrm{Im}(I - \Pi)$. (iv) Show that

$$\sum_{p=0}^{n-1} \binom{n-1}{p} \frac{x^p}{p+1} = \frac{(1+x)^n - 1}{nx} \leqslant \frac{(1+x)^n}{nx}.$$

k) Show that for all $\gamma \geqslant 0$ and $n \geqslant 0$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^n f(X_l) \geqslant \gamma\right) \leqslant \exp\left(\frac{1-\lambda_1}{5} - n\gamma\alpha + \frac{n\alpha^2}{1-\lambda_1-5\alpha}\right),$$

$\alpha \in [0, (1-\lambda_1)/5)$.
Hint. Use the inequality $\log(1+x) \leqslant x$, $x > 0$.

l) Show that for all $\gamma \geqslant 0$ and $n \geqslant 0$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^n f(X_l) \geqslant \gamma\right) \leqslant \mathrm{e}^{(1-\lambda_1)/5} \exp\left(-(1-\lambda_1)\frac{n\gamma^2}{12}\right).$$

Hint. Minimize the upper bound of Question (k) over $\alpha \in [0, (1-\lambda_1)/5)$.

Problem 7.38 (Wolfer and Kontorovich (2021)) Consider an *irreducible*, *reversible**, Markov chain $(X_n)_{n\geqslant 0}$ admitting a stationary distribution π on the finite state space $\mathbb{S} = \{1, 2, \ldots, d\}$, $d \geqslant 2$, and started in initial distribution π.

Our goal is to estimate the entries in transition matrix $P = (P_{i,j})_{1\leqslant i,j\leqslant d}$ of $(X_n)_{n\geqslant 0}$ using the estimator

$$\widehat{P}_{i,j}(m) := \frac{1}{N_i(m)} \sum_{k=1}^{m-1} \mathbf{1}_{\{X_k=i,X_{k+1}=j\}}, \qquad i,j = 1, \ldots, d,$$

where

$$N_i(m) := \sum_{k=1}^{m-1} \mathbf{1}_{\{X_k=i\}}$$

* *i.e.* $\pi_i P_{i,j} = \pi_j P_{j,i}$, $i,j = 1, \ldots, d$.

denotes the number of returns to state $\textcircled{i}$ until time $m-1$, $i=1,\ldots,d$.

a) For any $i=1,\ldots,d$, we let

$$(Z_i(k))_{k\geqslant 1}=(Z_i(1),Z_i(2),Z_i(3),\ldots)$$

denote a sequence of independent identically distributed random variables with distribution $P_{i,\cdot}$ on $\{1,\ldots,d\}$, *i.e.*

$$\mathbb{P}(Z_i(k)=j)=P_{i,j},\qquad j=1,\ldots,d,\quad k\geqslant 1.$$

Show that for all $i=1,\ldots,d$ we have

$$\mathbb{E}\left[\sum_{j=1}^d\left|\frac{1}{n}\sum_{k=1}^n\mathbf{1}_{\{Z_i(k)=j\}}-P_{i,j}\right|\right]\leqslant\sqrt{\frac{d}{n}},\qquad n\geqslant 1.$$

Hint. Use Jensen's inequality and the variance of the binomial distribution.

b) Show that for any $n\geqslant 1$, the function defined on $\mathbb{R}^n$ by

$$(z(1),\ldots,z(n))\mapsto\sum_{j=1}^d\left|\frac{1}{n}\sum_{k=1}^n\mathbf{1}_{\{z(k)=j\}}-P_{i,j}\right|$$

satisfies the bounded differences property with constant $c_i=2/n$, $i=1,\ldots,n$.

c) Show that for all $i=1,\ldots,d$ we have

$$\mathbb{P}\left(\sum_{j=1}^d\left|\frac{1}{n}\sum_{k=1}^n\mathbf{1}_{\{Z_i(k)=j\}}-P_{i,j}\right|>\varepsilon\right)\leqslant\exp\left(-\frac{n}{2}\operatorname{Max}\left(0,\varepsilon-\sqrt{\frac{d}{n}}\right)^2\right).$$

Hint. Use McDiarmid's inequality.

In what follows, starting from $\widetilde{X}_1$ in the distribution π we let $\widetilde{X}_2:=Z_{\widetilde{X}_1}(1)$, and

$$\widetilde{X}_{k+1}:=Z_{\widetilde{X}_k}\big(1+\widetilde{N}_{\widetilde{X}_k}(k)\big),\qquad k\geqslant 1,$$

where

$$\widetilde{N}_i(k):=\sum_{l=1}^{k-1}\mathbf{1}_{\{\widetilde{X}_l=i\}},\qquad k\geqslant 1.$$

We also let

$$\widetilde{P}_{i,j}(m):=\frac{1}{\widetilde{N}_i(m)}\sum_{k=1}^{m-1}\mathbf{1}_{\{\widetilde{X}_k=i,\widetilde{X}_{k+1}=j\}},\qquad i,j=1,\ldots,d.$$

d) Show that when $\widetilde{N}_i(m)=n\geqslant 1$ we have

$$\widetilde{P}_{i,j}(m) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}}, \qquad i,j = 1,\ldots,d.$$

e) Show that for $i = 1,\ldots,d$, the distribution of $(\widehat{P}_{i,1}(m),\ldots,\widehat{P}_{i,d}(m))$ on $\{N_i(m) = n\}$ is the same as the distribution of $(\widetilde{P}_{i,1}(m),\ldots,\widetilde{P}_{i,d}(m))$ on $\{\widetilde{N}_i(m) = n\}$.

f) Show that letting $n_i := \lceil m\pi_i/2\rceil$, $i = 1,\ldots,d$, for some constant $c_1 > 0$ we have

$$\sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } N_i(m) = n\right) \leqslant (2n_i+1)e^{-c_1 m\pi_i\varepsilon^2},$$

provided that $m \geqslant 4d/(\varepsilon^2\pi_i)$.

g) Show that

$$\sum_{i=1}^{d}\sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } N_i(m) = n\right) \leqslant \frac{2d}{c_1\varepsilon^2}e^{-c_1 m\pi_*\varepsilon^2/2},$$

provided that $m \geqslant 4d/(\varepsilon^2\pi_*)$ and $\varepsilon \in (0,1)$, where $\pi_* := \min_{1\leqslant j\leqslant d}\pi_j$.
Hint. Use the inequality $xe^{-x} \leqslant e^{-x/2}$, $x > 0$.

h) Show that for all $\varepsilon > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d}\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon\right) \\
&\quad \leqslant \sum_{i=1}^{d}\sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } N_i(m) = n\right) \\
&\quad + \mathbb{P}(\exists i \in \{1,\ldots,d\} \ : \ N_i(m) \notin [n_i, 3n_i]).
\end{aligned}$$

i) Using the bound in Question (l) of Problem 7.37, show that there exist two constants $c_2, c_3 > 0$ such that

$$\mathbb{P}(\exists i \in \{1,\ldots,d\} \ : \ N_i(m) \notin [n_i, 3n_i]) \leqslant c_2 d e^{-c_3(1-\lambda_1)m\pi_*^2}, \qquad m > 4/\pi_*.$$

j) Show that there is a constant $c > 0$ such that for any $\varepsilon, \delta \in (0,1)$, if

$$m \geqslant c\operatorname{Max}\left(\frac{1}{\varepsilon^2\pi_*}\operatorname{Max}\left(d, \log\frac{d}{\delta\varepsilon}\right), \frac{1}{(1-\lambda_1)\pi_*^2}\log\frac{d}{\delta}\right),$$

then we have

$$\mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d}\sum_{j=1}^{d}\left|\widehat{P}_{i,j}(m)-P_{i,j}\right|\leqslant\varepsilon\right)\geqslant 1-\delta.$$

Chapter 8
Branching Processes

Branching processes are used as a tool for modeling in genetics, biomolecular reproduction, population growth, genealogy, disease spread, photomultiplier cascades, nuclear fission, earthquake triggering, queueing models, viral phenomena, social networks, neuroscience, etc. This chapter mainly deals with the computation of probabilities of extinction and explosion in finite time for branching processes.

8.1 Construction and Examples

Consider a time-dependent population made of a number X_n of individuals at generation $n \geqslant 0$. In the branching process model, each of these X_n individuals may have a random number of descendants born at time $n+1$.

For each $k = 1, 2, \ldots, X_n$ we let Y_k denote the number of descendants of individual n^o k. That means, we have $X_0 = 1$, $X_1 = Y_1$, and at time $n+1$, the new population size X_{n+1} will be given by

$$X_{n+1} = Y_1 + \cdots + Y_{X_n} = \sum_{k=1}^{X_n} Y_k, \tag{8.1.1}$$

where the $(Y_k)_{k \geqslant 1}$ form a sequence of independent, identically distributed, nonnegative integer valued random variables which are assumed to be almost surely finite, *i.e.*

$$\mathbb{P}(Y_k < \infty) = \sum_{n \geqslant 0} \mathbb{P}(Y_k = n) = 1.$$

Note that new independent samples of $(Y_k)_{k\geqslant 1}$ are generated at each generation, and the X_n individuals of generation n "die" at time $n+1$ as they are not considered in the sum (8.1.1). In order to keep them at the next generation we would have to modify (8.1.1) into the *progeny*

$$X_{n+1} = X_n + Y_1 + \cdots + Y_{X_n},$$

however we will *not* adopt this convention, and we will rely on (8.1.1) instead.

As a consequence of (8.1.1), the branching process $(X_n)_{n\geqslant 0}$ is a Markov process with state space $\mathbb{S} = \mathbb{N}$ and transition matrix given by

$$P = \left[P_{i,j} \right]_{i,j\geqslant 0} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ \mathbb{P}(Y_1 = 0) & \mathbb{P}(Y_1 = 1) & \mathbb{P}(Y_1 = 2) & \cdots \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ P_{3,0} & P_{3,1} & P_{3,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \qquad (8.1.2)$$

Note that state ⓪ is absorbing since by construction we always have

$$P_{0,0} = \mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = 1, \qquad n \geqslant 0.$$

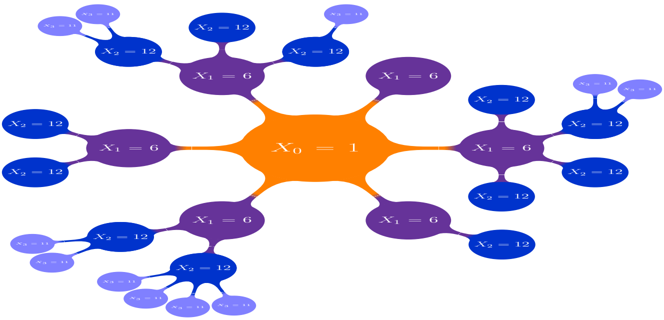


Fig. 8.1: Sample of a branching process.

Figure 8.1 represents an example of branching process with $X_0 = 1$ and $Y_1 = 6$, hence

$$X_1 = Y_{X_0} = Y_1 = 6,$$

then successively

$$(Y_k)_{k=1,2,\ldots,X_1} = (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6) = (0, 4, 1, 2, 2, 3)$$

and

$$\begin{aligned} X_2 &= Y_1 + \cdots + Y_{X_1} \\ &= Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 \\ &= 0 + 4 + 1 + 2 + 2 + 3 \\ &= 12, \end{aligned}$$

then

$$\begin{aligned} (Y_k)_{k=1,2,\ldots,X_2} &= (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9, Y_{10}, Y_{11}, Y_{12}) \\ &= (0, 2, 0, 0, 0, 4, 2, 0, 0, 2, 0, 1), \end{aligned}$$

and

$$\begin{aligned} X_3 &= Y_1 + \cdots + Y_{X_2} \\ &= Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + Y_9 + Y_{10} + Y_{11} + Y_{12} \\ &= 0 + 2 + 0 + 0 + 0 + 4 + 2 + 0 + 0 + 2 + 0 + 1 \\ &= 11. \end{aligned}$$

Figure 8.2 presents another sample tree for the path of a branching process.

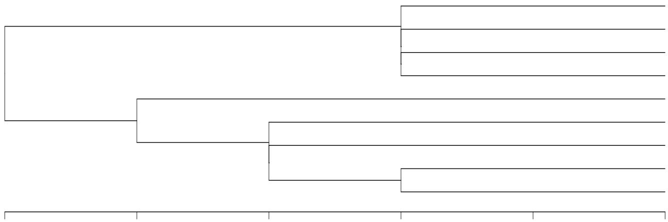


Fig. 8.2: Sample graph of a branching process.

In Figure 8.2 the branching process starts from $X_0 = 2$, with $X_1 = 3$, $X_2 = 5$, $X_3 = 9$, $X_4 = 9$, $X_5 = 9$. However, in the sequel and except if otherwise specified, all branching processes will start from $X_0 = 1$.

See Sellke et al. (2008) and Iribarren and Moro (2011) for results on the modeling of the offspring distribution of Y_1 based on social network and internet data and the use of power tail distributions. The use of power tail distributions leads to probability generating functions of polylogarithmic form.

8.2 Probability Generating Functions

Let now $G_1(s)$ denote the probability generating function of $X_1 = Y_1$, defined as

$$G_1(s) := \mathbb{E}\big[s^{X_1} \,\big|\, X_0 = 1\big] = \mathbb{E}\big[s^{Y_1}\big] = \sum_{k\geqslant 0} s^k \mathbb{P}(Y_1 = k), \quad -1 \leqslant s \leqslant 1,$$

cf. (1.7.3), denote the probability generating function of the (almost surely finite) random variable $X_1 = Y_1$, with

$$\begin{cases} G_1(0) = \mathbb{P}(Y_1 = 0), \\ \\ G_1(1) = \displaystyle\sum_{n\geqslant 0} \mathbb{P}(Y_1 = n) = \mathbb{P}(Y_1 < \infty) = 1, \\ \\ \mu := G_1'(1^-) = \displaystyle\sum_{k\geqslant 0} k\mathbb{P}(Y_1 = k) = \mathbb{E}[X_1 \mid X_0 = 1] = \mathbb{E}[Y_1]. \end{cases} \tag{8.2.1a}$$

More generally, letting $G_n(s)$ denote the probability generating function of X_n, defined as

$$G_n(s) := \mathbb{E}\big[s^{X_n} \mid X_0 = 1\big] = \sum_{k\geqslant 0} s^k \mathbb{P}(X_n = k \mid X_0 = 1), \quad -1 \leqslant s \leqslant 1,$$

$n \in \mathbb{N}$, we have

$$\begin{cases} G_0(s) = s, \qquad -1 \leqslant s \leqslant 1, \\ \\ G_n(0) = \mathbb{P}(X_n = 0 \mid X_0 = 1), \qquad n \geqslant 0, & (8.2.2a) \\ \\ \mu_n := \mathbb{E}[X_n \mid X_0 = 1] = G_n'(1^-) = \displaystyle\sum_{k\geqslant 0} k\mathbb{P}(X_n = k \mid X_0 = 1), & (8.2.2b) \end{cases}$$

cf. (1.7.5). When $X_0 = k$ we can view the branching tree as the union of k independent trees started from $X_0 = 1$ and we can write X_n as the sum of independent random variables

$$X_n = \sum_{l=1}^{k} X_n^{(l)}, \qquad n \geqslant 0,$$

where $X_n^{(l)}$ denotes the size of the tree $n^o l$ at time n, with $X_n^{(l)} = 1$, $l = 1, 2, \ldots, k$. In this case, we have

$$\begin{aligned}
\mathbb{E}\big[s^{X_n} \mid X_0 = k\big] &= \mathbb{E}\Big[s^{X_n^{(1)}+\cdots+X_n^{(k)}} \;\Big|\; X_0^{(1)} = 1, \ldots, X_0^{(k)} = 1\Big] \\
&= \prod_{l=1}^{k} \mathbb{E}\big[s^{X_n^{(l)}} \mid X_0^{(l)} = 1\big] \\
&= \big(\mathbb{E}\big[s^{X_n} \mid X_0 = 1\big]\big)^k \\
&= (G_n(s))^k, \quad -1 \leqslant s \leqslant 1, \quad n \in \mathbb{N}.
\end{aligned}$$

The next proposition provides an algorithm for the computation of the probability generating function G_n.

Proposition 8.1. *We have the recurrence relation*

$$G_{n+1}(s) = G_n(G_1(s)) = G_1(G_n(s)), \quad -1 \leqslant s \leqslant 1, \qquad n \geqslant 0. \qquad (8.2.3)$$

Proof. By the identity (1.6.11) on random products we have

$$\begin{aligned}
G_{n+1}(s) &= \mathbb{E}\big[s^{X_{n+1}} \mid X_0 = 1\big] \\
&= \mathbb{E}\big[s^{Y_1+\cdots+Y_{X_n}} \mid X_0 = 1\big] \\
&= \mathbb{E}\left[\prod_{l=1}^{X_n} s^{Y_l} \;\middle|\; X_0 = 1\right] \\
&= \sum_{k\geqslant 0} \mathbb{E}\left[\prod_{l=1}^{X_n} s^{Y_l} \;\middle|\; X_n = k\right] \mathbb{P}(X_n = k \mid X_0 = 1) \\
&= \sum_{k\geqslant 0} \mathbb{E}\left[\prod_{l=1}^{k} s^{Y_l} \;\middle|\; X_n = k\right] \mathbb{P}(X_n = k \mid X_0 = 1) \\
&= \sum_{k\geqslant 0} \mathbb{E}\left[\prod_{l=1}^{k} s^{Y_l}\right] \mathbb{P}(X_n = k \mid X_0 = 1) \\
&= \sum_{k\geqslant 0} \left(\prod_{l=1}^{k} \mathbb{E}\left[s^{Y_l}\right]\right) \mathbb{P}(X_n = k \mid X_0 = 1) \\
&= \sum_{k\geqslant 0} \big(\mathbb{E}\big[s^{Y_1}\big]\big)^k \mathbb{P}(X_n = k \mid X_0 = 1) \\
&= G_n\big(\mathbb{E}\big[s^{Y_1}\big]\big) \\
&= G_n(G_1(s)), \qquad -1 \leqslant s \leqslant 1.
\end{aligned}$$

□

Instead of (8.2.3) we may also write

$$G_n(s) = G_1(G_1(\cdots(G_1(s),\cdots)), \qquad -1 \leqslant s \leqslant 1, \tag{8.2.4}$$

and

$$G_n(s) = G_1(G_{n-1}(s)) = G_{n-1}(G_1(s)), \qquad -1 \leqslant s \leqslant 1.$$

Mean population size

In case the random variable Y_k is equal to a deterministic constant $\mu \in \mathbb{N}$, the population size at generation $n \geqslant 0$ will clearly be equal to μ^n. The next proposition shows that for branching processes, this property admits a natural extension to the random case.

Proposition 8.2. *The mean population size μ_n at generation $n \geqslant 0$ is given by*

$$\mu_n = \mathbb{E}[X_n \mid X_0 = 1] = (\mathbb{E}[X_1 \mid X_0 = 1])^n = \mu^n, \quad n \geqslant 0, \tag{8.2.5}$$

where $\mu = \mathbb{E}[Y_1]$ is given by (8.2.1a).

Proof. By (8.2.4), (8.2.2b) and the chain rule of derivation, we have

$$\begin{aligned} \mu_n &= G_n'(1^-) \\ &= \frac{d}{ds} G_1(G_{n-1}(s))_{|s=1} \\ &= G_{n-1}'(1^-)G_1'(G_{n-1}(1^-)) \\ &= G_{n-1}'(1^-)G_1'(1^-) \\ &= \mu \times \mu_{n-1}, \end{aligned}$$

hence $\mu_1 = \mu$, $\mu_2 = \mu \times \mu_1 = \mu^2$, $\mu_3 = \mu \times \mu_2 = \mu^3$, and by induction on $n \geqslant 1$ we obtain (8.2.5). □

Similarly, we find

$$\mathbb{E}[X_n \mid X_0 = k] = k\mathbb{E}[X_n \mid X_0 = 1] = k\mu^n, \quad n \geqslant 1,$$

hence starting from $X_0 = k \geqslant 1$, the average of X_n goes to infinity when $\mu > 1$. On the other hand, μ_n converges to 0 when $\mu < 1$, and in this case the *total* mean population size, including all ancestors, is finite and equal to

$$\sum_{n \geqslant 0} \mathbb{E}[X_n \mid X_0 = 1] = \sum_{n \geqslant 0} \mu^n = \frac{1}{1-\mu}.$$

Examples

i) Supercritical case. When $\mu > 1$ the average population size $\mu_n = \mu^n$ grows to infinity as n tends to infinity, and we say that the branching process $(X_n)_{n\geqslant 0}$ is *supercritical.*

This condition holds in particular when $\mathbb{P}(Y_1 \geqslant 1) = 1$ and Y_1 is not almost surely equal to 1, *i.e.* $\mathbb{P}(Y_1 = 1) < 1$. Indeed, under those conditions we have $\mathbb{P}(Y_1 \geqslant 2) > 0$ and

$$\begin{aligned}\mu = \mathbb{E}[Y_1] &= \sum_{n\geqslant 1} n\mathbb{P}(Y_1 = n)\\ &\geqslant \mathbb{P}(Y_1 = 1) + 2\sum_{n\geqslant 2} \mathbb{P}(Y_1 = n)\\ &= \mathbb{P}(Y_1 = 1) + 2\mathbb{P}(Y_1 \geqslant 2)\\ &> \mathbb{P}(Y_1 \geqslant 1)\\ &= 1,\end{aligned}$$

hence $\mu > 1$.

ii) Critical case. When $\mu = 1$ we have $\mu_n = (\mu)^n = 1$ for all $n \in \mathbb{N}$, and we say that the branching process $(X_n)_{n\geqslant 0}$ is *critical* .

iii) Subcritical case. In case $\mu < 1$, the average population size $\mu_n = \mu^n$ tends to 0 as n tends to infinity and we say that the branching process $(X_n)_{n\geqslant 0}$ is *subcritical.* In this case we necessarily have $\mathbb{P}(Y_1 = 0) > 0$ as

$$\begin{aligned}1 > \mu = \mathbb{E}[Y_1] &= \sum_{n\geqslant 1} n\mathbb{P}(Y_1 = n)\\ &\geqslant \sum_{n\geqslant 1} \mathbb{P}(Y_1 = n)\\ &= \mathbb{P}(Y_1 \geqslant 1)\\ &= 1 - \mathbb{P}(Y_0 = 0),\end{aligned}$$

although the converse is not true in general.

The variance $\sigma_n^2 = \mathrm{Var}[X_n \mid X_0 = 1]$ of X_n given that $X_0 = 1$ can be shown in a similar way to satisfy the recurrence relation

$$\sigma_{n+1}^2 = \sigma^2\mu^n + \mu\sigma_n^2,$$

cf. also Relation (1.7.8), where $\sigma^2 = \mathrm{Var}[Y_1]$, which shows by induction that

$$\sigma_n^2 = \mathrm{Var}[X_n \mid X_0 = 1] = \begin{cases} n\sigma^2, & \mu = 1,\\ \\ \sigma^2\mu^{n-1}\dfrac{1-\mu^n}{1-\mu} = \sigma^2\displaystyle\sum_{k=0}^{n-1}\mu^{n+k-1}, & \mu \neq 1,\end{cases}$$

$n \geqslant 1$ cf. *e.g.* pages 180-181 of Karlin and Taylor (1998), and Exercise 8.3-(a) for an application. We also have

$$\mathrm{Var}[X_n \mid X_0 = k] = k\mathrm{Var}[X_n \mid X_0 = 1] = k\sigma_n^2, \qquad k, n \geqslant 0,$$

due to Relation (1.6.10) for the variance of a sum of independent random variables.

8.3 Extinction Probabilities

Here we are interested in the time to extinction*

$$T_0 := \inf\{n \geqslant 0 \ : \ X_n = 0\},$$

and in the extinction probability

$$\alpha_k := \mathbb{P}(T_0 < \infty \mid X_0 = k)$$

within a finite time, after starting from $X_0 = k$. Note that the word "extinction" can have negative as well as positive meaning, for example when the branching process is used to model the spread of an infection.

Proposition 8.3. *The probability distribution of T_0 can be expressed using the probability generating function G_n as*

$$\mathbb{P}(T_0 = n \mid X_0 = 1) = G_n(0) - G_{n-1}(0) = G_1(G_{n-1}(0)) - G_{n-1}(0), \quad n \geqslant 1,$$

with $\mathbb{P}(T_0 = 0 \mid X_0 = 1) = 0$.

Proof. By the relation $\{X_{n-1} = 0\} \subset \{X_n = 0\}$, we have

$$\{T_0 = n\} = \{X_n = 0\} \cap \{X_{n-1} \geqslant 1\} = \{X_n = 0\} \setminus \{X_{n-1} = 0\}$$

and

$$\begin{aligned}\mathbb{P}(T_0 = n \mid X_0 = 1) &= \mathbb{P}(\{X_n = 0\} \cap \{X_{n-1} \geqslant 1\} \mid X_0 = 1)\\ &= \mathbb{P}(\{X_n = 0\} \setminus \{X_{n-1} = 0\} \mid X_0 = 1)\\ &= \mathbb{P}(X_n = 0 \mid X_0 = 1)) - \mathbb{P}(X_{n-1} = 0 \mid X_0 = 1)\\ &= G_n(0) - G_{n-1}(0)\\ &= G_1(G_{n-1}(0)) - G_{n-1}(0), \quad n \geqslant 1,\end{aligned}$$

where we applied Proposition 8.1. □

* We normally start from $X_0 \geqslant 1$.

First, we note that by the independence assumption, starting from $X_0 = k \geqslant 2$ independent individuals, we have

$$\alpha_k = \mathbb{P}(T_0 < \infty \mid X_0 = k) = (\mathbb{P}(T_0 < \infty \mid X_0 = 1))^k = (\alpha_1)^k, \quad k \geqslant 1. \tag{8.3.1}$$

Indeed, given k individuals at generation 0, each of them will start independently a new branch of offsprings, and in order to have extinction of the whole population, all of k branches should become extinct. Since the k branches behave independently, α_k is the product of the extinction probabilities for each branch, which yields $\alpha_k = (\alpha_1)^k$ since these extinction probabilities are all equal to α_1 and there are k of them.

The next proposition is a consequence of Lemma 5.2.

Proposition 8.4. *The extinction probability within a finite time after starting from $X_0 = k$ satisfies*

$$\alpha_1 = \mathbb{P}(T_0 < \infty \mid X_0 = 1) = \lim_{n\to\infty} G_n(0).$$

Proof. Since state ⓪ is absorbing, by Lemma 5.2 with $j = 0$ and $i = 1$ we find

$$\alpha_1 = \mathbb{P}(T_0 < \infty \mid X_0 = 1) = \lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 1) = \lim_{n\to\infty} G_n(0).$$

□

The next proposition shows that the extinction probability α_1 can be computed as the solution of an equation.

Proposition 8.5. *The extinction probability*

$$\alpha_1 := \mathbb{P}(T_0 < \infty \mid X_0 = 1)$$

is a solution of the equation

$$\alpha = G_1(\alpha). \tag{8.3.2}$$

Proof. By first step analysis, we have

$$\begin{aligned}
\alpha_1 &= \mathbb{P}(T_0 < \infty \mid X_0 = 1) \\
&= \mathbb{P}(X_1 = 0 \mid X_0 = 1) + \sum_{k\geqslant 1} \mathbb{P}(T_0 < \infty \mid X_1 = k)\mathbb{P}(X_1 = k \mid X_0 = 1) \\
&= \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \mathbb{P}(T_0 < \infty \mid X_0 = k)\mathbb{P}(Y_1 = k) \\
&= \sum_{k\geqslant 0} (\alpha_1)^k \mathbb{P}(Y_1 = k) \\
&= G_1(\alpha_1),
\end{aligned}$$

hence the extinction probability α_1 solves (8.3.2). □

Note that from the above proof we find

$$\alpha_1 \geqslant \mathbb{P}(X_1 = 0 \mid X_0 = 1) = \mathbb{P}(Y_1 = 0), \tag{8.3.3}$$

which shows that the extinction probability is non-zero whenever $\mathbb{P}(Y_1 = 0) > 0$. On the other hand, any solution α of (8.3.2) also satisfies

$$\alpha = G_1(G_1(\alpha)), \quad \alpha = G_1(G_1(G_1(\alpha))),$$

and more generally

$$\alpha = G_n(\alpha), \qquad n \geqslant 1, \tag{8.3.4}$$

by Proposition 8.1. On the other hand the solution of (8.3.2) may not be unique, for example $\alpha = 1$ is always solution of (8.3.2) since $G_1(1) = 1$, and it may not be equal to the extinction probability. The next proposition clarifies this point.

Proposition 8.6. *The extinction probability*

$$\alpha_1 := \mathbb{P}(T_0 < \infty \mid X_0 = 1)$$

is the smallest nonnegative solution of the equation $\alpha = G_1(\alpha)$*.*

Proof. By Lemma 8.4 we have $\alpha_1 = \lim_{n\to\infty} G_n(0)$. Next, we note that the function $s \longmapsto G_1(s)$ is increasing because

$$G_1'(s) = \mathbb{E}[Y_1 s^{Y_1 - 1}] > 0, \qquad s \in [0, 1).$$

Hence $s \longmapsto G_n(s)$ is also increasing by Proposition 8.1, and for any solution $\alpha \geqslant 0$ of (8.3.2) we have, by (8.3.4),

$$0 \leqslant G_n(0) \leqslant G_n(\alpha) = \alpha, \qquad n \geqslant 1,$$

and taking the limit in this inequality as n goes to infinity we get

$$0 \leqslant \alpha_1 = \lim_{n\to\infty} G_n(0) \leqslant \alpha,$$

by Proposition 8.4, hence the extinction probability α_1 is always smaller than any solution α of (8.3.2). This fact can also be recovered from Proposition 8.4 and

$$\begin{aligned}\alpha &= \lim_{n\to\infty} G_n(\alpha) \\ &= \lim_{n\to\infty} \left(G_n(0) + \sum_{k\geqslant 1} \alpha^k \mathbb{P}(X_n = k \mid X_0 = 1) \right)\end{aligned}$$

$$\geqslant \lim_{n\to\infty} G_n(0)$$
$$= \alpha_1.$$

Therefore α_1 is the smallest solution of (8.3.2). □

Since $G_1(0) = \mathbb{P}(Y_1 = 0)$ we have

$$\mathbb{P}(Y_1 = 0) = G_1(0) \leqslant G_1(\alpha_1) = \alpha_1,$$

which recovers (8.3.3).

On the other hand, if $\mathbb{P}(Y_1 \geqslant 1) = 1$ then we have $G_1(0) = 0$, which implies $\alpha_1 = 0$ by Proposition 8.6.

Note that from Lemma 5.2, Proposition 8.4, and (8.3.1), the transition matrix (8.1.2) satisfies

$$\lim_{n\to\infty} \left([P^n]_{i,0}\right)_{i\geqslant 0} = \begin{bmatrix} 1 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha_1 \\ (\alpha_1)^2 \\ (\alpha_1)^3 \\ \vdots \end{bmatrix}.$$

Examples

i) Assume that Y_1 has a Bernoulli distribution with parameter $p \in (0,1)$, *i.e.*

$$\mathbb{P}(Y_1 = 1) = p, \qquad \mathbb{P}(Y_1 = 0) = 1 - p.$$

Compute the extinction probability of the associated branching process.

In this case the branching process is actually a two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 & 0 \\ 1-p & p \end{bmatrix},$$

and we have

$$G_n(0) = \mathbb{P}(X_n = 0 \mid X_0 = 1) = (1-p)\sum_{k=0}^{n-1} p^k = 1 - p^n, \quad (8.3.5)$$

where we used the geometric series (A.2), hence as in (5.1.6) the extinction probability α_1 is given by

$$\alpha_1 = \mathbb{P}(T_0 < \infty \mid X_0 = 1) = \mathbb{P}\left(\bigcup_{n\geqslant 1}\{X_n = 0\} \;\middle|\; X_0 = 1\right)$$

$$= \lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 1)$$
$$= \lim_{n\to\infty} G_n(0) = 1,$$

provided that $p = \mathbb{E}[Y_1] < 1$, otherwise we have $\alpha_1 = 0$ when $p = 1$. The value of α_1 can be recovered using the generating function

$$G_1(s) = \mathbb{E}\big[s^{Y_1}\big] = \sum_{k\geqslant 0} s^k \mathbb{P}(Y_1 = k) = 1 - p + ps, \qquad (8.3.6)$$

for which the unique solution of $G_1(\alpha) = \alpha$ is the extinction probability $\alpha_1 = 1$, as shown in the next Figure 8.3.

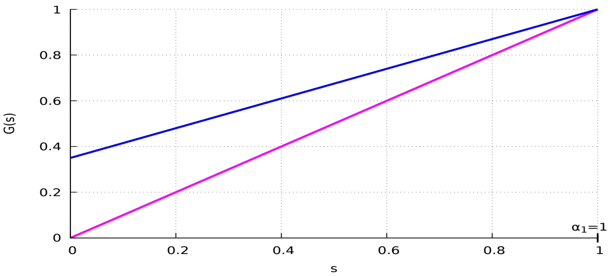


Fig. 8.3: Generating function of Y_1 with $p = 0.65$.

From (8.2.4) and (8.3.6) we can also show by induction on $n \geqslant 1$ as in Exercise 8.2 that

$$G_n(s) = p^n s + (1-p)\sum_{k=0}^{n-1} p^k = 1 - p^n + p^n s,$$

which recovers (8.3.5) from (1.7.5) or (8.2.2a) as

$$\mathbb{P}(X_n = 0 \mid X_0 = 1) = G_n(0) = 1 - p^n.$$

We also have $\mathbb{E}[X_n] = p^n$, $n \geqslant 0$.

ii) Same question as in (i) above for

$$\mathbb{P}(Y_1 = 2) = p, \qquad \mathbb{P}(Y_1 = 0) = q = 1 - p.$$

Here, we will directly use the probability generating function

$$G_1(s) = \mathbb{E}\left[s^{Y_1}\right] = \sum_{k\geqslant 0} s^k \mathbb{P}(Y_1 = k)$$
$$= s^0\mathbb{P}(Y_1 = 0) + s^2\mathbb{P}(Y_1 = 2) = 1 - p + ps^2.$$

We check that the solutions of

$$G_1(\alpha) = 1 - p + p\alpha^2 = \alpha,$$

i.e. *

$$p\alpha^2 - \alpha + q = p(\alpha - 1)(\alpha - q/p) = 0, \qquad (8.3.7)$$

with $q = 1 - p$, are given by

$$\left\{\frac{1+\sqrt{1-4pq}}{2p}, \frac{1-\sqrt{1-4pq}}{2p}\right\} = \left\{1, \frac{q}{p}\right\}, \qquad p \in (0, 1]. \qquad (8.3.8)$$

Hence the extinction probability is $\alpha_1 = 1$ if $q \geqslant p$, and it is equal to $\alpha_1 = q/p < 1$ if $q < p$, or equivalently if $\mathbb{E}[Y_1] > 1$, due to the relation $\mathbb{E}[Y_1] = 2p$.

iii) Assume that Y_1 has the geometric distribution with parameter $p \in (0, 1)$, *i.e.*

$$\mathbb{P}(Y_1 = n) = (1 - p)p^n, \qquad n \geqslant 0,$$

with $\mu = \mathbb{E}[Y_1] = p/q$. We have

$$G_1(s) = \mathbb{E}\left[s^{Y_1}\right] = \sum_{n\geqslant 0} s^n \mathbb{P}(Y_1 = n) = (1-p)\sum_{n\geqslant 0} p^n s^n = \frac{1-p}{1-ps}. \qquad (8.3.9)$$

The equation $G_1(\alpha) = \alpha$ reads

$$\frac{1-p}{1-p\alpha} = \alpha,$$

i.e.

$$p\alpha^2 - \alpha + q = p(\alpha - 1)(\alpha - q/p) = 0,$$

which is identical to (2.2.16) and (8.3.7) with $q = 1 - p$, and has for solutions (8.3.8). Hence the finite time extinction probability is

$$\alpha_1 = \mathbb{P}(T_0 < \infty \mid X_0 = 1)$$
$$= \min\left(1, \frac{q}{p}\right) = \begin{cases} \dfrac{q}{p}, & p \geqslant 1/2, \text{ (super)critical case,} \\ 1, & p \leqslant 1/2, \text{ (sub)critical case.} \end{cases}$$

* Remark that (8.3.7) is identical to the characteristic equation (2.2.16).

Note that we have $\alpha_1 < 1$ if and only if $\mathbb{E}[Y_1] > 1$, due to the equality $\mathbb{E}[Y_1] = p/q$. As can be seen from Figures 8.4 and 8.5, the extinction probability α_1 is equal to 1 when $p \leqslant 1/2$, meaning that extinction within a finite time is certain in that case. Note that we also find

$$\mathbb{P}(T_0 < \infty \mid X_0 = k) = \min\left(1, \left(\frac{q}{p}\right)^k\right)$$
$$= \begin{cases} \left(\dfrac{q}{p}\right)^k, & p \geqslant 1/2, \ \text{(super)critical case,} \\ \\ 1, & p \leqslant 1/2, \ \text{(sub)critical case,} \end{cases}$$

which incidentally coincides with the finite time hitting probability found in (3.4.16) for the simple random walk started from $k \geqslant 1$.

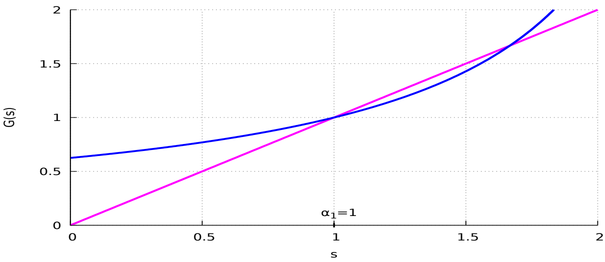


Fig. 8.4: Generating function of Y_1 with $p = 3/8 < 1/2$ and $\alpha_1 = 1$.

Next, in Figure 8.5 is a graph of the generating function $s \longmapsto G_1(s)$ for $p = 1/2$.

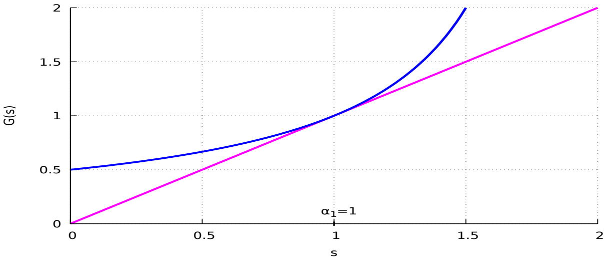


Fig. 8.5: Generating function of Y_1 with $p = 1/2$ and $\alpha_1 = 1$.

The generating function graph in Figure 8.6 corresponds to $p = 3/4$.

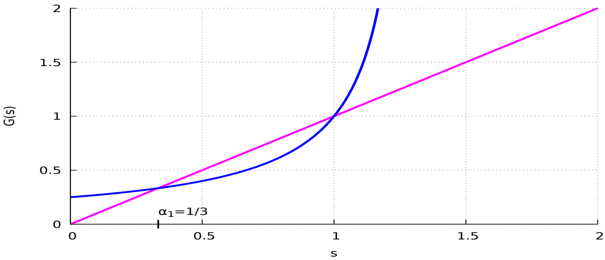


Fig. 8.6: Generating function of Y_1 with $p = 3/4 > 1/2$ and $\alpha_1 = q/p = 1/3$.

We also have $\mu_n = (\mathbb{E}[Y_1])^n = (p/q)^n$, $n \geqslant 1$.

iv) Assume now that Y_1 is the sum of two independent geometric variables with parameter p, *i.e.* it has the negative binomial distribution

$$\mathbb{P}(Y_1 = n) = \binom{n+r-1}{r-1} q^r p^n = (n+1) q^r p^n = (n+1) q^2 p^n, \qquad n \geqslant 0,$$

with $\underline{r=2}$, cf. (1.5.12).

In this case we have*

$$G_1(s) = \mathbb{E}\big[s^{Y_1}\big] = \sum_{n \geqslant 0} s^n \mathbb{P}(Y_1 = n)$$

* Here, Y_1 is the sum of two independent geometrically distributed random variables, and G_1 is the square of the generating function (8.3.9) of the geometric distribution.

$$= q^2 \sum_{n \geqslant 0} (n+1) p^n s^n = \left(\frac{1-p}{1-ps} \right)^2, \qquad -1 \leqslant s \leqslant 1,$$

see *here*.* When $p = 1/2$ we check that $G_1(\alpha) = \alpha$ reads

$$s^3 - 4s^2 + 4s - 1 = 0,$$

which is an equation of degree 3 in the unknown s. Now, since $\alpha = 1$ is solution of this equation we can factorise it as follows:

$$(s-1)(s^2 - 3s + 1) = 0,$$

and we check that the smallest nonnegative solution of this equation is given by

$$\alpha_1 = \frac{1}{2}(3 - \sqrt{5}) \simeq 0.382$$

which is the extinction probability, as illustrated in the next Figure 8.7. Here we have $\mathbb{E}[Y_1] = 2$.

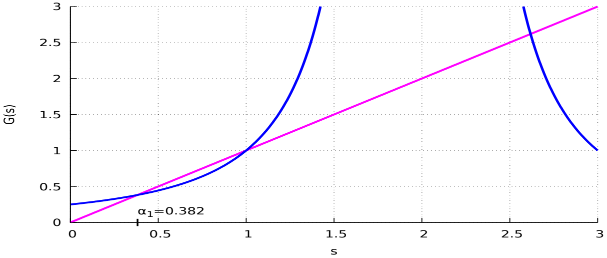


Fig. 8.7: Probability generating function of Y_1.

* We used the identity $\sum_{n \geqslant 0} (n+1) r^n = (1-r)^{-2}$, cf. (A.4).

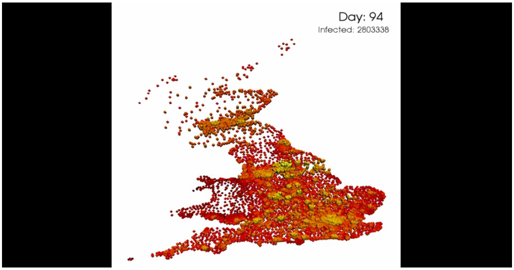


Fig. 8.8: Simulation of influenza spread.*

The next graph in Figure 8.9 illustrates the extinction of a branching process in finite time when Y_1 has the geometric distribution with $p = 1/2$, in which case there is extinction within finite time with probability 1.

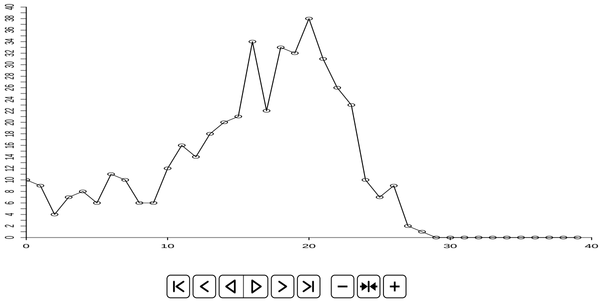


Fig. 8.9: Sample path of a branching process $(X_n)_{n\geqslant 0}$.†

* Click on the figure to play the video (works in Acrobat Reader).

† Animated figure (works in Acrobat Reader). Download the corresponding R code or the **IPython notebook** that can be run ***here***.

In Table 8.1 we summarize some questions and their associated solution methods introduced in this chapter and the previous ones.

Table 8.1: Summary of computing methods.

How to compute	Method
Expected value $\mathbb{E}[X]$	Sum the values of X weighted by their probabilities.
Uses of $G_X(s)$	$G_X(0) = \mathbb{P}(X = 0)$, $G_X(1) = \mathbb{P}(X < \infty)$, $G'_X(1^-) = \mathbb{E}[X]$.
Hitting probabilities $g(k)$	Solve* $g = Pg$ for $g(k)$.
Mean hitting times $h(k)$	Solve* $h = \mathbb{1} + Ph$ for $h(k)$.
Stationary distribution π	Solve† $\pi = \pi P$ for π.
Extinction probability α_1	Solve $G_1(\alpha) = \alpha$ for α, and choose the smallest solution.
$\lim_{n\to\infty} \left(\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \right)^n$	$\begin{bmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{bmatrix}$

Exercises

Exercise 8.1 A parent particle can be divided into 0, 1 or 2 particles with probabilities 1/5, 3/5, and 1/5, respectively. It disappears after splitting.

* Be sure to write only the relevant rows of the system under the appropriate boundary conditions.

† Remember that the values of $\pi(k)$ have to add up to 1.

Starting with one particle, the ancestor, let us denote by X_n the size of the corresponding branching process at the nth generation.

a) Find $P(X_2 > 0)$.
b) Find $P(X_2 = 1)$.
c) Find the probability that $X_1 = 2$ given that $X_2 = 1$.

Exercise 8.2 Each individual in a population has a random number Y of offsprings, with

$$\mathbb{P}(Y = 0) = 1/2, \qquad \mathbb{P}(Y = 1) = 1/2.$$

Let X_n denote the size of the population at time $n \geqslant 0$, with $X_0 = 1$.

a) Compute the generating function $G_1(s) = \mathbb{E}[s^Y]$ of Y for $s \in \mathbb{R}_+$.
b) Let $G_n(s) := \mathbb{E}[s^{X_n}]$ denote the generating function of X_n. Show that

$$G_n(s) = 1 - \frac{1}{2^n} + \frac{s}{2^n}, \qquad s \in \mathbb{R}. \tag{8.3.10}$$

c) Compute the probability $\mathbb{P}(X_n = 0 \mid X_0 = 1)$ that the population is extinct at time n.
d) Compute the average size $\mathbb{E}[X_n \mid X_0 = 1]$ of the population at step n.
e) Compute the extinction probability of the population starting from one individual at time 0.

Exercise 8.3 Each individual in a population has a random number ξ of offsprings, with distribution

$$\mathbb{P}(\xi = 0) = 0.2, \qquad \mathbb{P}(\xi = 1) = 0.5, \qquad \mathbb{P}(\xi = 2) = 0.3.$$

Let X_n denote the number of individuals in the population at the nth generation, with $X_0 = 1$.

a) Compute the mean and variance of X_2.
b) Give the probability distribution of the random variable X_2.
c) Compute the probability that the population is extinct by the fourth generation.
d) Compute the expected number of offsprings at the tenth generation.
e) What is the probability of extinction of this population?

Exercise 8.4 Each individual in a population has a random number Y of offsprings, with

$$\mathbb{P}(Y = 0) = c, \qquad \mathbb{P}(Y = 1) = b, \qquad \mathbb{P}(Y = 2) = a,$$

where $a + b + c = 1$.

a) Compute the generating function $G_1(s)$ of Y for $s \in [-1, 1]$.
b) Compute the probability that the population is extinct at time 2, starting from 1 individual at time 0.
c) Compute the probability that the population is extinct at time 2, starting from 2 individuals at time 0.
d) Show that when $0 < c \leqslant a$ the probability of eventual extinction of the population, starting from 2 individuals at time 0, is $(c/a)^2$.
e) What is this probability equal to when $0 < a < c$?

Exercise 8.5 Consider a branching process $(Z_n)_{n \geqslant 0}$ in which the offspring distribution at each generation is binomial with parameter $(2, p)$, *i.e.*

$$\mathbb{P}(Y = 0) = q^2, \quad \mathbb{P}(Y = 1) = 2pq, \quad \mathbb{P}(Y = 2) = p^2,$$

with $q := 1 - p$.

a) Compute the probability generating function G_Y of Y.
b) Compute the extinction probability of this process, starting from $Z_0 = 1$.
c) Compute the probability that the population becomes extinct for the first time in the second generation ($n = 2$), starting from $Z_0 = 1$.
d) Suppose that the initial population size Z_0 is a Poisson random variable with parameter $\lambda > 0$. Compute the extinction probability in this case.

Exercise 8.6 A cell culture is started with one red cell at time 0. After one minute the red cell dies and two new cells are born according to the following probability distribution:

Color configuration	Probability
2 red cells	1/4
1 red cell + 1 white cell	2/3
2 white cells	1/12

The above procedure is repeated minute after minute for any red cell present in the culture. On the other hand, the white cells can only live for one minute, and disappear after that without reproducing. We assume that the cells behave independently.

a) What is the probability that no white cells have been generated until time n included?
b) Compute the extinction probability of the whole cell culture.
c) Same questions as above for the following probability distribution:

Color configuration	Probability
2 red cells	1/3
1 red cell + 1 white cell	1/2
2 white cells	1/6

Exercise 8.7 Using first step analysis, show that if $(X_n)_{n\geqslant 0}$ is a subcritical branching process, *i.e.* $\mu = \mathbb{E}[Y_1] < 1$, the time to extinction $T_0 := \inf\{n \geqslant 0 \ : \ X_n = 0\}$ satisfies $\mathbb{E}[T_0 \mid X_0 = 1] < \infty$.

Exercise 8.8 Consider a branching process $(X_n)_{n\geqslant 0}$ started at $X_0 = 1$, in which the numbers Y_k of descendants of individual $n^o k$ form an *i.i.d.* sequence with the negative binomial distribution

$$\mathbb{P}(Y_k = n) = (n+1)q^2 p^n, \qquad n \geqslant 0, \quad k \geqslant 1,$$

where $0 < q = 1 - p < 1$.

a) Compute the probability generating function

$$G_1(s) := \mathbb{E}\big[s^{Y_k}\big] = \sum_{n\geqslant 0} s^n \mathbb{P}(Y_1 = n)$$

of Y_k, $k \geqslant 1$.

b) Compute the extinction probability $\alpha_1 := \mathbb{P}(T_0 < \infty \mid X_0 = 1)$ of the branching process $(X_n)_{n\geqslant 0}$ in finite time.

Exercise 8.9 Families in a given society have children until the birth of the first girl, after which the family stops having children. Let X denote the number of male children of a given husband.

a) Assuming that girls and boys are equally likely to be born, compute the probability distribution of X.
b) Compute the probability generating function $G_X(s)$ of X.
c) What is the probability that a given man has no male descendant (patrilineality) by the time of the third generation?
d) Suppose now that one fourth of the married couples have no children at all while the others continue to have children until the first girl, and then cease childbearing. What is the probability that the wife's female line of descent (matrilineality) will cease to exist by the third generation?

Exercise 8.10 Consider a branching process $(Z_k)_{k\geqslant 0}$ with $Z_0 = 1$ and offspring distribution given by

$$\mathbb{P}(Z_1 = 0) = \frac{1-p-q}{1-p} \quad \text{and} \quad \mathbb{P}(Z_1 = k) = qp^{k-1}, \qquad k = 1, 2, 3, \ldots,$$

where $0 \leqslant p < 1$ and $0 \leqslant q \leqslant 1 - p$.

a) Find the probability generating function of Z_1.
b) Compute $\mathbb{E}[Z_1]$.
c) Find the value of q for which $\mathbb{E}[Z_1] = 1$, known as the *critical value.*

d) Using the critical value of q, show by induction that the probability generating function of Z_k is given by

$$G_{Z_k}(s) = \frac{kp - (kp + p - 1)s}{1 - p + kp - kps}, \qquad -1 < s < 1,$$

for all $k \geqslant 1$.

Exercise 8.11 Borel distribution. Consider a branching process started from one individual, in which every individual gives rise to a Poisson number N of offsprings with the Poisson distribution with mean $\mu > 0$, *i.e.*

$$\mathbb{P}(N = n) = \mathrm{e}^{-\mu} \frac{\mu^n}{n!}, \qquad n \geqslant 0.$$

We let X denote the total number of descendants of a given ancestor, including this ancestor and all subsequent ancestors.

a) Show that the Probability Generating Function (PGF) of X satisfies the induction relation

$$G(s) = sG_\mu(G(s)), \qquad -1 \leqslant s \leqslant 1,$$

where

$$G_\mu(s) = \mathrm{e}^{-\mu} \sum_{n=0}^{\infty} \frac{\mu^n}{n!} s^n = \mathrm{e}^{\mu(s-1)}$$

denotes the PGF of the Poisson distribution with mean $\mu > 0$.

b) Show that the mean count of all descendants including all ancestors is given by

$$\mathbb{E}[X] = \frac{1}{1 - \mu}.$$

c) Show that the variance of the count of all descendants including all ancestors is given by

$$\mathrm{Var}[X] = \frac{\mu}{(1 - \mu)^3}.$$

Problem 8.12 Consider a branching process with *i.i.d.* offspring sequence $(Y_k)_{k \geqslant 1}$. The number of individuals in the population at generation $n + 1$ is given by the relation $X_{n+1} = Y_1 + \cdots + Y_{X_n}$, with $X_0 = 1$.

a) Let

$$Z_n = \sum_{k=1}^{n} X_k,$$

denote the total number of individuals generated from time 1 to n. Compute $\mathbb{E}[Z_n]$ as a function of $\mu = \mathbb{E}[Y_1]$.

b) Let $Z = \sum_{k \geqslant 1} X_k$. denote the total number of individuals generated from time 1 to infinity. Compute $\mathbb{E}[Z]$ and show that it is finite when $\mu < 1$.

In what follows we work under the condition $\mu < 1$.

c) Let

$$H(s) = \mathbb{E}\big[s^Z\big], \qquad -1 \leqslant s \leqslant 1,$$

denote the generating function of Z.

Show, by first step analysis, that the relation

$$H(s) = G_1(sH(s)), \qquad 0 \leqslant s \leqslant 1,$$

holds, where $G_1(x)$ is the probability generating function of Y_1.

d) In what follows we assume that Y_1 has the geometric distribution $\mathbb{P}(Y_1 = k) = qp^k$, $k \in \mathbb{N}$, with $p \in (0,1)$ and $q = 1 - p$. Compute $H(s)$ for $s \in [0,1]$.

e) Using the expression of the generating function $H(s)$ computed in Question (d), check that we have $H(0) = \lim_{s \searrow 0} H(s)$, where $H(0) = \mathbb{P}(Z = 0) = \mathbb{P}(Y_1 = 0) = G_1(0)$.

f) Using the generating function $H(s)$ computed in Question (d), recover the value of $\mathbb{E}[Z]$ found in Question (b).

g) Assume that each of the Z individuals earns an income U_k, $k = 1, 2, \ldots, Z$, where $(U_k)_{k \geqslant 1}$ is an *i.i.d.* sequence of random variables with finite expectation $\mathbb{E}[U]$ and distribution function $F(x) = \mathbb{P}(U \leqslant x)$.

Compute the expected value of the sum of gains of all the individuals in the population.

h) Compute the probability that none of the individuals earns an income higher than $x > 0$.

i) Evaluate the results of Questions (g) and (h) when U_k has the exponential distribution with $F(x) = 1 - \mathrm{e}^{-x}$, $x \geqslant 0$.

Hints and comments on Problem 8.12.

a) Use the expression (8.2.5) of $\mathbb{E}[X_k]$.

b) $\mathbb{E}[Z] < \infty$ implies that $Z < \infty$ almost surely.

c) Given that $X_1 = k$, Z can be decomposed into the sum of k independent population sizes, each of them started with 1 individual.

d) Compute $G_1(s)$ and μ in this model, and check the condition $\mu < 1$. When computing $H(s)$ you should have to solve a quadratic equation, and to choose the relevant solution out of two possibilities.

e) The goal of this question is to confirm the result of Question (d) by checking the value of $H(0) = \lim_{s \searrow 0} H(s)$. For this, find the Talyor expansion of $\sqrt{1 - 4pqs}$ as s tends to 0.

f) The identity $1 - 4pq = (q - p)^2$ can be useful, and the sign of $q - p$ has to be taken into account when computing $\sqrt{1 - 4pq}$.

g) Use conditioning on $Z = n$, $n \in \mathbb{N}$.

h) The answer will use both H and F.
i) Compute the value of $\mathbb{E}[U_1]$.

Chapter 9
Continuous-Time Markov Chains

In this chapter we start the study of *continuous-time* stochastic processes, which are families $(X_t)_{t\in\mathbb{R}_+}$ of random variables indexed by $\mathbb{R}_+$. Our aim is to make the transition from discrete to continuous-time Markov chains, the main difference between the two settings being the replacement of the transition matrix with the continuous-time *infinitesimal generator* of the process. We will start with the two fundamental examples of the Poisson and birth and death processes, followed by the construction of continuous-time Markov chains and their generators in more generality. From the point of view of simulations, the use of continuous-time Markov chains does not bring any special difficulty as any continuous-time simulation is actually based on discrete-time samples. From a theoretical point of view, however, the rigorous treatment of the continuous-time Markov property is much more demanding than its discrete-time counterpart, notably due to the use of the strong Markov property. Here we focus on the understanding of the continuous-time case by simple calculations, and we will refer to the literature for the use of the strong Markov property.

9.1 The Poisson Process

The *standard Poisson process* $(N_t)_{t\in\mathbb{R}_+}$ is a continuous-time *counting* process, *i.e.* $(N_t)_{t\in\mathbb{R}_+}$ has jumps of size $+1$ only, and its paths are constant (and right-continuous) in between jumps. The next Figure 9.1 represents a sample path of a Poisson process.

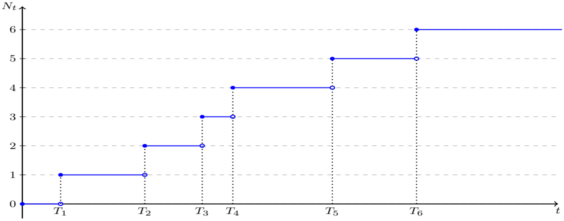


Fig. 9.1: Sample path of a Poisson process $(N_t)_{t\in\mathbb{R}_+}$.

We denote by $(T_k)_{k\geqslant 1}$ the increasing sequence of jump times of $(N_t)_{t\in\mathbb{R}_+}$, which can be defined from the (right-continuous) Poisson process path $(N_t)_{t\in\mathbb{R}_+}$ by noting that T_k is the first hitting time of state k, *i.e.*

$$T_k = \inf\{t\in\mathbb{R}_+ \ : \ N_t = k\}, \qquad k\geqslant 1,$$

with

$$\lim_{k\to\infty} T_k = \infty.$$

The value N_t at time t of the Poisson process can be recovered from its jump times $(T_k)_{k\geqslant 1}$ as

$$N_t = \sum_{k\geqslant 1} k \mathbb{1}_{[T_k,T_{k+1})}(t) = \sum_{k\geqslant 1} \mathbb{1}_{[T_k,\infty)}(t), \qquad t\geqslant 0,$$

where

$$\mathbb{1}_{[T_k,\infty)}(t) = \begin{cases} 1 \text{ if } t\geqslant T_k, \\ \\ 0 \text{ if } 0\leqslant t < T_k, \end{cases}$$

and

$$\mathbb{1}_{[T_k,T_{k+1})}(t) = \begin{cases} 1 \text{ if } T_k\leqslant t < T_{k+1}, \qquad k\geqslant 0, \\ \\ 0 \text{ if } 0\leqslant t < T_k \text{ or } t\geqslant T_{k+1}, \qquad k\geqslant 0. \end{cases}$$

with $T_0 = 0$.

In addition, $(N_t)_{t\in\mathbb{R}_+}$ is assumed to satisfy the following conditions:

i) Independence of increments: for all $0 \leqslant t_0 < t_1 < \cdots < t_n$ and $n \geqslant 1$ the increments

$$N_{t_1} - N_{t_0}, \ldots, \ N_{t_n} - N_{t_{n-1}},$$

over the disjoint time intervals $[t_0, t_1)$, $[t_1, t_2)$, $\ldots$, $[t_{n-2}, t_{n-1})$, $[t_{n-1}, t_n]$ are mutually independent random variables.

ii) Stationarity of increments: $N_{t+h} - N_{s+h}$ has the same distribution as $N_t - N_s$ for all $h > 0$ and $0 \leqslant s \leqslant t$.

The meaning of the above stationarity condition is that for all fixed $k \in \mathbb{N}$ we have

$$\mathbb{P}(N_{t+h} - N_{s+h} = k) = \mathbb{P}(N_t - N_s = k),$$

for all $h > 0$ and $0 \leqslant s \leqslant t$.

The stationarity of increments means that for all $k \in \mathbb{N}$, the probability $\mathbb{P}(N_{t+h} - N_{s+h} = k)$ does not depend on $h > 0$.

Based on the above assumption, a natural question arises:

what is the distribution of N_t at time t?

We already know that N_t takes values in $\mathbb{N}$ and therefore it has a discrete distribution for all $t \in \mathbb{R}_+$. It is a remarkable fact that the distribution of the increments of $(N_t)_{t\in\mathbb{R}_+}$, can be completely determined from the above conditions, as shown in the following theorem.

As seen in the next result, the random variable $N_t - N_s$ has the Poisson distribution with parameter $\lambda(t - s)$.

Theorem 9.1. *Assume that the counting process $(N_t)_{t\in\mathbb{R}_+}$ satisfies the independence and stationarity Conditions (i) and (ii) above. Then we have*

$$\mathbb{P}(N_t - N_s = k) = \mathrm{e}^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}, \qquad k \in \mathbb{N}, \quad 0 \leqslant s \leqslant t,$$

for some constant $\lambda > 0$.

Theorem 9.1 shows in particular that

$$\mathbb{E}[N_t - N_s] = \lambda(t - s) \quad \text{and} \quad \mathrm{Var}[N_t - N_s] = \lambda(t - s),$$

$0 \leqslant s \leqslant t$, cf. Relations (S.4) and (S.5) in the solution of Exercise 1.4-(a).

The parameter $\lambda > 0$ is called the *intensity* of the process and it can be recovered given from $\mathbb{P}(N_h = 1) = \lambda h \mathrm{e}^{-\lambda h}$ as the limit

$$\lambda = \lim_{h \searrow 0} \frac{1}{h} \mathbb{P}(N_h = 1). \tag{9.1.1}$$

Proof of Theorem 9.1. We only quote the main steps of the proof and we refer to Bosq and Nguyen (1996) for the complete argument. Using the independence and stationarity of increments, we show that the probability generating function

$$G_t(u) := \mathbb{E}[u^{N_t}], \qquad -1 \leqslant u \leqslant 1,$$

satisfies

$$G_t(u) := (G_1(u))^t, \qquad -1 \leqslant u \leqslant 1,$$

which implies that

$$G_t(u) := \mathrm{e}^{-tf(u)}, \qquad -1 \leqslant u \leqslant 1,$$

for some function $f(u)$ of u. Still relying on the independence and stationarity of increments, it can be shown that $f(u)$ takes the form

$$f(u) = \lambda \times (1-u), \qquad -1 \leqslant u \leqslant 1,$$

where $\lambda > 0$ is given by (9.1.1). □

In particular, given that $N_0 = 0$, the random variable N_t has a Poisson distribution with parameter λt:

$$\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}, \qquad t \geqslant 0.$$

From (9.1.1) above we see that*

$$\begin{cases} \mathbb{P}(N_h = 0) = \mathrm{e}^{-\lambda h} = 1 - \lambda h + o(h), \qquad h \searrow 0, \\ \\ \mathbb{P}(N_h = 1) = h\lambda \mathrm{e}^{-\lambda h} \simeq \lambda h, \qquad h \searrow 0, \end{cases}$$

and more generally that

$$\begin{cases} \mathbb{P}(N_{t+h} - N_t = 0) = \mathrm{e}^{-\lambda h} = 1 - \lambda h + o(h), \qquad h \searrow 0, & \text{(9.1.2a)} \\ \\ \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h \mathrm{e}^{-\lambda h} \simeq \lambda h, \qquad h \searrow 0, & \text{(9.1.2b)} \\ \\ \mathbb{P}(N_{t+h} - N_t = 2) = h^2 \frac{\lambda^2}{2}\mathrm{e}^{-\lambda h} \simeq h^2 \frac{\lambda^2}{2}, \qquad h \searrow 0, & \text{(9.1.2c)} \end{cases}$$

for all $t \geqslant 0$. This means that within "short" time intervals $[kh, (k+1)h]$ of length $h = t/n > 0$, the increments $N_{(k+1)h} - N_{kh}$ can be approximated by independent Bernoulli random variables X_{kh} with parameter λh, whose sum

* The notation $f(h) \simeq h^k$ means $\lim_{h\to 0} f(h)/h^k = 1$, and $f(h) = o(h)$ means $\lim_{h\to 0} f(h)/h = 0$.

$$\sum_{k=0}^{n-1} X_{kh} \simeq \sum_{k=0}^{n-1} (N_{(k+1)h} - N_{kh}) = N_t - N_0 = N_t$$

converges in distribution as n goes to infinity to the Poisson random variable N_t with parameter λt. This remark can be used for the random simulation of Poisson process paths.

More generally, we have

$$\mathbb{P}(N_{t+h} - N_t = k) \simeq h^k \frac{\lambda^k}{k!}, \qquad h \searrow 0, \qquad t > 0.$$

In order to determine the distribution of the first jump time T_1 we note that we have the equivalence

$$\{T_1 > t\} \Longleftrightarrow \{N_t = 0\},$$

which implies

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0,$$

i.e. T_1 has an exponential distribution with parameter $\lambda > 0$.

In order to prove the next proposition we note that more generally, we have the equivalence

$$\{T_n > t\} = \{N_t < n\}, \qquad t \geqslant 0, \quad n \geqslant 1.$$

Indeed, stating that the *n-th* jump time T_n is strictly larger than t is equivalent to saying that at most $n-1$ jumps of the Poisson process have occurred over the interval $[0, t]$, *i.e.* $N_t \leqslant n-1$. The next proposition shows that T_n has a gamma distribution with parameter (λ, n) for $n \geqslant 1$, also called the Erlang distribution in queueing theory.

Proposition 9.2. *The random variable T_n has the gamma probability density function*

$$x \longmapsto \lambda^n \mathrm{e}^{-\lambda x} \frac{x^{n-1}}{(n-1)!}$$

$x \geqslant 0$, $n \geqslant 1$.

Proof. For $n = 1$ we have

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0,$$

and by induction on $n \geqslant 1$, assuming that

$$\mathbb{P}(T_{n-1} > t) = \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-2}}{(n-2)!} ds,$$

at the rank $n-1$ with $n \geqslant 2$, we obtain

$$\begin{aligned}
\mathbb{P}(T_n > t) &= \mathbb{P}(T_n > t \geqslant T_{n-1}) + \mathbb{P}(T_{n-1} > t) \\
&= \mathbb{P}(N_t = n-1) + \mathbb{P}(T_{n-1} > t) \\
&= \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} + \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-2}}{(n-2)!} ds \\
&= \lambda \int_t^\infty \mathrm{e}^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} ds, \qquad t \geqslant 0,
\end{aligned}$$

which proves the desired relation at the rank n, where we applied an integration by parts on $\mathbb{R}_+$ to derive the last line. □

Let now

$$\tau_k = T_{k+1} - T_k, \qquad k \geqslant 1,$$

denote the time spent in state $k \in \mathbb{N}$, with $T_0 = 0$. In addition to Proposition 9.2 we could show the following proposition which is based on the *strong Markov property*, see *e.g.* Theorem 6.5.4 of Norris (1998), (9.2.4) below and Exercise 5.9 in discrete time.

Proposition 9.3. *The random inter-jump times*

$$\tau_k := T_{k+1} - T_k$$

spent in state $k \in \mathbb{N}$ form a sequence of independent identically distributed random variables having the exponential distribution with parameter $\lambda > 0$, i.e.

$$\mathbb{P}(\tau_0 > t_0, \tau_1 > t_1, \ldots, \tau_n > t_n) = \mathrm{e}^{-\lambda(t_0 + t_1 + \cdots + t_n)}, \qquad t_0, t_1, \ldots, t_n \geqslant 0.$$

Random samples of Poisson process jump times can be generated using the following R code.

```
lambda = 2.0;n = 10
taun <- rexp(n)/lambda; Tn <- cumsum(taun)
taun
Tn
```

Similarly, random samples of Poisson process paths can be generated using the following code.

```
n<-100
x<-cumsum(rexp(50,rate=0.5))
y<-cumsum(c(0,rep(1,50)))
plot(stepfun(x,y),xlim = c(0,10),do.points = F,main="L=0.5")
```

In other words, we have

$$\mathbb{P}(\tau_0 > t_0, \tau_1 > t_1, \ldots, \tau_n > t_n) = \mathbb{P}(\tau_0 > t_0) \times \cdots \times \mathbb{P}(\tau_n > t_n) \quad (9.1.3)$$
$$= \prod_{k=0}^{n} \mathrm{e}^{-\lambda t_k}$$
$$= \mathrm{e}^{-\lambda(t_0 + \cdots + t_n)},$$

for all $t_0, t_1, \ldots, t_n \geqslant 0$. In addition, from Proposition 9.2 the sum

$$T_k = \tau_0 + \tau_1 + \cdots + \tau_{k-1}, \qquad k \geqslant 1,$$

has a gamma distribution with parameter (λ, k), cf. also Exercise 9.13 for a proof in the particular case $k = 2$.

As the expectation of the exponentially distributed random variable τ_k with parameter $\lambda > 0$ is given by

$$\mathbb{E}[\tau_k] = \lambda \int_0^\infty x \mathrm{e}^{-\lambda x} dx = \frac{1}{\lambda},$$

we can check that the higher the intensity λ (*i.e.* the higher the probability of having a jump within a small interval), the smaller is the time spent in each state $k \in \mathbb{N}$ on average. Poisson random samples on arbitrary spaces will be considered in Chapter 11.

9.2 Continuous-Time Markov Chains

A $\mathbb{S}$-valued continuous-time stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *Markov*, or to have the *Markov property* if, for all $t \in [s, \infty)$, the probability distribution of X_t given the past of the process up to time s is determined by the state X_s of the process at time s, and does not depend on the past values of X_u for $u < s$. In other words, for all

$$0 < s_1 < \cdots < s_{n-1} < s < t$$

we have

$$\mathbb{P}(X_t = j \mid X_s = i_n,\ X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_1} = i_0) = \mathbb{P}(X_t = j \mid X_s = i_n). \quad (9.2.1)$$

In particular, we have

$$\mathbb{P}(X_t = j \mid X_s = i_n \text{ and } X_{s_{n-1}} = i_{n-1}) = \mathbb{P}(X_t = j \mid X_s = i_n).$$

Example

The Poisson process $(N_t)_{t\in\mathbb{R}_+}$ considered in Section 9.1 is a continuous-time Markov chain because it has independent increments by Condition (i) page 306. The birth and death processes discussed below are also continuous-time Markov chains, although they may not have independent increments.

More generally, any continuous-time process $(X_t)_{t\in\mathbb{R}_+}$ with independent increments has the Markov property. Indeed, for all $j, i_n, \ldots, i_1 \in \mathbb{S}$ we have (note that $X_0 = 0$ here)

$$\begin{aligned}
&\mathbb{P}(X_t = j \mid X_s = i_n,\ X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_1} = i_1)\\
&= \frac{\mathbb{P}(X_t = j, X_s = i_n, X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_1} = i_1)}{\mathbb{P}(X_s = i_n,\ X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_1} = i_1)}\\
&= \frac{\mathbb{P}(X_t - X_s = j - i_n, X_s = i_n, \ldots, X_{s_2} = i_2, X_{s_1} = i_1)}{\mathbb{P}(X_s = i_n, , \ldots, X_{s_2} = i_2, X_{s_1} = i_1)}\\
&= \frac{\mathbb{P}(X_t - X_s = j - i_n)\mathbb{P}(X_s = i_n, X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_2} = i_2, X_{s_1} = i_1)}{\mathbb{P}(X_s = i_n, X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_2} = i_2, X_{s_1} = i_1)}\\
&= \mathbb{P}(X_t - X_s = j - i_n) = \frac{\mathbb{P}(X_t - X_s = j - i_n)\mathbb{P}(X_s = i_n)}{\mathbb{P}(X_s = i_n)}\\
&= \frac{\mathbb{P}(X_t - X_s = j - i_n \text{ and } X_s = i_n)}{\mathbb{P}(X_s = i_n)}\\
&= \frac{\mathbb{P}(X_t = j \text{ and } X_s = i_n)}{\mathbb{P}(X_s = i_n)} = \mathbb{P}(X_t = j \mid X_s = i_n),
\end{aligned}$$

cf. (4.1.6) for the discrete-time version of this argument. Hence, continuous-time processes with *independent increments* are Markov chains. However, *not all continuous-time Markov chains have independent increments*, and in fact the continuous-time Markov chains of interest in this chapter will not have independent increments.

Birth process

The pure birth process behaves similarly to the Poisson process, by making the parameter of every exponential inter-jump time dependent on the current state of the process.

In other words, a continuous-time Markov chain $(X^b_t)_{t\in\mathbb{R}_+}$ such that*

$$\begin{aligned}
\mathbb{P}(X^b_{t+h} = i+1 \mid X^b_t = i) &= \mathbb{P}(X^b_{t+h} - X^b_t = 1 \mid X^b_t = i)\\
&\simeq \lambda_i h, \qquad h \searrow 0, \quad i \in \mathbb{S},
\end{aligned}$$

and

* Recall that by definition $f(h) \simeq g(h)$, $h \to 0$, if and only if $\lim_{h\to 0} f(h)/g(h) = 1$.

$$\mathbb{P}(X^b_{t+h} = X^b_t \mid X^b_t = i) = \mathbb{P}(X^b_{t+h} - X^b_t = 0 \mid X^b_t = i)$$
$$= 1 - \lambda_i h + o(h), \qquad h \searrow 0, \quad i \in \mathbb{S}, \tag{9.2.2}$$

is called a *pure birth process* with (possibly) state-dependent birth rates $\lambda_i \geqslant 0$, $i \in \mathbb{S}$, see Figure 9.2. Its inter-jump times $(\tau_k)_{k\geqslant 0}$ form a sequence of exponential independent random variables with state-dependent parameters.

This process is stationary in time because the rates λ_i, $i \in \mathbb{N}$, are independent of time t. The Poisson process $(N_t)_{t\in\mathbb{R}_+}$ is a pure birth process with state-independent birth rates $\lambda_i = \lambda > 0$, $i \in \mathbb{N}$.

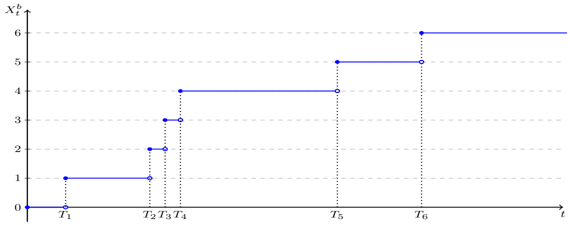


Fig. 9.2: Sample path of a birth process $(X^b_t)_{t\in\mathbb{R}_+}$.

As a consequence of (9.2.2) we can recover the fact that the time $\tau_{i,i+1}$ spent in state (i) by the pure birth process $(X^b_t)_{t\in\mathbb{R}_+}$ started at state (i) at time 0 before it moves to state $(i+1)$ has an exponential distribution with parameter λ_i. Indeed we have, using the Markov property in continuous time,

$$\begin{aligned}
\mathbb{P}(\tau_{i,i+1} > t+h \mid \tau_{i,i+1} > t \text{ and } X^b_0 = i) &= \frac{\mathbb{P}(\tau_{i,i+1} > t+h \mid X^b_0 = i)}{\mathbb{P}(\tau_{i,i+1} > t \mid X^b_0 = i)} \\
&= \frac{\mathbb{P}(X^b_{t+h} = i \mid X^b_0 = i)}{\mathbb{P}(X^b_t = i \mid X^b_0 = i)} \\
&= \frac{\mathbb{P}(X^b_{t+h} = i \text{ and } X^b_0 = i)\mathbb{P}(X^b_0 = i)}{\mathbb{P}(X^b_t = i \text{ and } X^b_0 = i)\mathbb{P}(X^b_0 = i)} \\
&= \frac{\mathbb{P}(X^b_{t+h} = i \text{ and } X^b_0 = i)}{\mathbb{P}(X^b_t = i \text{ and } X^b_0 = i)} \\
&= \frac{\mathbb{P}(X^b_{t+h} = i,\ X^b_t = i,\ X^b_0 = i)}{\mathbb{P}(X^b_t = i \text{ and } X^b_0 = i)} \\
&= \mathbb{P}(X^b_{t+h} = i \mid X^b_t = i \text{ and } X^b_0 = i) \\
&= \mathbb{P}(X^b_{t+h} = i \mid X^b_t = i) \\
&= \mathbb{P}(X^b_h = i \mid X^b_0 = i)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(\tau_{i,i+1} > h \mid X_0^b = i) \\
&= 1 - \lambda_i h + o(h), \qquad (9.2.3)
\end{aligned}
$$

which is often referred to as the *memoryless property* of Markov processes. In other words, since the above ratio is independent of $t > 0$ we get

$$
\mathbb{P}(\tau_{i,i+1} > t + h \mid \tau_{i,i+1} > t \text{ and } X_0^b = i) = \mathbb{P}(\tau_{i,i+1} > h \mid X_0^b = i),
$$

which means that the distribution of the waiting time after time t does not depend on t, cf. (12.1.1) in Chapter 12 for a similar argument.

From (9.2.3) we have

$$
\begin{aligned}
&\frac{\mathbb{P}(\tau_{i,i+1} > t + h \mid X_0^b = i) - \mathbb{P}(\tau_{i,i+1} > t \mid X_0^b = i)}{h\mathbb{P}(\tau_{i,i+1} > t \mid X_0^b = i)} \\
&= \frac{1}{h}\left(\frac{\mathbb{P}(\tau_{i,i+1} > t + h \mid X_0^b = i)}{\mathbb{P}(\tau_{i,i+1} > t \mid X_0^b = i)} - 1\right) \\
&= \frac{\mathbb{P}(\tau_{i,i+1} > t + h \mid \tau_{i,i+1} > t \text{ and } X_0^b = i) - 1}{h} \\
&\simeq -\lambda_i, \qquad h \to 0,
\end{aligned}
$$

which can be read as the differential equation

$$
\frac{d}{dt} \log \mathbb{P}(\tau_{i,i+1} > t \mid X_0^b = i) = -\lambda_i,
$$

where "log" denotes the *natural logarithm* "ln", with solution

$$
\mathbb{P}(\tau_{i,i+1} > t \mid X_0^b = i) = \mathrm{e}^{-\lambda_i t}, \qquad t \geqslant 0, \qquad (9.2.4)
$$

i.e. $\tau_{i,i+1}$ is an exponentially distributed random variable with parameter λ_i, and the mean time spent at state (i) before switching to state $(i+1)$ is given by

$$
\mathbb{E}[\tau_{i,i+1} \mid X_0^b = i] = \frac{1}{\lambda_i}, \qquad i \in \mathbb{S},
$$

see (9.3.12) below for the general case of continuous-time Markov chains. More generally, and similarly to (9.1.3) it can also be shown as a consequence of the *strong Markov property* that the sequence $(\tau_{j,j+1})_{j\geqslant i}$ is made of independent random variables which are respectively exponentially distributed with parameters λ_j, $j \geqslant i$.

Letting $T_{i,j}^b = \tau_{i,i+1} + \cdots + \tau_{j-1,j}$ denote the hitting time of state (j) starting from state (i) by the birth process $(X_t^b)_{t\in\mathbb{R}_+}$, we have the representation

$$
X_t^b = i + \sum_{i<j<\infty} \mathbb{1}_{[T_{i,j}^b,\infty)}(t), \qquad t \geqslant 0.
$$

Note that since the pure birth process has stationary increments, by Theorem 9.1 it has independent increments when the rates $\lambda_i = \lambda$ are state independent, *i.e.* when $(X_t^b)_{t\in\mathbb{R}_+}$ is a standard Poisson process with intensity $\lambda > 0$.

Death process

A continuous-time Markov chain $(X_t^d)_{t\in\mathbb{R}_+}$ such that

$$\begin{cases} \mathbb{P}(X_{t+h}^d - X_t^d = -1 \mid X_t^d = i) \simeq \mu_i h, & h \searrow 0, \quad i \in \mathbb{S}, \\ \mathbb{P}(X_{t+h}^d - X_t^d = 0 \mid X_t^d = i) = 1 - \mu_i h + o(h), & h \searrow 0, \quad i \in \mathbb{S}, \end{cases}$$

is called a *pure death process* with (possibly) state-dependent death rates $\mu_i \geqslant 0$, $i \in \mathbb{S}$. Its inter-jump times $(\tau_k)_{k\geqslant 0}$ form a sequence of exponential independent random variables with state-dependent parameters, see Figure 9.3.

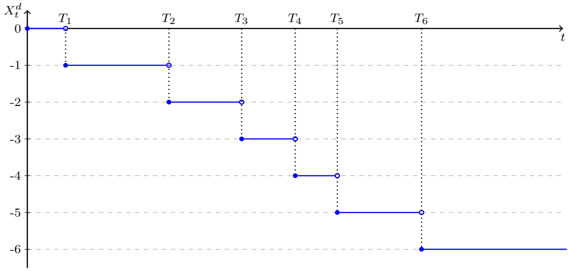


Fig. 9.3: Sample path of a death process $(X_t^d)_{t\in\mathbb{R}_+}$.

In the case of a pure death process $(X_t^d)_{t\in\mathbb{R}_+}$ we denote by $\tau_{i,i-1}$ the time spent in state (i) by $(X_t^d)_{t\in\mathbb{R}_+}$ before it moves to state $\boxed{i-1}$. Similarly to the pure birth process, that the sequence $(\tau_{j,j-1})_{j\leqslant i}$ is made of independent random variables which are exponentially distributed with parameter μ_j, $j \leqslant i$, with

$$\mathbb{P}(\tau_{j,j-1} > t) = \mathrm{e}^{-\mu_j t}, \qquad t \geqslant 0,$$

and

$$\mathbb{E}[\tau_{i,i-1}] = \frac{1}{\mu_i}, \qquad i \in \mathbb{S}.$$

Letting $T_{i,j}^d = \tau_{i,i-1} + \cdots + \tau_{j+1,j}$ denote the hitting time of state (j) starting from state (i) by the death process $(X_t^d)_{t\in\mathbb{R}_+}$ we have the representation

$$X_t^d = i - \sum_{-\infty<j<i} \mathbb{1}_{[T_{i,j}^d,\infty)}(t), \qquad t \geqslant 0.$$

When $(N_t)_{t\in\mathbb{R}_+}$ is a Poisson process, the process $(-N_t)_{t\in\mathbb{R}_+}$ is a pure death process with state-independent death rates $\mu_n = \lambda > 0$, $n \in \mathbb{N}$.

Birth and death process

A continuous-time Markov chain $(X_t)_{t\in\mathbb{R}_+}$ such that, for all $i \in \mathbb{S}$,

$$\begin{cases} \mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = i) \simeq \lambda_i h, \quad h \searrow 0, & (9.2.5a) \\ \mathbb{P}(X_{t+h} - X_t = -1 \mid X_t = i) \simeq \mu_i h, \quad h \searrow 0, \text{ and} \\ \mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = i) = 1 - (\lambda_i + \mu_i)h + o(h), \quad h \searrow 0, & (9.2.5b) \end{cases}$$

is called a *birth and death process* with (possibly) state-dependent birth rates $\lambda_i \geqslant 0$ and death rates $\mu_i \geqslant 0$, $i \in \mathbb{S}$, see Figure 9.4.

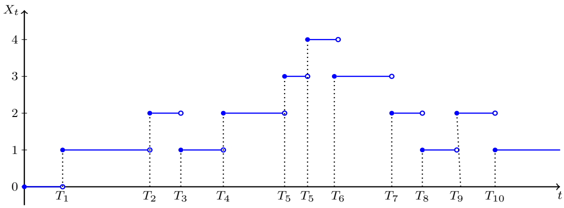


Fig. 9.4: Sample path of a birth and death process $(X_t)_{t\in\mathbb{R}_+}$.

The birth and death process $(X_t)_{t\in\mathbb{R}_+}$ can be built as

$$X_t = X_t^b + X_t^d, \qquad t \geqslant 0,$$

in which case the time τ_i spent in state (i) by $(X_t)_{t\in\mathbb{R}_+}$ satisfies the identity in distribution

$$\tau_i = \min(\tau_{i,i+1}, \tau_{i,i-1})$$

i.e. τ_i is an exponentially distributed random variable with parameter $\lambda_i + \mu_i$ and

$$\mathbb{E}[\tau_i] = \frac{1}{\lambda_i + \mu_i}.$$

Indeed, since $\tau_{i,i+1}$ and $\tau_{i,i-1}$ are two independent exponentially distributed random variables with parameters λ_i and μ_i, we have

$$\begin{aligned}\mathbb{P}(\min(\tau_{i,i+1} \text{ and } \tau_{i,i-1}) > t) &= \mathbb{P}(\tau_{i,i+1} > t \text{ and } \tau_{i,i-1} > t)\\ &= \mathbb{P}(\tau_{i,i+1} > t)\mathbb{P}(\tau_{i,i-1} > t)\\ &= \mathrm{e}^{-t(\lambda_i+\mu_i)}, \qquad t \geqslant 0,\end{aligned}$$

hence $\tau_i = \min(\tau_{i,i+1}, \tau_{i,i-1})$ is an exponentially distributed random variable with parameter $\lambda_i + \mu_i$, cf. also (1.5.8) in Chapter 1.

9.3 Transition Semigroup and Infinitesimal Generator

The transition semigroup of the continuous time Markov process $(X_t)_{t\in\mathbb{R}_+}$ is the family $(P(t))_{t\in\mathbb{R}_+}$ of matrices determined by

$$P_{i,j}(t) := \mathbb{P}(X_{t+s} = j \mid X_s = i), \qquad i, j \in \mathbb{S}, \quad s, t \geqslant 0,$$

where we assume that the probability $\mathbb{P}(X_{t+s} = j \mid X_s = i)$ does not depend on $s \in \mathbb{R}_+$. In this case the Markov process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *time homogeneous*.

Definition 9.4. *A continuous-time Markov chain* $(X_t)_{t\in\mathbb{R}_+}$ *is* irreducible *if for all* $t > 0$, $P(t)$ *is the transition matrix of an irreducible discrete-time chain.*

Note that we always have

$$P(0) = \mathrm{I}.$$

These data can be recorded as a time-dependent matrix indexed by $\mathbb{S}^2 = \mathbb{S} \times \mathbb{S}$, called the *transition semigroup* of the Markov process:

$$[\ P_{i,j}(t)\ ]_{i,j\in\mathbb{S}} = [\ \mathbb{P}(X_{t+s} = j \mid X_s = i)\ ]_{i,j\in\mathbb{S}},$$

also written as

$$[\ P_{i,j}(t)\ ]_{i,j\in\mathbb{S}} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & ⋰ \\ \cdots & P_{-2,-2}(t) & P_{-2,-1}(t) & P_{-2,0}(t) & P_{-2,1}(t) & P_{-2,2}(t) & \cdots \\ \cdots & P_{-1,-2}(t) & P_{-1,-1}(t) & P_{-1,0}(t) & P_{-1,1}(t) & P_{-1,2}(t) & \cdots \\ \cdots & P_{0,-2}(t) & P_{0,-1}(t) & P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & \cdots \\ \cdots & P_{1,-2}(t) & P_{1,-1}(t) & P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & \cdots \\ \cdots & P_{2,-2}(t) & P_{2,-1}(t) & P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & \cdots \\ ⋰ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

As in the discrete-time case, note the inversion of the order of indices (i, j) between $\mathbb{P}(X_{t+s} = j \mid X_s = i)$ and $P_{i,j}(t)$. In particular, the initial state $\textcircled{i}$ correspond to a *row number* in the matrix $P(t)$, while the final state $\textcircled{j}$ corresponds to a *column number*.

Due to the relation

$$\sum_{j\in\mathbb{S}} \mathbb{P}(X_{t+s} = j \mid X_s = i) = 1, \qquad i \in \mathbb{S}, \tag{9.3.1}$$

all *rows* of the transition matrix semigroup $(P(t))_{t\in\mathbb{R}_+}$ satisfy the condition

$$\sum_{j\in\mathbb{S}} P_{i,j}(t) = 1,$$

for $i \in \mathbb{S}$. In the sequel we will only consider $\mathbb{N}$-valued Markov process, and in this case the *transition semigroup* $(P(t))_{t\in\mathbb{R}_+}$ of the Markov process is written as

$$P(t) = [\ P_{i,j}(t)\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & \cdots \\ P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & \cdots \\ P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

From (9.3.1) we have

$$\sum_{j\geqslant 0} P_{i,j}(t) = 1,$$

for all $i \in \mathbb{N}$ and $t \geqslant 0$.

Exercise: Write down the transition semigroup $[\ P_{i,j}(t)\ ]_{i,j\in\mathbb{N}}$ of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$.

We can show that

$$[\ P_{i,j}(t)\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & \frac{\lambda^2 t^2}{2} e^{-\lambda t} & \cdots \\ 0 & e^{-\lambda t} & \lambda t e^{-\lambda t} & \cdots \\ 0 & 0 & e^{-\lambda t} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Indeed we have

$$\begin{aligned} P_{i,j}(t) &= \mathbb{P}(N_{s+t} = j \mid N_s = i) = \frac{\mathbb{P}(N_{s+t} = j \text{ and } N_s = i)}{\mathbb{P}(N_s = i)} \\ &= \frac{\mathbb{P}(N_{s+t} - N_s = j - i \text{ and } N_s = i)}{\mathbb{P}(N_s = i)} = \frac{\mathbb{P}(N_{s+t} - N_s = j - i)\mathbb{P}(N_s = i)}{\mathbb{P}(N_s = i)} \\ &= \mathbb{P}(N_{s+t} - N_s = j - i) = \begin{cases} e^{-\lambda t}\dfrac{(\lambda t)^{j-i}}{(j-i)!} & \text{if } j \geqslant i, \\ 0 & \text{if } j < i. \end{cases} \end{aligned}$$

In case the Markov process $(X_t)_{t\in\mathbb{R}_+}$ takes values in the finite state space $\mathbb{S} = \{0, 1, \ldots, N\}$ its transition semigroup will simply have the form

$$P(t) = [\ P_{i,j}(t)\ ]_{0\leqslant i,j\leqslant N} = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & \cdots & P_{0,N}(t) \\ P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & \cdots & P_{1,N}(t) \\ P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & \cdots & P_{2,N}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0}(t) & P_{N,1}(t) & P_{N,2}(t) & \cdots & P_{N,N}(t) \end{bmatrix}.$$

As noted above, the semigroup matrix $P(t)$ is a convenient way to record the values of $\mathbb{P}(X_{t+s} = j \mid X_s = i)$ in a table.

Proposition 9.5. *The family* $(P(t))_{t\in\mathbb{R}_+}$ *satisfies the relation*

$$P(s+t) = P(s)P(t) = P(t)P(s), \qquad (9.3.2)$$

which is called the semigroup property*.*

Proof. Using the Markov property and denoting by $\mathbb{S}$ the state space of the process, by standard arguments based on the *law of total probability* (1.3.1) for the probability measure $\mathbb{P}(\cdot \mid X_0 = i)$ and the Markov property (9.2.1), we have

$$\begin{aligned}
&P_{i,j}(t+s) = \mathbb{P}(X_{t+s} = j \mid X_0 = i) \\
&= \sum_{l \in \mathbb{S}} \mathbb{P}(X_{t+s} = j \text{ and } X_s = l \mid X_0 = i) = \sum_{l \in \mathbb{S}} \frac{\mathbb{P}(X_{t+s} = j,\ X_s = l,\ X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{l \in \mathbb{S}} \frac{\mathbb{P}(X_{t+s} = j,\ X_s = l,\ X_0 = i)}{\mathbb{P}(X_s = l \text{ and } X_0 = i)} \frac{\mathbb{P}(X_s = l \text{ and } X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{l \in \mathbb{S}} \mathbb{P}(X_{t+s} = j \mid X_s = l \text{ and } X_0 = i) \mathbb{P}(X_s = l \mid X_0 = i) \\
&= \sum_{l \in \mathbb{S}} \mathbb{P}(X_{t+s} = j \mid X_s = l) \mathbb{P}(X_s = l \mid X_0 = i) = \sum_{l \in \mathbb{S}} P_{i,l}(s) P_{l,j}(t) \\
&= [P(s)P(t)]_{i,j},
\end{aligned}$$

$i, j \in \mathbb{S}$, $s, t \geqslant 0$. We have shown the relation

$$P_{i,j}(s+t) = \sum_{l \in \mathbb{S}} P_{i,l}(s) P_{l,j}(t),$$

which leads to (9.3.2). □

From (9.3.2) property one can check in particular that the matrices $P(s)$ and $P(t)$ *commute*, *i.e.* we have

$$P(s)P(t) = P(t)P(s), \qquad s, t \geqslant 0.$$

Example

For the transition semigroup $(P(t))_{t \in \mathbb{R}_+}$ of the Poisson process we can check by hand computation that

$$P(s)P(t) = \begin{bmatrix} \mathrm{e}^{-\lambda s} & \lambda s \mathrm{e}^{-\lambda s} & \frac{\lambda^2}{2} s^2 \mathrm{e}^{-\lambda s} & \cdots \\ 0 & \mathrm{e}^{-\lambda s} & \lambda s \mathrm{e}^{-\lambda s} & \cdots \\ 0 & 0 & \mathrm{e}^{-\lambda s} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \times \begin{bmatrix} \mathrm{e}^{-\lambda t} & \lambda t \mathrm{e}^{-\lambda t} & \frac{\lambda^2}{2} t^2 \mathrm{e}^{-\lambda t} & \cdots \\ 0 & \mathrm{e}^{-\lambda t} & \lambda t \mathrm{e}^{-\lambda t} & \cdots \\ 0 & 0 & \mathrm{e}^{-\lambda t} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} e^{-\lambda(s+t)} & \lambda(s+t)e^{-\lambda(s+t)} & \frac{\lambda^2}{2}(s+t)^2 e^{-\lambda(s+t)} & \cdots \\ 0 & e^{-\lambda(s+t)} & \lambda(s+t)\, e^{-\lambda(s+t)} & \cdots \\ 0 & 0 & e^{-\lambda(s+t)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$
$$= P(s+t).$$

The above identity can be recovered by the following calculation, for all $0 \leqslant i \leqslant j$, which amounts to saying that the sum of two independent Poisson random variables with parameters s and t has a Poisson distribution with parameter $s+t$, cf. (S.9) in the solution of Exercise 1.7-(a). We have $P_{i,j}(s) = 0$, $i > j$, and $P_{l,j}(t) = 0$, $l > j$, hence

$$[P(s)P(t)]_{i,j} = \sum_{l \geqslant 0} P_{i,l}(s) P_{l,j}(t) = \sum_{l=i}^{j} P_{i,l}(s) P_{l,j}(t)$$
$$= e^{-\lambda s - \lambda t} \sum_{l=i}^{j} \frac{(\lambda s)^{l-i}}{(l-i)!} \frac{(\lambda t)^{j-l}}{(j-l)!} = e^{-\lambda s - \lambda t} \frac{1}{(j-i)!} \sum_{l=i}^{j} \binom{j-i}{l-i} (\lambda s)^{l-i} (\lambda t)^{j-l}$$
$$= e^{-\lambda s - \lambda t} \frac{1}{(j-i)!} \sum_{l=0}^{j-i} \binom{j-i}{l} (\lambda s)^{l} (\lambda t)^{j-i-l} = e^{-\lambda(s+t)} \frac{1}{(j-i)!} (\lambda s + \lambda t)^{j-i}$$
$$= P_{i,j}(s+t), \qquad s, t \geqslant 0.$$

Infinitesimal Generator

The infinitesimal generator of a continuous-time Markov process allows us to encode all properties of the process $(X_t)_{t \in \mathbb{R}_+}$ in a single matrix.

By differentiating the semigroup relation (9.3.2) with respect to t we get, by componentwise differentiation and assuming a finite state space $\mathbb{S}$,

$$\begin{aligned} P'(t) &= \lim_{h \searrow 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \searrow 0} \frac{P(t)P(h) - P(t)}{h} \\ &= P(t) \lim_{h \searrow 0} \frac{P(h) - P(0)}{h} \\ &= P(t)Q, \end{aligned} \tag{9.3.3}$$

where

$$Q := P'(0) = \lim_{h \searrow 0} \frac{P(h) - P(0)}{h}$$

is called the *infinitesimal generator* of $(X_t)_{t\in\mathbb{R}_+}$.

When $\mathbb{S} = \{0, 1, \ldots, N\}$ we will denote by $\lambda_{i,j}$, $i, j \in \mathbb{S}$, the entries of the infinitesimal generator matrix $Q = (\lambda_{i,j})_{i,j\in\mathbb{S}}$, *i.e.*

$$Q = \frac{dP(t)}{dt}_{|t=0} = [\ \lambda_{i,j}\ ]_{0\leqslant i,j\leqslant N} = \begin{bmatrix} \lambda_{0,0} & \lambda_{0,1} & \lambda_{0,2} & \cdots & \lambda_{0,N} \\ \lambda_{1,0} & \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,N} \\ \lambda_{2,0} & \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{N,0} & \lambda_{N,1} & \lambda_{N,2} & \cdots & \lambda_{N,N} \end{bmatrix}. \qquad (9.3.4)$$

Denoting $Q = [\lambda_{i,j}]_{i,j\in\mathbb{S}}$, for all $i \in \mathbb{S}$ we have

$$\sum_{j\in\mathbb{S}} \lambda_{i,j} = \sum_{j\in\mathbb{S}} P'_{i,j}(0) = \frac{d}{dt}\sum_{j\in\mathbb{S}} P_{i,j}(t)_{|t=0} = \frac{d}{dt}\mathbb{1}_{|t=0} = 0,$$

hence the *rows* of the infinitesimal generator matrix $Q = [\lambda_{i,j}]_{i,j\in\mathbb{S}}$ always add up to 0, *i.e.*

$$\sum_{l\in\mathbb{S}} \lambda_{i,l} = \lambda_{i,i} + \sum_{l\neq i} \lambda_{i,l} = 0,$$

or

$$\lambda_{i,i} = -\sum_{l\neq i} \lambda_{i,l}. \qquad (9.3.5)$$

Note that a state (i) such that $\lambda_{i,j} = 0$ for all $j \in \mathbb{S}$ is absorbing.

Equation (9.3.3), *i.e.*

$$P'(t) = P(t)Q, \qquad t > 0, \qquad (9.3.6)$$

is called the *forward Kolmogorov equation*, cf. (1.2.4). In a similar way we can show the relation

$$P'(t) = QP(t), \qquad t > 0, \qquad (9.3.7)$$

which is called the *backward Kolmogorov equation*.

The forward and backward Kolmogorov equations (9.3.6)-(9.3.7) can be solved either using the matrix exponential e^{tQ} defined as

$$\exp(tQ) := \sum_{n\geqslant 0} \frac{t^n}{n!} Q^n = \mathrm{I} + \sum_{n\geqslant 1} \frac{t^n}{n!} Q^n, \tag{9.3.8}$$

or by viewing the Kolmogorov equations (9.3.6)-(9.3.7) component by component as systems of differential equations.

In (9.3.8) above, $Q^0 = \mathrm{I}$ is the identity matrix, written as

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

when the state space is $\mathbb{S} = \{0, 1, \ldots, N\}$. Using matrix exponentials, the solution of (9.3.6) is given by

$$P(t) = P(0)\exp(tQ) = \exp(tQ), \quad t \geqslant 0.$$

We will often use the first order approximation in $h \to 0$ of

$$P(h) = \exp(hQ) = \mathrm{I} + \sum_{n\geqslant 1} \frac{h^n}{n!} Q^n = \mathrm{I} + hQ + \frac{h^2}{2!} Q^2 + \frac{h^3}{3!} Q^3 + \frac{h^4}{4!} Q^4 + \cdots,$$

given by the first order Taylor expansion

$$P(h) = \mathrm{I} + hQ + o(h), \qquad h \searrow 0, \tag{9.3.9}$$

where $o(h)$ is a function such that $\lim_{h\to 0} o(h)/h = 0$, *i.e.*

$$P(h) = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} + h \begin{bmatrix} \lambda_{0,0} & \lambda_{0,1} & \lambda_{0,2} & \cdots & \lambda_{0,N} \\ \lambda_{1,0} & \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,N} \\ \lambda_{2,0} & \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{N,0} & \lambda_{N,1} & \lambda_{N,2} & \cdots & \lambda_{N,N} \end{bmatrix} + o(h), \quad h \searrow 0.$$

Relation (9.3.9) yields the transition probabilities over a small time interval of length $h > 0$, as:

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = P_{i,j}(h) = \begin{cases} \lambda_{i,j} h + o(h), & i \neq j, \quad h \searrow 0, \\ 1 + \lambda_{i,i} h + o(h), & i = j, \quad h \searrow 0, \end{cases}$$

and by (9.3.5) we also have

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = P_{i,j}(h) = \begin{cases} \lambda_{i,j} h + o(h), & i \neq j, \quad h \searrow 0, \\ \\ 1 - h \sum_{l \neq i} \lambda_{i,l} + o(h), & i = j, \quad h \searrow 0. \end{cases} \tag{9.3.10}$$

For example, in the case of a two-state continuous-time Markov chain we have

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \\ \beta & -\beta \end{bmatrix},$$

with $\alpha, \beta \geqslant 0$, and

$$\begin{aligned} P(h) &= \mathrm{I} + hQ + o(h) \\ &= \begin{bmatrix} 1 & 0 \\ \\ 0 & 1 \end{bmatrix} + h \begin{bmatrix} -\alpha & \alpha \\ \\ \beta & -\beta \end{bmatrix} + o(h) \\ &= \begin{bmatrix} 1 - \alpha h & \alpha h \\ \\ \beta h & 1 - \beta h \end{bmatrix} + o(h), \end{aligned} \tag{9.3.11}$$

as $h \searrow 0$. In this case, $P(h)$ above has the same form as the transition matrix (4.5.1) of a *discrete-time* Markov chain with "small" time step $h > 0$ and "small" transition probabilities $h\alpha$ and $h\beta$, namely $h\alpha$ is the probability of switching from state (0) to state (1), and $h\beta$ is the probability of switching from state (1) to state (0) within a short period of time $h > 0$.

We note that since

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) \simeq \lambda_{i,j} h, \qquad h \searrow 0, \quad i \neq j,$$

and

$$\mathbb{P}(X_{t+h} \neq j \mid X_t = i) = 1 - \lambda_{i,j} h + o(h), \qquad h \searrow 0, \quad i \neq j,$$

the transition of the process $(X_t)_{t \in \mathbb{R}_+}$ from state (i) to state (j) behaves identically to that of a Poisson process with intensity $\lambda_{i,j}$, cf. (9.1.2a)-(9.1.2b) above. Similarly to the Poisson, birth and death processes, the relation

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = \lambda_{i,j} h + o(h), \qquad h \searrow 0, \quad i \neq j,$$

shows that the time $\tau_{i,j}$ spent in state (i) "before moving to state $(j) \neq$ (i)", *i.e.* given the first jump is to state (j), is an *exponentially distributed random variable* with parameter $\lambda_{i,j}$, *i.e.*

$$\mathbb{P}(\tau_{i,j} > t) = \mathrm{e}^{-\lambda_{i,j}t}, \qquad t \geqslant 0, \tag{9.3.12}$$

and we have

$$\mathbb{E}[\tau_{i,j}] = \lambda_{i,j} \int_0^\infty t\mathrm{e}^{-t\lambda_{i,j}} dt = \frac{1}{\lambda_{i,j}}, \qquad i \neq j.$$

When $i = j$ we have

$$\mathbb{P}(X_{t+h} \neq i \mid X_t = i) \simeq h \sum_{l \neq i} \lambda_{i,l} = -\lambda_{i,i} h, \qquad h \searrow 0,$$

and

$$\mathbb{P}(X_{t+h} = i \mid X_t = i) = 1 - h \sum_{l \neq i} \lambda_{i,l} + o(h) = 1 + \lambda_{i,i} h + o(h), \quad h \searrow 0,$$

hence, by the same Poisson process analogy, the time τ_i spent in state $\textcircled{i}$ before the next transition to a different state is an exponentially distributed random variable with parameter $\sum_{j \neq i} \lambda_{i,j}$, *i.e.*

$$\mathbb{P}(\tau_i > t) = \exp\left(-t \sum_{j \neq i} \lambda_{i,j}\right) = \mathrm{e}^{t\lambda_{i,i}}, \qquad t \geqslant 0.$$

In other words, we can also write the time τ_i spent in state $\textcircled{i}$ as

$$\tau_i = \min_{\substack{j \in \mathbb{S} \\ j \neq i}} \tau_{i,j},$$

and this recovers the fact that τ_i is an exponential random variable with parameter $\sum_{j \neq i} \lambda_{i,j}$, since

$$\begin{aligned}
\mathbb{P}(\tau_i > t) &= \mathbb{P}\left(\min_{\substack{j \in \mathbb{S} \\ j \neq i}} \tau_{i,j} > t\right) \\
&= \prod_{\substack{j \in \mathbb{S} \\ j \neq i}} \mathbb{P}(\tau_{i,j} > t) \\
&= \exp\left(-t \sum_{j \neq i} \lambda_{i,j}\right) = \mathrm{e}^{t\lambda_{i,i}}, \qquad t \geqslant 0.
\end{aligned}$$

cf. (1.5.8) in Chapter 1. In addition we have

$$\mathbb{E}[\tau_i] = \sum_{l \neq i} \lambda_{i,l} \int_0^\infty t \exp\left(-t \sum_{l \neq i} \lambda_{i,l}\right) dt = \frac{1}{\sum_{l \neq i} \lambda_{i,l}} = -\frac{1}{\lambda_{i,i}},$$

and the times $(\tau_k)_{k \in \mathbb{S}}$ spent in each state $k \in \mathbb{S}$ form a sequence of independent random variables.

Examples

i) Two-state continuous-time Markov chain.

For the two-state continuous-time Markov chain with generator

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \\ \beta & -\beta \end{bmatrix},$$

the mean time spent at state ⓪ is $1/\alpha$, whereas the mean time spent at state ① is $1/\beta$. We will come back to this example in more detail in the following Section 9.4.

ii) Poisson process.

The generator of the Poisson process is given by $\lambda_{i,j} = \mathbb{1}_{\{j=i+1\}}\lambda$, $i \neq j$, *i.e.*

$$Q = [\ \lambda_{i,j}\ ]_{i,j \in \mathbb{N}} = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ \\ 0 & -\lambda & \lambda & \cdots \\ \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

From the relation $P(h) = \mathrm{I} + hQ + o(h)$ we recover the infinitesimal transition probabilities of the Poisson process as

$$\mathbb{P}(N_{t+h} - N_t = 1) = \mathbb{P}(N_{t+h} = i + 1 \mid N_t = i) \simeq \lambda h,$$

$h \searrow 0$, $i \in \mathbb{N}$, and

$$\mathbb{P}(N_{t+h} - N_t = 0) = \mathbb{P}(N_{t+h} = i \mid N_t = i) = 1 - \lambda h + o(h),$$

$h \searrow 0$, $i \in \mathbb{N}$.

iii) Pure birth process.

The generator of the pure birth process on $\mathbb{N} = \{0, 1, 2, \ldots\}$ is

$$Q = [\ \lambda_{i,j}\ ]_{i,j \geqslant 0} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots \\ \\ 0 & -\lambda_1 & \lambda_1 & \cdots \\ \\ 0 & 0 & -\lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

in which the rate λ_i is (possibly) state-dependent. From the relation

$$P(h) = \mathrm{I} + hQ + o(h), \qquad h \searrow 0,$$

i.e.

$$P(h) = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} + h \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots \\ \\ 0 & -\lambda_1 & \lambda_1 & \cdots \\ \\ 0 & 0 & -\lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} + o(h), \quad h \searrow 0,$$

we recover the infinitesimal transition probabilities of the pure birth process as

$$\mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = i) = \mathbb{P}(X_{t+h} = i + 1 \mid X_t = i) \simeq \lambda_i h,$$

$h \searrow 0$, $i \in \mathbb{N}$, and

$$\mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = i) = \mathbb{P}(X_{t+h} = i \mid X_t = i) = 1 - \lambda_i h + o(h),$$

$h \searrow 0$, $i \in \mathbb{N}$.

iv) Pure death process.

The generator of the pure death process on $-\mathbb{N} = \{\ldots, -2, -1, 0\}$ is

$$Q = [\ \lambda_{i,j}\ ]_{i,j \leqslant 0} = \begin{bmatrix} \cdots & 0 & \mu_0 & -\mu_0 \\ \\ \cdots & \mu_1 & -\mu_1 & 0 \\ \\ \cdots & -\mu_2 & 0 & 0 \\ \iddots & \vdots & \vdots & \vdots \end{bmatrix}.$$

From the relation

$$P(h) = \begin{bmatrix} \cdots & 0 & 0 & 1 \\ \cdots & 0 & 1 & 0 \\ \cdots & 1 & 0 & 0 \\ \iddots & \vdots & \vdots & \vdots \end{bmatrix} + h \begin{bmatrix} \cdots & 0 & \mu_0 & -\mu_0 \\ \cdots & \mu_1 & -\mu_1 & 0 \\ \cdots & -\mu_2 & 0 & 0 \\ \iddots & \vdots & \vdots & \vdots \end{bmatrix} + o(h), \quad h \searrow 0,$$

we recover the infinitesimal transition probabilities

$$\mathbb{P}(X_{t+h} - X_t = -1 \mid X_t = i) = \mathbb{P}(X_{t+h} = i-1 \mid X_t = i) \simeq \mu_i h, \quad h \searrow 0,$$

$i \in \mathbb{S}$, and

$$\mathbb{P}(X_{t+h} = i \mid X_t = i) = \mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = i) = 1 - \mu_i h + o(h), \quad h \searrow 0,$$

$i \in \mathbb{S}$, of the pure death process.

v) Birth and death process on $\{0, 1, \ldots, N\}$.

By (9.2.5a)-(9.2.5b) and (9.3.10), the generator of the birth and death process on $\{0, 1, \ldots, N\}$ is

$$[\,\lambda_{i,j}\,]_{0 \leqslant i,j \leqslant N} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots\cdots & \cdots & 0 & 0 \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & 0 & \cdots\cdots & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \ \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \ \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \ \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots \ 0 & \mu_{N-1} & -\lambda_{N-1}-\mu_{N-1} & \lambda_{N-1} \\ 0 & 0 & 0 & 0 & \cdots \ 0 & 0 & \mu_N & -\mu_N \end{bmatrix},$$

with $\mu_0 = \lambda_N = 0$.

From the Relation (9.3.9) we have

$$P(h) = \mathrm{I} + hQ + o(h), \qquad h \searrow 0,$$

and we recover the infinitesimal transition probabilities

$$\mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = i) \simeq \lambda_i h, \qquad h \searrow 0, \quad i = 0, 1, \ldots, N,$$

and

$$\mathbb{P}(X_{t+h} - X_t = -1 \mid X_t = i) \simeq \mu_i h, \qquad h \searrow 0, \quad i = 0, 1, \ldots, N,$$

and

$$\mathbb{P}(X_{t+h} - X_t = 0 \mid X_t = i) = 1 - (\lambda_i + \mu_i)h + o(h), \qquad h \searrow 0, \quad i = 0, 1, \ldots, N,$$

of the birth and death process on $\{0, 1, \ldots, N\}$, with $\mu_0 = \lambda_N = 0$.

Recall that the time τ_i spent in state (i) is an exponentially distributed random variable with parameter $\lambda_i + \mu_i$ and we have

$$P_{i,i}(t) \geqslant \mathbb{P}(\tau_i > t) = \mathrm{e}^{-t(\lambda_i+\mu_i)}, \qquad t \geqslant 0,$$

and

$$\mathbb{E}[\tau_i] = \frac{1}{\lambda_i + \mu_i}.$$

In the case of a pure birth process we find

$$P_{i,i}(t) = \mathbb{P}(\tau_i > t) = \mathrm{e}^{-t\lambda_i}, \qquad t \geqslant 0,$$

and similarly for a pure death process. This allows us in particular to compute the diagonal entries in the matrix exponential $P(t) = \exp(tQ)$, $t \geqslant 0$.

When $\mathbb{S} = \{0, 1, \ldots, N\}$ with $\lambda_i = \lambda$ and $\mu_i = \mu$, $i = 1, 2, \ldots, N-1$, and $\lambda_0 = \mu_N = 0$, the above birth and death process becomes a continuous-time analog of the discrete-time gambling process.

9.4 Two-State Continuous-Time Markov Chain

In this section we consider a continuous-time Markov process with state space $\mathbb{S} = \{0, 1\}$, in the same way as in Section 4.5 which covered the two-state Markov chain in discrete time. This continuous-time Markov chain is also called a *telegraph process*.

In this case the infinitesimal generator Q of $(X_t)_{t\in\mathbb{R}_+}$ has the form

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}, \tag{9.4.1}$$

with $\alpha, \beta \geqslant 0$. The forward Kolmogorov equation (9.3.6) reads

$$P'(t) = P(t) \times \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}, \qquad t > 0, \tag{9.4.2}$$

i.e.

$$\begin{bmatrix} P'_{0,0}(t) & P'_{0,1}(t) \\ P'_{1,0}(t) & P'_{1,1}(t) \end{bmatrix} = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) \\ P_{1,0}(t) & P_{1,1}(t) \end{bmatrix} \times \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}, \qquad t > 0,$$

or

$$\begin{cases} P'_{0,0}(t) = -\alpha P_{0,0}(t) + \beta P_{0,1}(t), \ P'_{0,1}(t) = \alpha P_{0,0}(t) - \beta P_{0,1}(t), \\ \\ P'_{1,0}(t) = -\alpha P_{1,0}(t) + \beta P_{1,1}(t), \ P'_{1,1}(t) = \alpha P_{1,0}(t) - \beta P_{1,1}(t), \end{cases}$$

$t > 0$, which is a system of four differential equations * with initial condition

$$P(0) = \begin{bmatrix} P_{0,0}(0) & P_{0,1}(0) \\ \\ P_{1,0}(0) & P_{1,1}(0) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \\ 0 & 1 \end{bmatrix} = \mathrm{I}.$$

The solution of the forward Kolmogorov equation (9.4.2) is given by the matrix exponential

$$P(t) = P(0)\exp(tQ) = \exp(tQ) = \exp\left(t \begin{bmatrix} -\alpha & \alpha \\ \\ \beta & -\beta \end{bmatrix} \right),$$

which is computed in the next Proposition 9.6, see also the command

MatrixExp[t*-a,a,b,-b].

Proposition 9.6. *The solution $P(t)$ of the forward Kolmogorov equation (9.4.2) is given by*

$$P(t) = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) \\ \\ P_{1,0}(t) & P_{1,1}(t) \end{bmatrix} \tag{9.4.3}$$
$$= \begin{bmatrix} \dfrac{\beta}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} & \dfrac{\alpha}{\alpha+\beta} - \dfrac{\alpha}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} \\ \\ \dfrac{\beta}{\alpha+\beta} - \dfrac{\beta}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} & \dfrac{\alpha}{\alpha+\beta} + \dfrac{\beta}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} \end{bmatrix},$$

$t \geqslant 0$.

Proof. We will compute the matrix exponential e^{tQ} by the diagonalization technique. The matrix Q has two eigenvectors†

$$\begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -\alpha \\ \\ \beta \end{bmatrix},$$

* Please refer to MH3110 - Ordinary Differential Equations for more on first order linear systems of differential equations.

† Please refer to MH1201 - Linear Algebra II for more on eigenvectors, eigenvalues, and diagonalization.

with respective eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -\alpha - \beta$.* Hence it can be transformed into the *diagonal form*

$$Q = M \times D \times M^{-1}$$

as follows:

$$Q = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix} \times \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \times \begin{bmatrix} \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta} \\ -\dfrac{1}{\alpha+\beta} & \dfrac{1}{\alpha+\beta} \end{bmatrix}.$$

Consequently we have

$$\begin{aligned}
P(t) &= \exp(tQ) = \sum_{n\geqslant 0} \frac{t^n}{n!} Q^n = \sum_{n\geqslant 0} \frac{t^n}{n!} (M \times D \times M^{-1})^n \\
&= \sum_{n\geqslant 0} \frac{t^n}{n!} M \times D^n \times M^{-1} = M \times \left(\sum_{n\geqslant 0} \frac{t^n}{n!} D^n \right) \times M^{-1} \\
&= M \times \exp(tD) \times M^{-1} \\
&= \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix} \times \begin{bmatrix} \mathrm{e}^{t\lambda_1} & 0 \\ 0 & \mathrm{e}^{t\lambda_2} \end{bmatrix} \times \begin{bmatrix} \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta} \\ -\dfrac{1}{\alpha+\beta} & \dfrac{1}{\alpha+\beta} \end{bmatrix} \\
&= \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{e}^{-t(\alpha+\beta)} \end{bmatrix} \times \begin{bmatrix} \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta} \\ -\dfrac{1}{\alpha+\beta} & \dfrac{1}{\alpha+\beta} \end{bmatrix} \\
&= \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{\mathrm{e}^{-t(\alpha+\beta)}}{\alpha+\beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\beta}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} & \dfrac{\alpha}{\alpha+\beta} - \dfrac{\alpha}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} \\ \dfrac{\beta}{\alpha+\beta} - \dfrac{\beta}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} & \dfrac{\alpha}{\alpha+\beta} + \dfrac{\beta}{\alpha+\beta}\mathrm{e}^{-t(\alpha+\beta)} \end{bmatrix},
\end{aligned}$$

$t > 0$. □

From Proposition 9.6 we obtain the probabilities

* See for example the command Eigensystem[-a,a,b,-b].

$$\begin{cases} \mathbb{P}(X_h = 0 \mid X_0 = 0) = \dfrac{\beta + \alpha \mathrm{e}^{-(\alpha+\beta)h}}{\alpha+\beta}, \\ \\ \mathbb{P}(X_h = 1 \mid X_0 = 0) = \dfrac{\alpha}{\alpha+\beta}(1 - \mathrm{e}^{-(\alpha+\beta)h}), \\ \\ \mathbb{P}(X_h = 0 \mid X_0 = 1) = \dfrac{\beta}{\alpha+\beta}(1 - e^{-h(\alpha+\beta)h}), \\ \\ \mathbb{P}(X_h = 1 \mid X_0 = 1) = \dfrac{\alpha + \beta \mathrm{e}^{-(\alpha+\beta)h}}{\alpha+\beta}, \qquad h \geqslant 0. \end{cases}$$

In other words, (9.4.3) can be rewritten as

$$P(h) = \begin{bmatrix} 1 - \dfrac{\alpha}{\alpha+\beta}(1 - \mathrm{e}^{-(\alpha+\beta)h}) & \dfrac{\alpha}{\alpha+\beta}(1 - \mathrm{e}^{-(\alpha+\beta)h}) \\ \\ \dfrac{\beta}{\alpha+\beta}(1 - \mathrm{e}^{-(\alpha+\beta)h}) & 1 - \dfrac{\beta}{\alpha+\beta}(1 - \mathrm{e}^{-(\alpha+\beta)h}) \end{bmatrix}, \qquad h > 0, \tag{9.4.4}$$

hence, since

$$1 - \mathrm{e}^{-(\alpha+\beta)h} \simeq h(\alpha+\beta), \qquad h \searrow 0,$$

the expression (9.4.4) above recovers (9.3.11) as $h \searrow 0$, *i.e.* we have

$$P(h) = \begin{bmatrix} 1 - h\alpha & h\alpha \\ \\ h\beta & 1 - h\beta \end{bmatrix} + o(h) = \mathrm{I} + hQ + o(h), \qquad h \searrow 0,$$

which recovers (9.4.1).

From these expressions we can determine the large time behavior of the continuous-time Markov chain by taking limits as t goes to infinity:

$$\lim_{t\to\infty} P(t) = \lim_{t\to\infty} \begin{bmatrix} \mathbb{P}(X_t = 0 \mid X_0 = 0) & \mathbb{P}(X_t = 1 \mid X_0 = 0) \\ \\ \mathbb{P}(X_t = 0 \mid X_0 = 1) & \mathbb{P}(X_t = 1 \mid X_0 = 1) \end{bmatrix} = \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \\ \beta & \alpha \end{bmatrix},$$

whenever $\alpha > 0$ or $\beta > 0$, whereas if $\alpha = \beta = 0$ we simply have

$$P(t) = \begin{bmatrix} 1 & 0 \\ \\ 0 & 1 \end{bmatrix} = \mathrm{I}, \qquad t \geqslant 0,$$

and the chain is constant. Note that in continuous time the limiting distribution of the two-state chain always exists (unlike in the discrete-time case), and convergence will be faster when $\alpha + \beta$ is larger. Hence we have

$$\lim_{t\to\infty} \mathbb{P}(X_t = 1 \mid X_0 = 0) = \lim_{t\to\infty} \mathbb{P}(X_t = 1 \mid X_0 = 1) = \frac{\alpha}{\alpha+\beta}$$

and

$$\lim_{t\to\infty} \mathbb{P}(X_t = 0 \mid X_0 = 0) = \lim_{t\to\infty} \mathbb{P}(X_t = 0 \mid X_0 = 1) = \frac{\beta}{\alpha+\beta}$$

and

$$\pi = [\pi_0, \pi_1] = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) = \left(\frac{1/\alpha}{1/\alpha+1/\beta}, \frac{1/\beta}{1/\alpha+1/\beta}\right)$$

appears as a *limiting distribution* as t goes to infinity, provided that $(\alpha, \beta) \neq (0, 0)$. This means that whatever the starting point X_0, the probability of being at ① after a "large" time is close to $\alpha/(\alpha+\beta)$, while the probability of being at ⓪ is close to $\beta/(\alpha+\beta)$.

Next, we consider a simulation of the two-state continuous Markov chain with infinitesimal generator

$$Q = \begin{bmatrix} -20 & 20 \\ \\ 40 & -40 \end{bmatrix},$$

i.e. $\alpha = 20$ and $\beta = 40$. Figure 9.5 represents a sample path $(x_t)_{t\in\mathbb{R}_+}$ of the continuous-time chain, while Figure 9.6 represents the sample average

$$y_t = \frac{1}{t}\int_0^t x_s ds, \qquad t \in [0, 1],$$

which counts the proportion of values of the chain in the state ①. This proportion is found to converge to $\alpha/(\alpha+\beta) = 1/3$, as a consequence of the Ergodic Theorem in continuous time, see (9.5.4) below.*

* Download the corresponding R **code** or the **IPython notebook** that can be run ***here***.

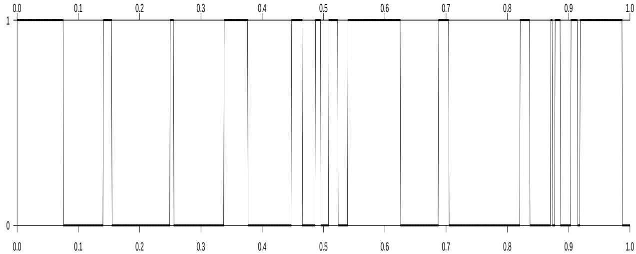


Fig. 9.5: Sample path of a continuous-time two-state chain with $\alpha = 20$ and $\beta = 40$.

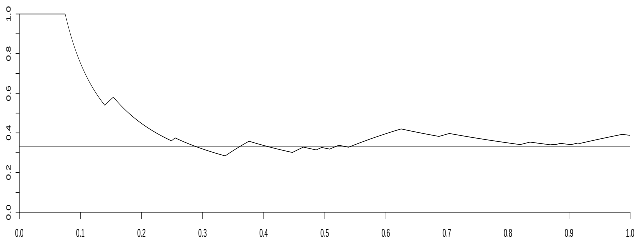


Fig. 9.6: The proportion of process values in the state 1 converges to $1/3 = \alpha/(\alpha+\beta)$.

9.5 Limiting and Stationary Distributions

A probability distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ on $\mathbb{S}$ is said to be *stationary* for $P(t)$ if it satisfies the equation

$$\pi P(t) = \pi, \qquad t \geqslant 0.$$

In the next proposition, we show that the notion of stationary distribution admits a simpler characterization.

Proposition 9.7. *The probability distribution $\pi = (\pi_i)_{i\in\mathbb{S}}$ is stationary if and only if it satisfies the equation*

$$\pi Q = 0.$$

Proof. Assuming that $\pi Q = 0$, we have

$$\begin{aligned} \pi P(t) &= \pi \exp(tQ) = \pi \sum_{n \geqslant 0} \frac{t^n}{n!} Q^n \\ &= \pi \left(\mathrm{I} + \sum_{n \geqslant 1} \frac{t^n}{n!} Q^n \right) \\ &= \pi + \sum_{n \geqslant 1} \frac{t^n}{n!} \pi Q^n \\ &= \pi, \end{aligned}$$

since $\pi Q^n = \pi Q Q^{n-1} = 0$, $n \geqslant 1$. Conversely, the relation $\pi = \pi P(t)$ shows, by differentiation at $t = 0$, that

$$0 = \pi P'(0) = \pi Q.$$

□

When $\mathbb{S} = \{0, 1, \ldots, N\}$ and the generator Q has the form (9.3.4) the relation $\pi Q = 0$ reads

$$\pi_0 \lambda_{0,j} + \pi_1 \lambda_{1,j} + \cdots + \pi_N \lambda_{N,j} = 0, \qquad j = 0, 1, \ldots, N,$$

i.e.,

$$\sum_{\substack{i=0 \\ i \neq j}}^{N} \pi_i \lambda_{j,i} = -\pi_j \lambda_{j,j}, \qquad j = 0, 1, \ldots, N,$$

hence from (9.3.5) we find the balance condition

$$\sum_{\substack{i=0 \\ i \neq j}}^{N} \pi_i \lambda_{i,j} = \sum_{\substack{k=0 \\ k \neq j}}^{N} \pi_j \lambda_{j,k},$$

which can be interpreted by stating the equality between incoming and outgoing "flows" into and from state $\textcircled{j}$ are equal for all $j = 0, 1, \ldots, N$.

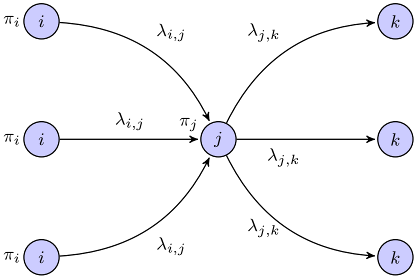


Fig. 9.7: Global balance condition (continuous time).

Next is the continuous-time analog of Proposition 7.7 in Section 7.2.

Proposition 9.8. *Consider a continuous-time Markov chain $(X_t)_{t\in\mathbb{R}_+}$ on a* finite state space $\mathbb{S}$, *which admits a limiting distribution given by*

$$\pi_j := \lim_{t\to\infty} \mathbb{P}(X_t = j \mid X_0 = i) = \lim_{t\to\infty} P_{i,j}(t), \qquad j \in \mathbb{S}, \tag{9.5.1}$$

independent of the initial state $i \in \mathbb{S}$. Then we have

$$\pi Q = 0, \tag{9.5.2}$$

i.e. π is a stationary distribution for the chain $(X_t)_{t\in\mathbb{R}_+}$.

Proof. Taking $\mathbb{S} = \{0, 1, \ldots, N\}$ we note that by (9.5.1) since the limiting distribution is independent of the initial state it satisfies

$$\lim_{t\to\infty} P(t) = \begin{bmatrix} \lim_{t\to\infty} P_{0,0}(t) & \cdots & \lim_{t\to\infty} P_{0,N}(t) \\ \vdots & \ddots & \vdots \\ \lim_{t\to\infty} P_{N,0}(t) & \cdots & \lim_{t\to\infty} P_{N,N}(t) \end{bmatrix}$$
$$= \begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_N \\ \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \cdots & \pi_N \end{bmatrix} = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix},$$

where π is the row vector

$$\pi = [\pi_0, \pi_1, \ldots, \pi_n].$$

By the forward Kolmogorov equation (9.3.6) and (9.5.1) we find that the limit of $P'(t)$ exists as $t \to \infty$ since

$$\lim_{t\to\infty} P'(t) = \lim_{t\to\infty} P(t)Q = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix} Q.$$

On the other hand, since $P'(t)$ converges as $t \to \infty$ we should have

$$\lim_{t\to\infty} P'(t) = 0,$$

for the integral

$$P(t) = P(0) + \int_0^t P'(s)ds \tag{9.5.3}$$

to converge as $t \to \infty$. This shows that

$$\begin{bmatrix} \pi Q \\ \vdots \\ \pi Q \end{bmatrix} = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix} Q = 0$$

by (9.3.6), hence we have $\pi Q = 0$, or $\sum_{i\in\mathbb{S}} \pi_i \lambda_{i,j} = 0$, $j \in \mathbb{S}$. □

Equation (9.5.2) is actually equivalent to

$$\pi = \pi(\mathrm{I} + hQ), \qquad h > 0,$$

which yields the stationary distribution of a discrete-time Markov chain with transition matrix $P(h) = \mathrm{I} + hQ + o(h)$ on "small" discrete intervals of length $h \searrow 0$.

Proposition 9.8 admits the following partial converse. More generally it can be shown, cf. Corollary 6.2 of Lalley (2013), that an irreducible continuous-time Markov chain admits its stationary distribution π as limiting distribution, similarly to the discrete-time Theorem 7.8, cf. also Proposition 1.1 page 9 of Sigman (2014) in the positive recurrent case.

Proposition 9.9. *Consider an irreducible, continuous-time Markov chain* $(X_t)_{t\in\mathbb{R}_+}$ *on a* finite state space $\mathbb{S}$, *with stationary distribution* π, *i.e.*

$$\pi Q = 0,$$

and assume that the matrix Q *is diagonalizable. Then* $(X_t)_{t\in\mathbb{R}_+}$ *admits* π *as limiting distribution, i.e.*

$$\lim_{t\to\infty} P_{i,j}(t) = \lim_{t\to\infty} \mathbb{P}(X_t = j \mid X_0 = i) = \pi_j, \quad j \in \mathbb{S},$$

which is independent of the initial state $i \in \mathbb{S}$.

Proof. By *e.g.* Theorem 2.1 in Chapter 10 of Karlin and Taylor (1981), since the chain is irreducible, $\lambda_1 = 0$ is an eigenvalue of Q with multiplicity one and eigenvector $u^{(1)} = (1, 1, \ldots, 1)$. In addition, the remaining eigenvectors $u^{(2)}, \ldots, u^{(n)} \in \mathbb{R}^n$ with eigenvalues $\lambda_2, \ldots, \lambda_n$ are orthogonal to the invariant (or stationary) distribution $[\pi_1, \pi_2, \ldots, \pi_n]$ of $(X_t)_{t\in\mathbb{R}_+}$ as we have $\lambda_k \langle u^{(k)}, \pi\rangle_{\mathbb{R}^n} = \langle Qu^{(k)}, \pi\rangle_{\mathbb{R}^n} = \langle u^{(k)}, Q^\top \pi\rangle_{\mathbb{R}^n} = \pi^\top Q u^{(k)} = 0$, $k = 2, \ldots, n$. Hence by diagonalization we have $Q = M^{-1}DM$ where the matrices M and M^{-1} take the form

$$M = \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \quad \text{and} \quad M^{-1} = \begin{bmatrix} 1 & u_1^{(2)} & \cdots & u_1^{(n)} \\ 1 & u_2^{(2)} & \cdots & u_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_n^{(2)} & \cdots & u_n^{(n)} \end{bmatrix},$$

and D is the diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. This allows us to compute the transition probabilities of $(X_t)_{t\in\mathbb{R}_+}$ as

$$P_{i,j}(t) = \mathbb{P}(X_t = j \mid X_0 = i) = [\exp(tQ)]_{i,j} = [M^{-1}\exp(tD)M]_{i,j}$$

where $\exp(tD)$ is the diagonal matrix

$$\exp(tD) = \mathrm{diag}(1, \mathrm{e}^{t\lambda_2}, \ldots, \mathrm{e}^{t\lambda_n}),$$

and the eigenvalues $\lambda_2, \ldots, \lambda_n$ have to be strictly negative, hence we have

$$\begin{aligned} \lim_{t\to\infty} [\mathbb{P}(X_t = j \mid X_0 = i)]_{1\leqslant i,j\leqslant n} &= \lim_{t\to\infty} [M^{-1}\exp(tD)M]_{1\leqslant i,j\leqslant n} \\ &= \begin{bmatrix} 1 & u_1^{(2)} & \cdots & u_1^{(n)} \\ 1 & u_2^{(2)} & \cdots & u_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_n^{(2)} & \cdots & u_n^{(n)} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \\ &= \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \pi_1 & \cdots & \pi_n \\ \vdots & \ddots & \vdots \\ \pi_1 & \cdots & \pi_n \end{bmatrix}. \end{aligned}$$

□

The discrete-time Ergodic Theorem 7.12 also admits a continuous-time version with a similar proof, stating that if the chain $(X_t)_{t\in\mathbb{R}_+}$ is *irreducible* then the sample average of the number of visits to state (i) converges almost surely to π_i, *i.e.*,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \mathbb{1}_{\{X_t=i\}}dt = \pi_i, \qquad i\in\mathbb{S}. \tag{9.5.4}$$

Examples

i) Two-state Markov chain.

Consider a two-state Markov chain with infinitesimal generator

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \\ \beta & -\beta \end{bmatrix},$$

the limiting distribution solves $\pi Q = 0$, *i.e.*

$$\begin{cases} 0 = -\alpha\pi_0 + \beta\pi_1 \\ \\ 0 = \alpha\pi_0 - \beta\pi_1, \end{cases}$$

with $\pi_0 + \pi_1 = 1$, *i.e.*

$$\pi = [\pi_0, \pi_1] = \left[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right]. \tag{9.5.5}$$

ii) Birth and death process on $\mathbb{N}$.

Next, consider the birth and death process on $\mathbb{N}$ with infinitesimal generator

$$Q = [\ \lambda_{i,j}\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -\lambda_2-\mu_2 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -\lambda_3-\mu_3 & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix},$$

the stationary distribution solves $\pi Q = 0$, *i.e.*

$$\begin{cases} 0 = -\lambda_0\pi_0 + \mu_1\pi_1 \\ \\ 0 = \lambda_0\pi_0 - (\lambda_1+\mu_1)\pi_1 + \mu_2\pi_2 \\ \\ 0 = \lambda_1\pi_1 - (\lambda_2+\mu_2)\pi_2 + \mu_3\pi_3 \\ \quad \vdots \\ 0 = \lambda_{j-1}\pi_{j-1} - (\lambda_j+\mu_j)\pi_j + \mu_{j+1}\pi_{j+1}, \\ \quad \vdots \end{cases}$$

i.e.

$$\begin{cases} \pi_1 = \dfrac{\lambda_0}{\mu_1}\pi_0 \\ \\ \pi_2 = -\dfrac{\lambda_0}{\mu_2}\pi_0 + \dfrac{\lambda_1+\mu_1}{\mu_2}\pi_1 = -\dfrac{\lambda_0}{\mu_2}\pi_0 + \dfrac{\lambda_1+\mu_1}{\mu_2}\dfrac{\lambda_0}{\mu_1}\pi_0 = \dfrac{\lambda_1}{\mu_2}\dfrac{\lambda_0}{\mu_1}\pi_0 \\ \\ \pi_3 = -\dfrac{\lambda_1}{\mu_3}\pi_1 + \dfrac{\lambda_2+\mu_2}{\mu_3}\pi_2 = -\dfrac{\lambda_1}{\mu_3}\dfrac{\lambda_0}{\mu_1}\pi_0 + \dfrac{\lambda_2+\mu_2}{\mu_3}\dfrac{\lambda_1}{\mu_2}\dfrac{\lambda_0}{\mu_1}\pi_0 = \dfrac{\lambda_2}{\mu_3}\dfrac{\lambda_1}{\mu_2}\dfrac{\lambda_0}{\mu_1}\pi_0 \\ \quad \vdots \\ \pi_{j+1} = \dfrac{\lambda_j\cdots\lambda_0}{\mu_{j+1}\cdots\mu_1}\pi_0. \\ \quad \vdots \end{cases}$$

Using the convention

$$\lambda_{j-1}\cdots\lambda_0 = \prod_{l=0}^{j-1}\lambda_l = 1 \quad \text{and} \quad \mu_j\cdots\mu_1 = \prod_{l=1}^{j}\mu_l = 1$$

in the case $j = 0$, we have

$$\begin{aligned} 1 = \sum_{j\geqslant 0}\pi_j &= \pi_0 + \pi_0\sum_{j\geqslant 0}\frac{\lambda_j\cdots\lambda_0}{\mu_{j+1}\cdots\mu_1} = \pi_0 + \pi_0\sum_{j\geqslant 1}\frac{\lambda_{j-1}\cdots\lambda_0}{\mu_j\cdots\mu_1} \\ &= \pi_0\sum_{i\geqslant 0}\frac{\lambda_0\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}, \end{aligned}$$

hence

$$\pi_0 = \frac{1}{\displaystyle\sum_{i\geqslant 0}\frac{\lambda_0\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}},$$

and

$$\pi_j = \frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j \sum_{i \geqslant 0} \frac{\lambda_0 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i}}, \qquad j \in \mathbb{N}.$$

When $\lambda_i = \lambda$, $i \in \mathbb{N}$, and $\mu_i = \mu$, $i \geqslant 1$, this gives

$$\pi_j = \frac{\lambda^j}{\mu^j \sum_{i \geqslant 0} (\lambda/\mu)^i} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^j, \qquad j \in \mathbb{N}.$$

provided that $\lambda < \mu$, hence in this case the stationary distribution is the geometric distribution with parameter λ/μ, otherwise the stationary distribution does not exist.

iii) Birth and death process on $\mathbb{S} = \{0, 1, \ldots, N\}$.

The birth and death process on $\mathbb{S} = \{0, 1, \ldots, N\}$ has the infinitesimal generator

$$Q = [\ \lambda_{i,j}\ ]_{0 \leqslant i,j \leqslant N} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots\cdots & 0 & 0 & 0 & 0 \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & \cdots\cdots & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots\ \ddots\ \vdots & & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots\ \ddots\ \ddots & & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots\ \ddots\ \ddots & & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots\ \vdots\ \ddots & & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0\ \vdots\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots\cdots & 0 & \mu_{N-1} & -\lambda_{N-1} - \mu_{N-1} & \lambda_{N-1} \\ 0 & 0 & 0 & \cdots\cdots & 0 & 0 & \mu_N & -\mu_N \end{bmatrix},$$

we can apply (1.6.6) with $\lambda_j = 0$, $j \geqslant N$, and $\mu_j = 0$, $j \geqslant N+1$, which yields

$$\pi_j = \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1 \sum_{i=0}^{N} \frac{\lambda_{i-1} \cdots \lambda_0}{\mu_i \cdots \mu_1}}, \qquad j \in \{0, 1, \ldots, N\},$$

and coincides with (9.5.5) when $N = 1$.

When $\lambda_i = \lambda$, $i \in \mathbb{N}$, and $\mu_i = \mu$, $i \geqslant 1$, this gives

$$\pi_j = \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{N+1}} \left(\frac{\lambda}{\mu}\right)^j, \qquad j = 0, 1, \ldots, N,$$

which is a truncated geometric distribution since $\pi_j = 0$ for all $j > N$ and any $\lambda, \mu > 0$.

9.6 Discrete-Time Embedded Chain

Consider the sequence $(T_n)_{n\geqslant 0}$ the sequence of jump times of the continuous-time Markov process $(X_t)_{t\in\mathbb{R}_+}$, defined recursively by $T_0 = 0$, then

$$T_1 = \inf\{t > 0 \ : \ X_t \neq X_0\},$$

and

$$T_{n+1} = \inf\{t > T_n \ : \ X_t \neq X_{T_n}\}, \qquad n \geqslant 0.$$

The *embedded chain* of $(X_t)_{t\in\mathbb{R}_+}$ is the *discrete-time* Markov chain $(Z_n)_{n\geqslant 0}$ defined by $Z_0 = X_0$ and

$$Z_n := X_{T_n}, \qquad n \geqslant 1.$$

The next Figure 9.8 shows the graph of the embedded chain of a birth and death process.*

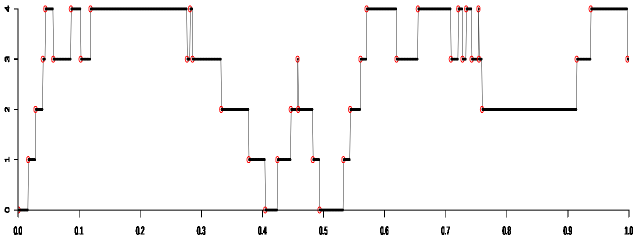


Fig. 9.8: Birth and death process $(X_t)_{t\in\mathbb{R}_+}$ with its embedded chain $(Z_n)_{n\geqslant 0}$.

The results of Chapters 2-8 can now be applied to the discrete-time embedded chain. The next Figure 9.9 represents the discrete-time embedded chain associated to the path of Figure 9.8, in which we have $Z_0 = 0$, $Z_1 = 1$, $Z_2 = 2$, $Z_3 = 3$, $Z_4 = 4$, $Z_5 = 3$, $Z_6 = 4$, $Z_7 = 3$, ...

* Download the corresponding **code** or the **IPython notebook** that can be run ***here***.

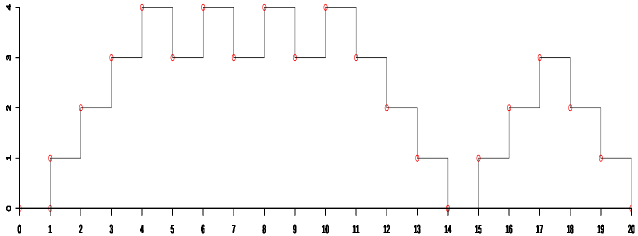


Fig. 9.9: Discrete-time embedded chain $(Z_n)_{n\geqslant 0}$ based on the path of Figure 9.8.

For example, if $\lambda_0 > 0$ and $\mu_1 > 0$, the embedded chain of the two-state continuous-time Markov chain has the transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{9.6.1}$$

which switches permanently between the states ⓪ and ①.

In case one of the states $\{0,1\}$ is absorbing, the transition matrix becomes

$$P = \begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \text{ if } \lambda_0 = 0, \ \mu_1 > 0, \\ \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \text{ if } \lambda_0 > 0, \ \mu_1 = 0, \\ \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ if } \lambda_0 = 0, \ \mu_1 = 0. \end{cases}$$

Birth and death embedded chain

More generally, consider now the birth and death process with infinitesimal generator

$$Q = [\ \lambda_{i,j}\ ]_{0\leqslant i,j\leqslant N} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots\cdots & 0 & 0 & 0 & 0 \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & \cdots\cdots & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots\cdots & 0 & \mu_{N-1} & -\lambda_{N-1}-\mu_{N-1} & \lambda_{N-1} \\ 0 & 0 & 0 & \cdots\cdots & 0 & 0 & \mu_N & -\mu_N \end{bmatrix}.$$

Given that a transition occurs from state (i) in a "short" time interval $[t, t+h]$, the probability that the chain switches to state $\boxed{i+1}$ is given by

$$\begin{aligned}
\mathbb{P}(X_{t+h} = i+1 \mid X_t = i \text{ and } X_{t+h} - X_t \neq 0) &= \frac{\mathbb{P}(X_{t+h} = i+1 \text{ and } X_t = i)}{\mathbb{P}(X_{t+h} - X_t \neq 0 \text{ and } X_t = i)} \\
&= \frac{\mathbb{P}(X_{t+h} = i+1 \text{ and } X_t = i)}{\mathbb{P}(X_{t+h} - X_t \neq 0 \mid X_t = i)\mathbb{P}(X_t = i)} \\
&= \frac{\mathbb{P}(X_{t+h} = i+1 \mid X_t = i)}{\mathbb{P}(X_{t+h} - X_t \neq 0 \mid X_t = i)} \\
&= \frac{\mathbb{P}(X_{t+h} - X_t = 1 \mid X_t = i)}{\mathbb{P}(X_{t+h} - X_t \neq 0 \mid X_t = i)} \\
&\simeq \frac{\lambda_i h}{\lambda_i h + \mu_i h} \\
&= \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad h \searrow 0, \quad i \in \mathbb{S},
\end{aligned}$$

where we applied (9.3.10), hence the transition matrix of the embedded chain satisfies

$$P_{i,i+1} = \lim_{h\searrow 0} \mathbb{P}(X_{t+h} = i+1 \mid X_t = i \text{ and } X_{t+h} - X_t \neq 0) = \frac{\lambda_i}{\lambda_i + \mu_i}, \quad i \in \mathbb{S}. \tag{9.6.2}$$

This result can also be obtained from (1.5.9) which states that

$$\mathbb{P}(\tau_{i,i+1} < \tau_{i,i-1}) = \frac{\lambda_i}{\lambda_i + \mu_i}. \tag{9.6.3}$$

Similarly the probability that a given transition occurs from (i) to $\boxed{i-1}$ is

$$\mathbb{P}(X_{t+h} = i-1 \mid X_t = i \text{ and } X_{t+h} - X_t \neq 0) = \frac{\mu_i}{\lambda_i + \mu_i}, \qquad h \searrow 0, \quad i \in \mathbb{S},$$

which can also be obtained from (1.5.9) which states that

$$\mathbb{P}(\tau_{i,i-1} < \tau_{i,i+1}) = \frac{\mu_i}{\lambda_i + \mu_i}.$$

Hence we have

$$P_{i,i-1} = \lim_{h \searrow 0} \mathbb{P}(X_{t+h} = i-1 \mid X_t = i \text{ and } X_{t+h} - X_t \neq 0) = \frac{\mu_i}{\lambda_i + \mu_i}, \quad i \in \mathbb{S},$$

and the embedded chain $(Z_n)_{n \geqslant 0}$ has the transition matrix

$$P = \big[P_{i,j} \big]_{i,j \in \mathbb{S}}$$

$$= \left[\begin{array}{cccccccccc} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \frac{\mu_1}{\lambda_1+\mu_1} & 0 & \frac{\lambda_1}{\lambda_1+\mu_1} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{\mu_{N-1}}{\lambda_{N-1}+\mu_{N-1}} & 0 & \frac{\lambda_{N-1}}{\lambda_{N-1}+\mu_{N-1}} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{array} \right],$$

provided that $\lambda_0 > 0$ and $\mu_N > 0$. When $N = 1$, this coincides with (9.6.1). In case $\lambda_0 = \mu_N = 0$, states $\textcircled{0}$ and $\textcircled{N}$ are both absorbing since the birth rate starting from $\textcircled{0}$ and the death rate starting from $\textcircled{N}$ are both 0, hence the transition matrix of the embedded chain can be written as

$$P = \big[P_{i,j} \big]_{0 \leqslant i,j \leqslant N}$$

$$= \left[\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \frac{\mu_1}{\lambda_1+\mu_1} & 0 & \frac{\lambda_1}{\lambda_1+\mu_1} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{\mu_{N-1}}{\lambda_{N-1}+\mu_{N-1}} & 0 & \frac{\lambda_{N-1}}{\lambda_{N-1}+\mu_{N-1}} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{array} \right],$$

which is the transition matrix of a gambling type process on $\{0, 1, \ldots, N\}$. When $N = 1$ this yields $P = \mathrm{I}$, which is consistent with the fact that a two-state Markov chain with two absorbing states is constant.

For example, for a continuous-time chain with infinitesimal generator

$$Q = [\ \lambda_{i,j}\ ]_{0 \leqslant i,j \leqslant 4} = \begin{bmatrix} -10 & 10 & 0 & 0 & 0 \\ 10 & -20 & 10 & 0 & 0 \\ 0 & 10 & -30 & 20 & 0 \\ 0 & 0 & 10 & -40 & 30 \\ 0 & 0 & 0 & 20 & -20 \end{bmatrix},$$

the transition matrix of the embedded chain is

$$P = \left[\, P_{i,j} \,\right]_{0 \leqslant i,j \leqslant 4} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 1/4 & 0 & 3/4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

In case the states ⓪ and ④ are absorbing, *i.e.*

$$Q = [\ \lambda_{i,j}\ ]_{0 \leqslant i,j \leqslant 4} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 10 & -20 & 10 & 0 & 0 \\ 0 & 10 & -30 & 20 & 0 \\ 0 & 0 & 10 & -40 & 30 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

the transition matrix of the embedded chain becomes

$$P = \left[\, P_{i,j} \,\right]_{0 \leqslant i,j \leqslant 4} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 1/4 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

More generally, by (9.3.10) and (9.3.5) we could also show that the embedded chain of a continuous-time Markov chain with generator Q of the form (9.3.4) has the transition matrix

$$P = \left[\, P_{i,j} \,\right]_{i,j \in \mathbb{S}}$$

$$= \begin{bmatrix} 0 & -\frac{\lambda_{0,1}}{\lambda_{0,0}} & -\frac{\lambda_{0,2}}{\lambda_{0,0}} & \cdots\cdots & -\frac{\lambda_{0,N-1}}{\lambda_{0,0}} & -\frac{\lambda_{0,N}}{\lambda_{0,0}} \\ -\frac{\lambda_{1,0}}{\lambda_{1,1}} & 0 & -\frac{\lambda_{1,2}}{\lambda_{1,1}} & \cdots\cdots & -\frac{\lambda_{1,N-1}}{\lambda_{1,1}} & -\frac{\lambda_{1,N}}{\lambda_{1,1}} \\ \vdots & \ddots & \vdots & \vdots\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots\ \vdots & \ddots & \vdots \\ -\frac{\lambda_{N-1,0}}{\lambda_{N-1,N-1}} & -\frac{\lambda_{N-1,1}}{\lambda_{N-1,N-1}} & \cdots & \cdots\cdots & 0 & -\frac{\lambda_{N-1,N}}{\lambda_{N-1,N-1}} \\ -\frac{\lambda_{N,0}}{\lambda_{N,N}} & -\frac{\lambda_{N,1}}{\lambda_{N,N}} & \cdots & \cdots\cdots & -\frac{\lambda_{N,N-1}}{\lambda_{N,N}} & 0 \end{bmatrix}$$

see (9.3.10), provided that $\lambda_{i,i} > 0$, $i = 0, 1, \ldots, N$.

9.7 Mean Absorption Time and Probabilities

Absorption probabilities

The absorption probabilities of the continuous-time process $(X_t)_{t\in\mathbb{R}_+}$ can be computed based on the behaviour of the embedded chain $(Z_n)_{n\geqslant 0}$. In fact the continuous waiting time between two jumps has no influence on the absorption probabilities. Here we consider only the simple example of birth and death processes, which can be easily generalized to more complex situations.

The basic idea is to perform a first step analysis on the underlying discrete-time embedded chain. Assume that state ⓪ is absorbing, *i.e.* $\lambda_0 = 0$, and let

$$T_0 = \inf\{t \geqslant 0 \ : \ X_t = 0\}$$

denote the absorption time of the chain into state ⓪. Let now

$$g_0(k) = \mathbb{P}(T_0 < \infty \mid X_0 = k), \qquad k = 0, 1, \ldots, N,$$

denote the probability of absorption in ⓪ starting from state $k \in \{0, 1, \ldots, N\}$. We have the boundary condition $g_0(0) = 1$, and by first step analysis on the chain $(Z_n)_{n\geqslant 1}$ we get

$$g_0(k) = \frac{\lambda_k}{\lambda_k + \mu_k} g_0(k+1) + \frac{\mu_k}{\lambda_k + \mu_k} g_0(k-1), \qquad k = 1, 2, \ldots, N-1.$$

When the rates $\lambda_k = \lambda$ and $\mu_k = \mu$ are independent of $k \in \{1, 2, \ldots, N-1\}$, this equation becomes

$$g_0(k) = p g_0(k+1) + q g_0(k-1), \qquad k = 1, 2, \ldots, N-1,$$

which is precisely Equation (2.2.6) for the gambling process with

$$p = \frac{\lambda}{\lambda + \mu} \quad \text{and} \quad q = \frac{\mu}{\lambda + \mu}.$$

When $\lambda_0 = \mu_N = 0$ we have the boundary conditions

$$g_0(0) = 1 \quad \text{and} \quad g_0(N) = 0$$

since the state $\textcircled{N}$ becomes absorbing, and the solution becomes

$$g_0(k) = \frac{(\mu/\lambda)^k - (\mu/\lambda)^N}{1 - (\mu/\lambda)^N}, \qquad k = 0, 1, \ldots, N,$$

when $\lambda \neq \mu$, according to (2.2.12). When $\lambda = \mu$, Relation (2.2.13) shows that

$$g_0(k) = \frac{N-k}{N} = 1 - \frac{k}{N}, \qquad k = 0, 1, \ldots, N.$$

Mean absorption time

We may still use the embedded chain $(Z_n)_{n\geqslant 0}$ to compute the mean absorption time, using the mean inter-jump times. Here, unlike in the case of absorption probabilities, the random time spent by the continuous-time process $(X_t)_{t\in\mathbb{R}_+}$ should be taken into account in the calculation.

Next, we consider a birth and death process on the state space $\mathbb{S} = \{0, 1, \ldots, N\}$ with absorbing states $\textcircled{0}$ and $\textcircled{N}$. Recall that the the mean time spent at state $\textcircled{i}$, given that the next transition is from $\textcircled{i}$ to $\boxed{i+1}$, is equal to

$$\mathbb{E}[\tau_{i,i+1}] = \frac{1}{\lambda_i}, \qquad i = 1, 2, \ldots, N-1,$$

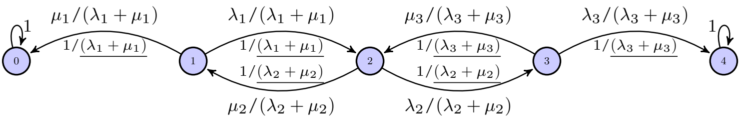
and the mean time spent at state $\textcircled{i}$, given that the next transition is from $\textcircled{i}$ to $\boxed{i-1}$, is equal to

$$\mathbb{E}[\tau_{i,i-1}] = \frac{1}{\mu_i}, \qquad i = 1, 2, \ldots, N-1.$$

We associate a *weighted* graph to the Markov chain $(Z_n)_{n\geqslant 0}$ that includes the average

$$\mathbb{E}[\tau_i] = \frac{1}{\lambda_i + \mu_i}$$

of the time $\tau_i = \min(\tau_{i,i-1}, \tau_{i,i+1})$ spent in state $\textcircled{i}$ before the next transition, $i = 1, 2, \ldots, N-1$. In the next graph, which is drawn for $N = 4$, the weights are *underlined*:



with $\lambda_0 = \mu_4 = 0$.

Proposition 9.10. *The mean absorption times*

$$h_{0,N}(i) = \mathbb{E}\left[T_{0,N} \mid X_0 = i\right], \qquad i = 0, 1, \ldots, N,$$

into states $\{0, N\}$ *starting from state* $i \in \{0, 1, \ldots, N\}$ *satisfy the boundary conditions* $h_{0,N}(0) = h_{0,N}(N) = 0$ *the first step analysis equation*

$$h_{0,N}(i) = \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i} h_{0,N}(i-1) + \frac{\lambda_i}{\lambda_i + \mu_i} h_{0,N}(i+1),$$

$i = 1, 2, \ldots, N-1$.

Proof. By first step analysis on the discrete-time embedded chain $(Z_n)_{n \geqslant 1}$ with transition matrix

$$P = \left[\, P_{i,j} \,\right]_{i,j \in \mathbb{S}} = \left[\begin{array}{cccccccc} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\mu_1}{\lambda_1 + \mu_1} & 0 & \frac{\lambda_1}{\lambda_1 + \mu_1} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{\lambda_{N-1}}{\lambda_{N-1} + \mu_{N-1}} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{array} \right],$$

we have

$$\begin{aligned} h_{0,N}(i) &= \frac{\mu_i}{\lambda_i + \mu_i} \left(\mathbb{E}[\tau_i] + h_{0,N}(i-1)\right) + \frac{\lambda_i}{\lambda_i + \mu_i} \left(\mathbb{E}[\tau_i] + h_{0,N}(i+1)\right) \\ &= \frac{\mu_i}{\lambda_i + \mu_i} \left(\frac{1}{\lambda_i + \mu_i} + h_{0,N}(i-1) \right) + \frac{\lambda_i}{\lambda_i + \mu_i} \left(\frac{1}{\lambda_i + \mu_i} + h_{0,N}(i+1) \right), \end{aligned} \tag{9.7.1}$$

$i = 1, 2, \ldots, N-1$. □

Note that by conditioning on the independent exponential random variables $\tau_{i,i-1}$ and $\tau_{i,i+1}$ we can also show that

$$\mathbb{E}[\tau_i \mid \tau_{i,i-1} < \tau_{i,i+1}] = \mathbb{E}[\tau_i \mid \tau_{i,i+1} < \tau_{i,i-1}] = \mathbb{E}[\tau_i] = \frac{1}{\lambda_i + \mu_i},$$

$i = 1, 2, \ldots, N-1$, cf. (1.8.10) in Exercise 1.5-(a), hence (9.7.1) can be rewritten as

$$\begin{aligned} h_{0,N}(i) &= \frac{\mu_i}{\lambda_i + \mu_i} \left(\mathbb{E}[\tau_i \mid \tau_{i,i-1} < \tau_{i,i+1}] + h_{0,N}(i-1) \right) \\ &\quad + \frac{\lambda_i}{\lambda_i + \mu_i} \left(\mathbb{E}[\tau_i \mid \tau_{i,i+1} < \tau_{i,i-1}] + h_{0,N}(i+1) \right). \end{aligned}$$

When the rates $\lambda_i = \lambda$ and $\mu_i = \mu$ are independent of $i \in \{1, 2, \ldots, N-1\}$, this equation becomes

$$h_{0,N}(i) = \frac{1}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} h_{0,N}(i+1) + \frac{\mu}{\lambda + \mu} h_{0,N}(i-1),$$

$i = 1, 2, \ldots, N-1$, which is a modification of Equation (2.3.7a), by replacing the discrete time step by the average time $1/(\lambda + \mu)$ spent at any state. Rewriting the equation as

$$h_{0,N}(i) = \frac{1}{\lambda + \mu} + p h_{0,N}(i+1) + q h_{0,N}(i-1),$$

$i = 1, 2, \ldots, N-1$, or

$$(\lambda + \mu) h_{0,N}(i) = 1 + p(\lambda + \mu) h_{0,N}(i+1) + q(\lambda + \mu) h_{0,N}(i-1),$$

$i = 1, 2, \ldots, N-1$, with

$$p = \frac{\lambda}{\lambda + \mu} \quad \text{and} \quad q = \frac{\mu}{\lambda + \mu},$$

we find from (2.3.12) that, with $r = q/p = \mu/\lambda$,

$$(\lambda + \mu) h_{0,N}(k) = \frac{1}{q - p} \left(k - N \frac{1 - r^k}{1 - r^N} \right),$$

i.e.

$$h_{0,N}(k) = \frac{1}{\mu - \lambda} \left(k - N \frac{1 - (\mu/\lambda)^k}{1 - (\mu/\lambda)^N} \right), \qquad k = 0, 1, \ldots, N, \quad (9.7.2)$$

when $\lambda \neq \mu$. In the limit $\lambda \to \mu$ we find by (2.3.19) that

$$h_{0,N}(k) = \frac{1}{2\mu} k(N - k), \qquad k = 0, 1, \ldots, N.$$

This solution is similar to that of the gambling problem with draw Exercise 2.1 as we multiply the solution of the gambling problem in the fair case by the average time $1/(2\mu)$ spent in any state in $\{1, 2, \ldots, N-1\}$.

The mean absorption time for the embedded chain $(Z_n)_{n\geqslant 0}$ can be recovered by dividing (9.7.2) by the mean time $\mathbb{E}[\tau_i] = 1/(\lambda + \mu)$ between two jumps, as

$$\frac{\lambda + \mu}{\mu - \lambda}\left(k - N\frac{1-(\mu/\lambda)^k}{1-(\mu/\lambda)^N}\right), \qquad k = 0, 1, \ldots, N, \tag{9.7.3}$$

which coincides with (2.3.12) in the non-symmetric case with $p = \lambda/(\lambda + \mu)$ and $p = \mu/(\lambda + \mu)$, and recovers (2.3.19), *i.e.*

$$k(N-k), \qquad k = 0, 1, \ldots, N,$$

in the symmetric case $\lambda = \mu$.

In Table 9.1 we gather some frequent questions with their corresponding solution methods.

Table 9.1: Summary of computing methods.

How to compute	Method
Infinitesimal generator $Q = (\lambda_{i,j})_{i,j\in \mathbb{S}}$	$Q = \frac{dP(t)}{dt}_{\mid t=0} = P'(0)$.
Semigroup $(P(t))_{t\in\mathbb{R}_+}$	$P(t) = \exp(tQ)$, $t \geqslant 0$, $P(h) = \mathrm{I} + hQ + o(h)$, $h \searrow 0$.
Stationary distribution π	Solve* $\pi Q = 0$ for π.
Probability distribution of the time $\tau_{i,j}$ spent in $i \to j$	Exponential distribution $(\lambda_{i,j})$.
Probability distribution of the time τ_i spent at state (i)	Exponential distribution $\left(\sum_{l\neq i} \lambda_{i,l}\right)$.
$\lim_{t\to\infty} \exp\left(t\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}\right)$	$\begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix}$
Hitting probabilities	Solve† $g = Pg$ for the embedded chain.
Mean hitting times	Use the embedded chain with weighted links using mean inter-jump times.

Exercises

Exercise 9.1 A workshop has five machines and one repairman. Each machine functions until it fails at an exponentially distributed random time with rate $\mu = 0.20$ per hour. On the other hand, it takes a exponentially distributed random time with (rate) $\lambda = 0.50$ per hour to repair a given machine. We assume that the machines behave independently of one another, and that

(i) up to five machines can operate at any given time,
(ii) at most one can be under repair at any time.

* Remember that the values of $\pi(k)$ have to add up to 1.

† Be sure to write only the relevant rows of the system under the appropriate boundary conditions.

Compute the proportion of time the repairman is idle in the long run.

Exercise 9.2 Two types of consultations occur at a database according to two independent Poisson processes: “read” consultations arrive at the rate (or intensity) λ_R and “write” consultations arrive at the rate (or intensity) λ_W.

a) What is the probability that the time interval between two consecutive “read” consultations is larger than $t > 0$?
b) What is the probability that during the time interval $[0, t]$, at most three “write” consultations arrive?
c) What is the probability that the next arriving consultation is a “read” consultation?
d) Determine the distribution of the number of arrived “read” consultations during $[0, t]$, given that in this interval a total number of n consultations occurred.

Exercise 9.3 Consider two machines, operating simultaneously and independently, where both machines have an exponentially distributed time to failure with mean $1/\mu$. There is a single repair facility, and the repair times are exponentially distributed with rate λ.

a) In the long run, what is the probability that no machines are operating when $\lambda = \mu = 1$?
b) We now assume that at most one machine can operate at any time. Namely, while one machine is working, the other one may be either under repair or already fixed and waiting to take over. How does this modify your answer to question (a)?

Exercise 9.4 Passengers arrive at a cable car station according to a Poisson process with intensity $\lambda > 0$. Each car contains at most 4 passengers, and a cable car arrives immediately and leaves with 4 passengers as soon as there are at least 4 people in the queue. We let $(X_t)_{t\in\mathbb{R}_+}$ denote the number of passengers in the waiting queue at time $t \geqslant 0$.

a) Explain why $(X_t)_{t\in\mathbb{R}_+}$ is a continuous-time Markov chain with state space $\mathbb{S} = \{0, 1, 2, 3\}$, and give its matrix infinitesimal generator Q.
b) Compute the limiting distribution $\pi = [\pi_0, \pi_1, \pi_2, \pi_3]$ of $(X_t)_{t\in\mathbb{R}_+}$.
c) Compute the mean time between two departures of cable cars.

Exercise 9.5 (Muni Toke (2015)). We consider a stock whose prices can only belong to the following five ticks:

$$\$10.01;\quad \$10.02;\quad \$10.03;\quad \$10.04;\quad \$10.05,$$

numbered $k = 1, 2, 3, 4, 5$.

At time t, the order book for this stock contains exactly $N_t^{(k)}$ sell orders at the price tick n$^o k$, $k = 1, 2, 3, 4, 5$, where $(N_t^{(k)})_{t\in\mathbb{R}_+}$ are independent Poisson processes with same intensity $\lambda > 0$. In addition,

- Any *sell* order can be cancelled after an exponentially distributed random time with parameter $\mu > 0$,
- *Buy* market orders are submitted according to another Poisson process with intensity $\theta > 0$, and are filled instantly at the lowest order price present in the book.

Order cancellations can occur as a result of various trading algorithms such as, *e.g.*, "spoofing", "layering", or "front running".

a) Show that the *total* number of sell orders L_t in the order book at time t forms a continuous-time Markov chain, and write down its infinitesimal generator Q.
b) It is estimated that 95% percent of high-frequency trader orders are later cancelled. What relation does this imply between μ and λ?

Exercise 9.6 The size of a fracture in a rock formation is modeled by a continuous-time pure birth process with parameters

$$\lambda_k = (1 + k)^\rho, \qquad k \geqslant 1,$$

i.e. the growth rate of the fracture is a power of $1 + k$, where k is the current fracture length. Show that when $\rho > 1$, the mean time for the fracture length to grow to infinity is finite. Conclude that the time to failure of the rock formation is almost-surely finite.*

Exercise 9.7 Customers arrive at a processing station according to a Poisson process with rate $\lambda = 0.1$, *i.e.* on average one customer per ten minutes. Processing of customer queries starts as soon as the third customer enters the queue.

a) Compute the expected time until the start of the customer service.
b) Compute the probability that no customer service occurs within the first hour.

Exercise 9.8 Online user logins are modeled according to a standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$ with intensity $\lambda > 0$. Each user spends the same amount of time $\theta > 0$ checking the site, and then quits. We assume that the process started at time $t = 0$.

a) Compute the average number of users checking the website at any given time t.

* Recall that a finite-valued random variable may have an infinite mean.

b) Compute the probability that no user is checking the website at a given time $t > 0$.

Exercise 9.9 Suppose that customers arrive at a facility according to a Poisson process having rate $\lambda = 3$. Let N_t be the number of customers that have arrived up to time t and let T_n be the arrival time of the nth customer, $n = 1, 2, \ldots$ Determine the following (conditional) probabilities and (conditional) expectations, where $0 < t_1 < t_2 < t_3 < t_4$.

a) $\mathbb{P}(N_{t_3} = 5 \mid N_{t_1} = 1)$.

b) $\mathbb{E}[N_{t_1} N_{t_4}(N_{t_3} - N_{t_2})]$.

c) $\mathbb{E}[N_{t_2} \mid T_2 > t_1]$.

Exercise 9.10 Let $(X_t)_{t\in\mathbb{R}_+}$ be a birth and death process on $\{0, 1, 2\}$ with birth and death parameters $\lambda_0 = 2\alpha$, $\lambda_1 = \alpha$, $\lambda_2 = 0$, and $\mu_0 = 0$, $\mu_1 = \beta$, $\mu_2 = 2\beta$. Determine the stationary distribution of $(X_t)_{t\in\mathbb{R}_+}$.

Exercise 9.11 Let $(X_t)_{t\in\mathbb{R}_+}$ be a birth and death process on $\{0, 1, \ldots, N\}$ with birth and death parameters $\lambda_n = \alpha(N - n)$ and $\mu_n = \beta n$, respectively. Determine the stationary distribution of $(X_t)_{t\in\mathbb{R}_+}$.

Exercise 9.12 Consider the pure birth process with birth rates $\lambda_0 = 1$, $\lambda_1 = 3$, $\lambda_2 = 2$, $\lambda_3 = 5$. Compute $P_{0,n}(t)$ for $n = 0, 1, 2, 3$.

Exercise 9.13 Consider a pure birth process $(X_t)_{t\in\mathbb{R}_+}$ started at $X_0 = 0$, and let T_k denote the time until the kth birth. Show that

$$\mathbb{P}(T_1 > t \text{ and } T_2 > t + s) = P_{0,0}(t)(P_{0,0}(s) + P_{0,1}(s)).$$

Determine the joint probability density function of (T_1, T_2), and then the joint density of $(\tau_0, \tau_1) := (T_1, T_2 - T_1)$.

Exercise 9.14 Cars pass a certain street location with identical speeds, according to a Poisson process with rate $\lambda > 0$. A woman at that location needs T units of time to cross the street, *i.e.* she waits until it appears that no car will cross that location within the next T time units.

a) Find the probability that her waiting time is 0.
b) Find her expected waiting time.
c) Find the total average time it takes to cross the street.

d) Assume that, due to other factors, the crossing time in the absence of cars is an independent exponentially distributed random variable with parameter $\mu > 0$. Find the total average time it takes to cross the street in this case.

Exercise 9.15 A machine is maintained at random times, such that the inter-service times $(\tau_k)_{k\geqslant 0}$ are *i.i.d.* with exponential distribution of parameter $\mu > 0$. The machine breaks down if it has not received maintenance for more than T units of time. After breaking down it is automatically repaired.

a) Compute the probability that the machine breaks down before its first maintenance after it is started.
b) Find the expected time until the machine breaks down.
c) Assuming that the repair time is exponentially distributed with parameter $\lambda > 0$, find the proportion of time the machine is working.

Exercise 9.16 A system consists of two machines and two repairmen. Each machine can work until failure at an exponentially distributed random time with parameter 0.2. A failed machine can be repaired only by one repairman, within an exponentially distributed random time with parameter 0.25. We model the number X_t of working machines at time $t \in \mathbb{R}_+$ as a continuous-time Markov process.

a) Complete the missing entries in the matrix

$$Q = \begin{bmatrix} \square & 0.5 & 0 \\ 0.2 & \square & \square \\ 0 & \square & -0.4 \end{bmatrix}$$

of its generator.
b) Calculate the long-run probability distribution $[\pi_0, \pi_1, \pi_2]$ of X_t.
c) Compute the average number of working machines in the long run.
d) Given that a working machine can produce 100 units every hour, how many units can the system produce per hour in the long run?
e) Assume now that in case a single machine is under failure then both repairmen can work on it, therefore dividing the mean repair time by a factor 2. Complete the missing entries in the matrix

$$Q = \begin{bmatrix} -0.5 & \square & \square \\ \square & -0.7 & \square \\ \square & \square & -0.4 \end{bmatrix}$$

of the modified generator and calculate the long run probability distribution $[\pi_0, \pi_1, \pi_2]$ for X_t.

Exercise 9.17 Let $X_1(t)$ and $X_2(t)$ be two independent two-state Markov chains on $\{0, 1\}$ and having the same infinitesimal matrix

$$\begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.$$

Argue that $Z(t) := X_1(t) + X_2(t)$ is a Markov chain on the state space $\mathbb{S} = \{0, 1, 2\}$ and determine the transition semigroup $P(t)$ of $Z(t)$.

Exercise 9.18 Consider a two-state discrete-time Markov chain $(\xi_n)_{n\geqslant 0}$ on $\{0, 1\}$ with transition matrix

$$\begin{bmatrix} 0 & 1 \\ 1-\alpha & \alpha \end{bmatrix}. \tag{9.7.4}$$

Let $(N_t)_{t\in\mathbb{R}_+}$ be a Poisson process with parameter $\lambda > 0$, and let the

$$X_t = \xi_{N_t}, \qquad t \in \mathbb{R}_+,$$

i.e. $(X_t)_{t\in\mathbb{R}_+}$ is a two-state birth and death process.

a) Compute the mean return time $\mathbb{E}[T_0^r \mid X_0 = 0]$ of X_t to state $\textcircled{0}$, where T_0^r is defined as

$$T_0^r = \inf\{t > T_1 \ : \ X_t = 0\}$$

and

$$T_1 = \inf\{t > 0 \ : \ X_t = 1\}$$

is the first hitting time of state $\textcircled{1}$.

b) Compute the mean return time $\mathbb{E}[T_1^r \mid X_0 = 1]$ of X_t to state $\textcircled{1}$, where T_1^r is defined as

$$T_1^r = \inf\{t > T_0 \ : \ X_t = 1\}$$

and

$$T_0 = \inf\{t > 0 \ : \ X_t = 0\}$$

is the first hitting time of state $\textcircled{0}$. The return time T_1^r to $\textcircled{1}$ starting from $\textcircled{1}$ is evaluated by switching first to state $\textcircled{0}$ before returning to state $\textcircled{1}$.

c) Show that $(X_t)_{t\in\mathbb{R}_+}$ is a two-state birth and death process and determine its generator matrix Q in terms of α and λ.

Problem 9.19 Let $(N_t^1)_{t\in\mathbb{R}_+}$ and $(N_t^2)_{t\in\mathbb{R}_+}$ be two independent Poisson processes with intensities $\lambda_1 > 0$ and $\lambda_2 > 0$.

a) Show that $(N_t^1 + N_t^2)_{t\in\mathbb{R}_+}$ is a Poisson process and find its intensity.

b) Consider the difference

$$M_t = N_t^1 - N_t^2, \qquad t \geqslant 0,$$

and that $(M_t)_{t\in\mathbb{R}_+}$ has stationary independent increments.

c) Find the distribution of $M_t - M_s$, $0 < s < t$.
d) Compute

$$\lim_{t\to\infty} \mathbb{P}(|M_t| \leqslant c)$$

for any $c > 0$.
e) Suppose that N_t^1 denotes the number of clients arriving at a taxi station during the time interval $[0,t]$, and that N_t^2 denotes the number of taxis arriving at that same station during the same time interval $[0,t]$.

How do you interpret the value of M_t depending on its sign?

How do you interpret the result of Question (d)?

Problem 9.20 We consider a birth and death process $(X_t)_{t\in\mathbb{R}_+}$ on $\{0,1,\ldots,N\}$ with transition semigroup $(P(t))_{t\in\mathbb{R}}$ and birth and death rates

$$\lambda_n = (N-n)\lambda, \qquad \mu_n = n\mu, \qquad n = 0,1,\ldots,N.$$

This process is used for the modeling of membrane channels in neuroscience.

a) Write down the infinitesimal generator Q of $(X_t)_{t\in\mathbb{R}_+}$.
b) From the forward Kolmogorov equation $P'(t) = P(t)Q$, show that for all $n = 0,1,\ldots,N$ we have

$$\begin{cases} P'_{n,0}(t) = -\lambda_0 P_{n,0}(t) + \mu_1 P_{n,1}(t), \\ \\ P'_{n,k}(t) = \lambda_{k-1} P_{n,k-1}(t) - (\lambda_k + \mu_k) P_{n,k}(t) + \mu_{k+1} P_{n,k+1}(t), \\ \\ P'_{n,N}(t) = \lambda_{N-1} P_{n,N-1}(t) - \mu_N P_{n,N}(t), \end{cases}$$

$k = 1,2,\ldots,N-1$.
c) Let

$$G_k(s,t) = \mathbb{E}\big[s^{X_t} \mid X_0 = k\big] = \sum_{n=0}^N s^n \mathbb{P}(X_t = n \mid X_0 = k) = \sum_{n=0}^N s^n P_{k,n}(t)$$

denote the generating function of X_t given that $X_0 = k \in \{0,1,\ldots,N\}$. From the result of Question (b), show that $G_k(s,t)$ satisfies the *partial differential equation*

$$\frac{\partial G_k}{\partial t}(s,t) = \lambda N(s-1)G_k(s,t) + (\mu + (\lambda-\mu)s - \lambda s^2)\frac{\partial G_k}{\partial s}(s,t), \quad (9.7.5)$$

with $G_k(s,0) = s^k$, $k = 0,1,\ldots,N$.
d) Verify that the solution of (9.7.5) is given by

$$G_k(s,t) = \frac{1}{(\lambda+\mu)^N}(\mu + \lambda s + \mu(s-1)\mathrm{e}^{-(\lambda+\mu)t})^k(\mu + \lambda s - \lambda(s-1)\mathrm{e}^{-(\lambda+\mu)t})^{N-k},$$

$k = 0, 1, \ldots, N$.

e) Show that

$$\mathbb{E}[X_t \mid X_0 = k] = \frac{k}{(\lambda+\mu)^N}(\lambda + \mu e^{-(\lambda+\mu)t})(\mu+\lambda)^{k-1}(\mu+\lambda)^{N-k} \\ + \frac{N-k}{(\lambda+\mu)^N}(\mu+\lambda)^k(\lambda - \lambda e^{-(\lambda+\mu)t})(\mu+\lambda)^{N-k-1}.$$

f) Compute

$$\lim_{t\to\infty} \mathbb{E}[X_t \mid X_0 = k]$$

and show that it does not depend on $k \in \{0, 1, \ldots, N\}$.

Problem 9.21 Let $T_1, T_2, \ldots$ be the first two jump times of a Poisson process $(N_t)_{t\in\mathbb{R}_+}$ with intensity λ. Given $f : \mathbb{R} \longrightarrow \mathbb{R}$ an integrable function, show that

$$\mathbb{E}\left[\sum_{k=1}^{N_t} f(T_i)\right] = \lambda \int_0^t f(s)ds. \tag{9.7.6}$$

Chapter 10
Discrete-Time Martingales

As mentioned in the introduction, stochastic processes can be classified into two main families, namely *Markov processes* on the one hand, and *martingales* on the other hand. Markov processes have been our main focus of attention so far, and in this chapter we turn to the notion of martingale. We will give a precise mathematical meaning to the description of martingales, which says that when $(X_n)_{n\geqslant 0}$ is a martingale, the best possible estimate at time $n \geqslant 0$ of the future value X_m at time $m > n$ is X_n itself. In this chapter, the main application of martingales will be to recover in an elegant way the previous results on gambling processes of Chapter 2. The concept of martingale has also many applications in stochastic modeling, for example in financial mathematics, where martingales are used to characterize the fairness and equilibrium in a market model.

10.1 Filtrations and Conditional Expectations

Before dealing with martingales we need to introduce the important notion of *filtration* generated by a discrete-time stochastic process $(S_n)_{n\geqslant 0}$. In what follows, we work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 10.1. *The* filtration $(\mathcal{F}_n)_{n\geqslant 0}$ *generated by a stochastic process* $(S_n)_{n\geqslant 0}$ *taking its values in a state space* $\mathbb{S}$*, is the family of* σ*-algebras*

$$\mathcal{F}_n := \sigma(S_0, S_1, \ldots, S_n), \qquad n \geqslant 0,$$

where $\sigma(S_0, S_1, \ldots, S_n)$ *denotes the collections of events generated by* $S_0, S_1, \ldots, S_n$, $n \geqslant 0$.

Examples of events in $\mathcal{F}_n$ include:

$$\{S_0 \leqslant a_0, \ S_1 \leqslant a_1, \ \ldots, \ S_n \leqslant a_n\}$$

for $a_0, a_1, \ldots, a_n$ a given fixed sequence of real numbers. Note that we have the inclusion $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $n \geqslant 0$, *i.e.* $(\mathcal{F}_n)_{n\geqslant 0}$ is *non-decreasing.*

One refers to $\mathcal{F}_n$ as the *information* generated by $(S_k)_{k\geqslant 0}$ up to time n, and to $(\mathcal{F}_n)_{n\geqslant 0}$ as the *information flow* generated by $(S_n)_{n\geqslant 0}$. We say that a random variable is $\mathcal{F}_n$-measurable whenever F can be written as a function $F = f(S_0, S_1, \ldots, S_n)$ of $(S_0, S_1, \ldots, S_n)$.

Example

Consider the simple random walk

$$S_n := X_1 + X_2 + \cdots + X_n, \qquad n \geqslant 0,$$

where $(X_k)_{k\geqslant 1}$ is a sequence of independent, identically distributed $\{-1, 1\}$-valued random variables, and $S_0 := 0$.

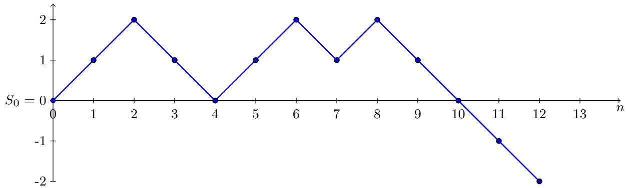


Fig. 10.1: Sample path of the random walk $(S_n)_{n\geqslant 0}$.

The filtration (or information flow) $(\mathcal{F}_n)_{n\geqslant 0}$ generated by $(S_n)_{n\geqslant 0}$ satisfies

$$\mathcal{F}_0 = \big\{\{S_0 \neq 0\}, \{S_0 = 0\}\big\} = \big\{\emptyset, \Omega\big\}, \quad \mathcal{F}_1 = \big\{\emptyset, \{X_1 = 1\}, \{X_1 = -1\}, \Omega\big\}, \tag{10.1.1}$$

and

$$\mathcal{F}_2 = \sigma\big(\big\{\emptyset, \{X_1 = 1, X_2 = 1\}, \{X_1 = 1, X_2 = -1\}, \{X_1 = -1, X_2 = 1\}, \{X_1 = -1, X_2 = -1\}, \Omega\big\}\big).$$

The notation $\mathcal{F}_n$ is useful to represent a quantity of information available at time n, and various sub σ-algebras of $\mathcal{F}_n$ can be defined. For example, the

σ-algebra $\mathcal{G}$ generated by S_2 satisfies

$$\begin{aligned}\mathcal{G} =& \sigma\big(\{\emptyset, \{S_2 = -2\}, \{S_2 = 2\}, \{S_2 = 0\}, \Omega\}\big)\\ =& \sigma\big(\{\emptyset, \{X_1 = -1, X_2 = -1\}, \{X_1 = 1, X_2 = 1\}\\ & \{X_1 = 1, X_2 = -1\} \cup \{X_1 = -1, X_2 = 1\}, \Omega\}\big),\end{aligned}$$

which contains less information than $\mathcal{F}_2$, as it only tells whether the increments X_1, X_2 are both equal to 1 or to -1.

Definition 10.2. *A stochastic process* $(Z_n)_{n\geqslant 0}$ *is said to be* $(\mathcal{F}_n)_{n\geqslant 0}$-adapted *if the value of* Z_n *depends on no more than the information available up to time* n *in* $\mathcal{F}_n$.

In other words, the value of an $(\mathcal{F}_n)_{n\geqslant 0}$-adapted process Z_n is determined by a function of $X_0, X_1, \ldots, X_n$ for all $n \geqslant 0$.

We now review the definition of *conditional expectation*, see Sections 1.6 and 1.8 for details. Given X a random variable with finite mean, the conditional expectation $\mathbb{E}[X \mid \mathcal{F}_n]$ refers to

$$\mathbb{E}[X \mid X_0, X_1, \ldots, X_n] = \mathbb{E}[X \mid X_0 = k_0, \ldots, X_n = k_n]_{k_0 = X_0, \ldots, k_n = X_n},$$

given that $X_0, X_1, \ldots, X_n$ are respectively equal to $k_0, k_1, \ldots, k_n \in \mathbb{S}$.

The conditional expectation $\mathbb{E}[X \mid \mathcal{F}_n]$ is itself a random variable that depends only on the values of $X_0, X_1, \ldots, X_n$, *i.e.* on the history of the process up to time $n \in \mathbb{N}$. It can also be interpreted as the best possible estimate of X in mean-square sense, given the values of $X_0, X_1, \ldots, X_n$, see Propositions 1.8.2 and 1.19.

By point (ii) page 47, any integrable $\mathcal{F}_n$-*adapted* process $(Z_n)_{n\geqslant 0}$ satisfies

$$\mathbb{E}[Z_n \mid \mathcal{F}_n] = Z_n, \qquad n \geqslant 0.$$

10.2 Martingales - Definition and Properties

We now turn to the definition of *martingale*.

Definition 10.3. *An integrable,* discrete-time stochastic process* $(M_n)_{n\geqslant 0}$ *is a* martingale *with respect to* $(\mathcal{F}_n)_{n\geqslant 0}$ *if* $(M_n)_{n\geqslant 0}$ *is* $(\mathcal{F}_n)_{n\geqslant 0}$*-adapted and satisfies the property*

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n, \qquad n \geqslant 0. \tag{10.2.1}$$

* Integrable means $\mathbb{E}[|M_n|] < \infty$ for all $n \geqslant 0$.

The process $(M_n)_{n\geqslant 0}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geqslant 0}$ if, given the information $\mathcal{F}_n$ known up to time n, the best possible estimate of M_{n+1} is simply M_n.

Remark 10.4. *Definition 10.3 can be equivalently stated by saying that*

$$\mathbb{E}[M_n \mid \mathcal{F}_k] = M_k, \qquad k = 0, 1, \ldots, n.$$

Proof. Let $k \geqslant 0$. We prove the statement by induction on $n \geqslant k$. For $n = k$ we have $M_k = \mathbb{E}[M_k \mid \mathcal{F}_k]$. Next, assuming that $M_k = \mathbb{E}[M_n \mid \mathcal{F}_k]$ for some $n \geqslant k$, by the *tower property* of conditional expectations, see point (iii) page 47, we have

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_k] = \mathbb{E}[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \mid \mathcal{F}_k] = \mathbb{E}[M_n \mid \mathcal{F}_k] = M_k.$$

□

A particular property of martingales is that their expectation is constant over time.

Proposition 10.5. *Let $(M_n)_{n\geqslant 0}$ be an $(\mathcal{F}_n)_{n\geqslant 0}$-martingale. We have*

$$\mathbb{E}[M_n] = \mathbb{E}[M_0], \qquad n \geqslant 0.$$

Proof. By (1.8.6) and the *tower property* of conditional expectations, see point (iii) page 47, we have

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_0] = \mathbb{E}[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \mid \mathcal{F}_0] = \mathbb{E}[M_n \mid \mathcal{F}_0], \qquad n \geqslant 0.$$

Hence, by induction on $n \geqslant 0$, we have

$$\begin{aligned}
\mathbb{E}[M_{n+1} \mid \mathcal{F}_0] &= \mathbb{E}[M_n \mid \mathcal{F}_0] \\
&= \mathbb{E}[M_{n-1} \mid \mathcal{F}_0] \\
&= \cdots \\
&= \mathbb{E}[M_1 \mid \mathcal{F}_0] \\
&= \mathbb{E}[M_0 \mid \mathcal{F}_0], \quad n \geqslant 0,
\end{aligned}$$

hence

$$\mathbb{E}[M_n \mid \mathcal{F}_0] = \mathbb{E}[M_0 \mid \mathcal{F}_0], \qquad n \geqslant 0,$$

and the conclusion follows from (1.8.5). □

Examples of martingales

1. Any *centered** integrable process $(S_n)_{n\geqslant 0}$ with mutually *independent* increments is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ generated by $(S_n)_{n\geqslant 0}$, see the example of page 362.

   Indeed, in this case we have

$$\begin{aligned}\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[S_n + S_{n+1} - S_n \mid \mathcal{F}_n]\\ &= \mathbb{E}[S_n \mid \mathcal{F}_n] + \mathbb{E}[S_{n+1} - S_n \mid \mathcal{F}_n]\\ &= \mathbb{E}[S_n \mid \mathcal{F}_n] + \underbrace{\mathbb{E}[S_{n+1} - S_n]}_{=0}\\ &= \mathbb{E}[S_n \mid \mathcal{F}_n] = S_n, \qquad n \geqslant 0.\end{aligned}$$

   In addition to being a martingale, a stochastic process $(S_n)_{n\geqslant 0}$ with centered independent increments is also a Markov process, cf. Section 4.1.

   However, not all martingales have the Markov property, and not all Markov processes are martingales. In addition, there are martingales and Markov processes which do not have independent increments.

2. Given $X \in L^2(\Omega)$ a square-integrable random variable and $(\mathcal{F}_n)_{n\geqslant 0}$ a filtration, the process $(X_n)_{n\geqslant 0}$ defined by

$$X_n := \mathbb{E}[X \mid \mathcal{F}_n], \qquad n \geqslant 0,$$

   is an $(\mathcal{F}_n)_{n\geqslant 0}$-martingale under the probability measure $\mathbb{P}$, as follows from the tower property (1.6.8):

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n] = X_n, \quad n \geqslant 0. \tag{10.2.2}$$

Figure 10.2 illustrates various estimates $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$ at times $n = 0$ ("Wed"), $n = 1$ ("Thu"), $n = 2$ ("Fri"), $n = 3$ ("Sat"), for a random outcome $X =$"Saturday temperature" known at time "Sat", *i.e.*

$$\begin{cases} X_{\text{Wed}} = \mathbb{E}[X \mid \mathcal{F}_0] = 26,\\ X_{\text{Thu}} = \mathbb{E}[X \mid \mathcal{F}_1] = 28,\\ X_{\text{Fri}} = \mathbb{E}[X \mid \mathcal{F}_2] = 26,\\ X_{\text{Sat}} = \mathbb{E}[X \mid \mathcal{F}_3] = 24.\end{cases}$$

* A process $(S_n)_{n\geqslant 0}$ is said to be *centered* if $\mathbb{E}[S_n] = 0$ for all $n \geqslant 0$.

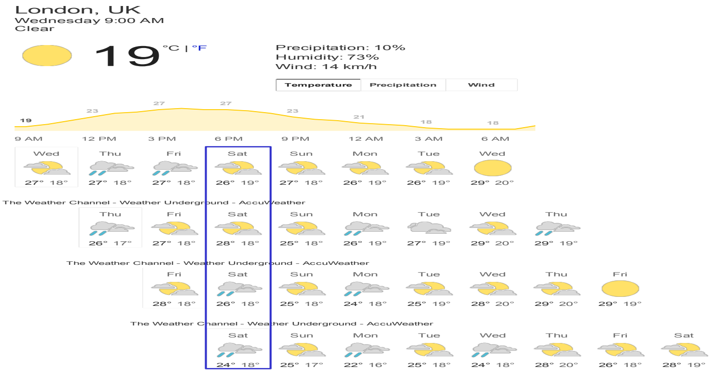


Fig. 10.2: Updated weather forecast.

10.3 Stopping Times

Next, we turn to the definition of stopping time. If an event occurs at a (random) *stopping time*, it should be possible, at any time $n \in \mathbb{N}$, to determine whether the event has already occured, based on the information available up to time n. This idea is formalized in the next definition.

Definition 10.6. *A random variable* $\tau : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ *is a* stopping time *with respect to* $(\mathcal{F}_n)_{n \geqslant 0}$ *if*

$$\{\tau > n\} \in \mathcal{F}_n, \qquad n \geqslant 0. \tag{10.3.1}$$

The meaning of Relation (10.3.1) is that the knowledge of $\{\tau > n\}$ depends only on the information present in $\mathcal{F}_n$ up to time n, *i.e.* on the knowledge of $X_0, X_1, \ldots, X_n$.

Note that condition (10.3.1) is equivalent to the condition

$$\{\tau \leqslant n\} \in \mathcal{F}_n, \qquad n \geqslant 0,$$

since $\{\tau \leqslant n\} = \{\tau > n\}^c$ and $\mathcal{F}_n$ is *stable by complement*.

Not every $\mathbb{N}$-valued random variable is a stopping time, however, hitting times provide natural examples of stopping times. We refer to Definition 10.1 for the construction of the filtration generated by a discre-time stochastic process.

Proposition 10.7. *Let* $(X_n)_{n \geqslant 0}$ *denote a stochastic process taking its values in* $\mathbb{S}$, *and generating a filtration* $(\mathcal{F}_n)_{n \geqslant 0}$ *The first hitting time*

$$\tau_x := \inf\{k \geqslant 0 \ : \ X_k = x\}$$

of $x \in \mathbb{S}$ *by* $(X_n)_{n\geqslant 0}$ *is a* stopping time *with respect to the filtration* $(\mathcal{F}_n)_{n\geqslant 0}$.

Proof. We have

$$\begin{aligned}\{\tau_x > n\} &= \{X_0 \neq x, \ X_1 \neq x, \ \ldots, X_n \neq x\} \\ &= \{X_0 \neq x\} \cap \{X_1 \neq x\} \cap \cdots \cap \{X_n \neq x\} \in \mathcal{F}_n, \quad n \geqslant 0,\end{aligned}$$

since

$$\{X_0 \neq x\} \in \mathcal{F}_0 \subset \mathcal{F}_n, \ \{X_1 \neq x\} \in \mathcal{F}_1 \subset \mathcal{F}_n, \ldots, \ \{X_n \neq x\} \in \mathcal{F}_n, \quad n \geqslant 0.$$

□

Hitting times can be used to trigger "buy limit" or "sell stop" orders in finance. On the other hand, the first time

$$\tau := \inf\left\{k \geqslant 0 \ : \ X_k = \operatorname*{Max}_{l=0,1,\ldots,N} X_l\right\}$$

the process $(X_k)_{k\geqslant 0}$ reaches its maximum over $\{0, 1, \ldots, N\}$ is not a stopping time. Indeed, it is not possible to decide whether $\{\tau \leqslant n\}$, *i.e.* the maximum has been reached before time n, based on the information available up to time n.

Exercise: Show from Definition 10.6 that the minimum $\tau \wedge \nu := \min(\tau, \nu)$ and the maximum $\tau \vee \nu := \operatorname{Max}(\tau, \nu)$ of two stopping times are themselves stopping times.

Definition 10.8. *Given* $(Z_n)_{n\geqslant 0}$ *a stochastic process and* $\tau : \Omega \longrightarrow \mathbb{N}$ *a stopping time, the* stopped process

$$(Z_{\tau\wedge n})_{n\geqslant 0} = (Z_{\min(\tau,n)})_{n\geqslant 0}$$

is defined as

$$Z_{\tau\wedge n} = Z_{\min(\tau,n)} = \begin{cases} Z_n & \text{if } n < \tau, \\ Z_\tau & \text{if } n \geqslant \tau, \end{cases}$$

Using indicator functions, we may also write

$$Z_{\tau\wedge n} = Z_n \mathbb{1}_{\{n<\tau\}} + Z_\tau \mathbb{1}_{\{n\geqslant\tau\}}, \qquad n \geqslant 0.$$

Figure 10.3 is an illustration of the path of a stopped process.

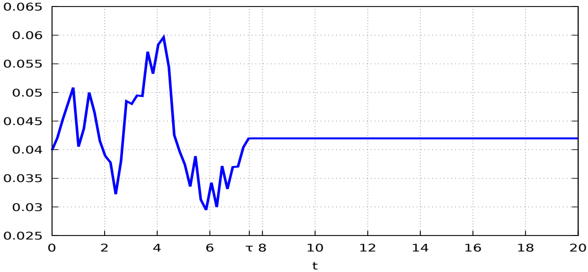


Fig. 10.3: Stopped process.

The following Theorem 10.9 is called the *Stopping Time Theorem*, and is due to J.L. Doob (1910-2004).

Theorem 10.9. *Assume that* $(M_n)_{n\geqslant 0}$ *is a martingale with respect to* $(\mathcal{F}_n)_{n\geqslant 0}$ *and that* $\tau : \Omega \longrightarrow \mathbb{N}\cup\{+\infty\}$ *is a stopping time with respect to* $(\mathcal{F}_n)_{n\geqslant 0}$*. Then the* stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$ *is also a martingale with respect to* $(\mathcal{F}_n)_{n\geqslant 0}$*.*

Proof. Writing

$$M_n = M_0 + \sum_{l=1}^{n}(M_l - M_{l-1}) = M_0 + \sum_{l\geqslant 1} \mathbb{1}_{\{l\leqslant n\}}(M_l - M_{l-1}),$$

we have

$$M_{\tau\wedge n} = M_0 + \sum_{l=1}^{\tau\wedge n}(M_l - M_{l-1}) = M_0 + \sum_{l=1}^{n} \mathbb{1}_{\{l\leqslant \tau\}}(M_l - M_{l-1}),$$

and for $k \leqslant n$ we find

$$\begin{aligned}
\mathbb{E}[M_{\tau\wedge n} \mid \mathcal{F}_k] &= \mathbb{E}\left[M_0 + \sum_{l=1}^{n} \mathbb{1}_{\{l\leqslant \tau\}}(M_l - M_{l-1}) \,\middle|\, \mathcal{F}_k\right] \\
&= M_0 + \sum_{l=1}^{n} \mathbb{E}\big[\mathbb{1}_{\{l\leqslant \tau\}}(M_l - M_{l-1}) \,\big|\, \mathcal{F}_k\big] \\
&= M_0 + \sum_{l=1}^{k} \mathbb{E}\big[\mathbb{1}_{\{l\leqslant \tau\}}(M_l - M_{l-1}) \,\big|\, \mathcal{F}_k\big] + \sum_{l=k+1}^{n} \mathbb{E}\big[\mathbb{1}_{\{l\leqslant \tau\}}(M_l - M_{l-1}) \,\big|\, \mathcal{F}_k\big] \\
&= M_0 + \sum_{l=1}^{k}(M_l - M_{l-1})\mathbb{E}\big[\mathbb{1}_{\{\tau > l-1\}} \,\big|\, \mathcal{F}_k\big]
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{l=k+1}^{n} \mathbb{E}\big[\mathbb{E}\big[(M_l - M_{l-1})\mathbb{1}_{\{l-1<\tau\}} \,\big|\, \mathcal{F}_{l-1}\big] \,\big|\, \mathcal{F}_k\big] \\
&= M_0 + \sum_{l=1}^{k} (M_l - M_{l-1})\mathbb{1}_{\{l\leqslant\tau\}} \\
&\quad + \sum_{l=k+1}^{n} \mathbb{E}\Big[\mathbb{1}_{\{l-1<\tau\}} \underbrace{\mathbb{E}\big[M_l - M_{l-1} \,\big|\, \mathcal{F}_{l-1}\big]}_{=0} \,\Big|\, \mathcal{F}_k\Big] \\
&= M_0 + \sum_{l=1}^{\tau\wedge k} (M_l - M_{l-1}) \\
&= M_{\tau\wedge k},
\end{aligned}
$$

$k = 0, 1, \ldots, n$, where we used the tower property (1.6.8), the martingale property

$$
\begin{aligned}
\mathbb{E}\big[M_l - M_{l-1} \,\big|\, \mathcal{F}_{l-1}\big] &= \mathbb{E}\big[M_l \,\big|\, \mathcal{F}_{l-1}\big] - \mathbb{E}\big[M_{l-1} \,\big|\, \mathcal{F}_{l-1}\big] \\
&= M_{l-1} - M_{l-1} = 0
\end{aligned}
$$

of $(M_l)_{l\geqslant 0}$, and the fact that

$$
\{\tau \geqslant l\} = \{\tau > l-1\} \in \mathcal{F}_{l-1} \supset \mathcal{F}_k, \qquad l \geqslant k+1.
$$

□

By the Stopping Time Theorem 10.9 we know that the stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$ is a martingale, hence its expectation is constant over time by Proposition 10.5.

i) If τ is a stopping time bounded by a constant $N \geqslant 1$, *i.e.* $\tau \leqslant N$, by Proposition 10.5 we have

$$
\mathbb{E}[M_\tau] = \mathbb{E}[M_{\tau\wedge N}] = \mathbb{E}[M_{\tau\wedge 0}] = \mathbb{E}[M_0]. \tag{10.3.2}
$$

ii) As a consequence of (10.3.2), if $(M_n)_{n\geqslant 0}$ is a martingale and $\tau \leqslant N$ and $\nu \leqslant N$ are two *bounded* stopping times bounded by a constant $N \geqslant 1$, we have

$$
\mathbb{E}[M_\tau] = \mathbb{E}[M_\nu] = \mathbb{E}[M_0]. \tag{10.3.3}
$$

iii) In case τ is only *a.s.* finite, *i.e.* $\mathbb{P}(\tau < \infty) = 1$, we may also write

$$
\mathbb{E}[M_\tau] = \mathbb{E}\left[\lim_{n\to\infty} M_{\tau\wedge n}\right] = \lim_{n\to\infty} \mathbb{E}[M_{\tau\wedge n}] = \mathbb{E}[M_0],
$$

provided that the limit and expectation signs can be exchanged, however this may not be always the case.

In some situations the exchange of limit and expectation signs may not be valid.* Nevertheless, the exchange is possible when the stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$ is bounded in absolute value, *i.e.* $|M_{\tau\wedge n}| \leqslant K$ *a.s.*, $n \geqslant 0$, for some constant $K > 0$, as a consequence of the *dominated convergence theorem*.

Analog statements can be proved for *sub*martingales, see *e.g.* Exercise 10.6 for this notion.

10.4 Ruin Probabilities

In the sequel we will show that, as an application of the Stopping Time Theorem 10.9, the ruin probabilities computed for random walks in Chapter 2 can be recovered in a simple and elegant way.

Consider the standard random walk (or gambling process) $(S_n)_{n\geqslant 0}$ on $\{0, 1, \ldots, B\}$ with independent $\{-1, 1\}$-valued increments, and

$$\mathbb{P}(S_{n+1} - S_n = +1) = p \quad \text{and} \quad \mathbb{P}(S_{n+1} - S_n = -1) = q, \quad n \geqslant 0,$$

as introduced in Section 2.1. Let

$$\tau_{0,B} : \Omega \longrightarrow \mathbb{N}$$

be the first hitting time of the boundary $\{0, B\}$, defined by

$$\tau = \tau_{0,B} := \inf\{n \geqslant 0 \ : \ S_n = B \text{ or } S_n = 0\}. \qquad (10.4.1)$$

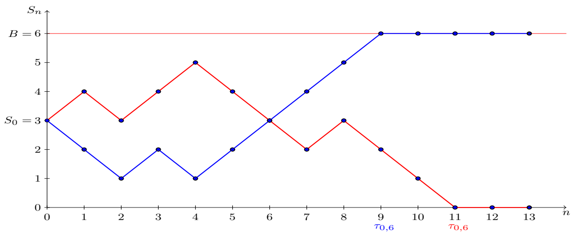


Fig. 10.4: Sample paths of the random walk $(S_n)_{n\geqslant 0}$.

One can easily check that the event $\{\tau > n\}$ depends only on the history of $(S_k)_{k\geqslant 0}$ up to time n since

* Consider for example the sequence $M_n := n\mathbb{1}_{\{U<1/n\}}$, $n \geqslant 1$, where $U \simeq U(0, 1]$ is a uniformly distributed random variable on $(0, 1]$, see also Exercise 10.5.

$$\{\tau > n\} = \{0 < S_0 < B\} \cap \{0 < S_1 < B\} \cap \cdots \cap \{0 < S_n < B\} \subset \mathcal{F}_n,$$

hence τ is a stopping time.

We will recover the ruin probabilities

$$\mathbb{P}(S_\tau = 0 \mid S_0 = k), \qquad k = 0, 1, \ldots, B,$$

computed in Chapter 2 in three steps, first in the unbiased case $p = q = 1/2$ (note that the hitting time τ can be shown to be *a.s.* finite, *i.e.* $\mathbb{P}(\tau < \infty) = 1$, cf. *e.g.* the identity (2.2.30)).

Unbiased case $p = q = 1/2$

Step 1. The process $(S_n)_{n \geqslant 0}$ is a martingale.

We note that the process $(S_n)_{n \geqslant 0}$ has independent increments, and in the unbiased case $p = q = 1/2$ those increments are centered:

$$\mathbb{E}[S_{n+1} - S_n] = 1 \times p + (-1) \times q = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0, \qquad (10.4.2)$$

hence $(S_n)_{n \geqslant 0}$ is a *martingale* by Point 1 page 365.

Step 2. The stopped process $(S_{\tau \wedge n})_{n \geqslant 0}$ is also a martingale, as a consequence of the Stopping Time Theorem 10.9.

Step 3. Since the stopped process $(S_{\tau \wedge n})_{n \geqslant 0}$ is a martingale by the Stopping Time Theorem 10.9, we find that its expectation $\mathbb{E}[S_{\tau \wedge n} \mid S_0 = k]$ is constant in $n \geqslant 0$ by Proposition 10.5, which gives

$$k = \mathbb{E}[S_0 \mid S_0 = k] = \mathbb{E}[S_{\tau \wedge n} \mid S_0 = k], \qquad k = 0, 1, \ldots, B.$$

Letting n go to infinity and noting that $|S_{\tau \wedge n}| \leqslant B$, $n \geqslant 0$, by dominated convergence we get

$$\begin{aligned}
\mathbb{E}[S_\tau \mid S_0 = k] &= \mathbb{E}\Big[\lim_{n \to \infty} S_{\tau \wedge n} \,\Big|\, S_0 = k\Big] \\
&= \lim_{n \to \infty} \mathbb{E}[S_{\tau \wedge n} \mid S_0 = k] \\
&= \lim_{n \to \infty} \mathbb{E}[S_{\tau \wedge 0} \mid S_0 = k] \\
&= \lim_{n \to \infty} \mathbb{E}[S_0 \mid S_0 = k] \\
&= k,
\end{aligned}$$

provided that $\mathbb{P}(\tau < \infty) = 1$, see Exercise 10.1, where the exchange between limit and expectation is justified by the boundedness $|S_{\tau \wedge n}| \leqslant B$ *a.s.*, $n \geqslant 0$. Hence we have

$$\begin{cases} 0 \times \mathbb{P}(S_\tau = 0 \mid S_0 = k) + B \times \mathbb{P}(S_\tau = B \mid S_0 = k) = \mathbb{E}[S_\tau \mid S_0 = k] = k \\ \mathbb{P}(S_\tau = 0 \mid S_0 = k) + \mathbb{P}(S_\tau = B \mid S_0 = k) = \mathbb{P}(\tau < \infty) = 1, \end{cases}$$

which shows that

$$\mathbb{P}(S_\tau = B \mid S_0 = k) = \frac{k}{B} \quad \text{and} \quad \mathbb{P}(S_\tau = 0 \mid S_0 = k) = 1 - \frac{k}{B},$$

$k = 0, 1, \ldots, B$, which recovers (2.2.22) without relying on boundary conditions, and with shorter calculations. Namely, the solution has been obtained in a simple way without solving any finite difference equation, demonstrating the power of the martingale approach.

Biased case $p \neq q$

Next, we turn to the biased case where $p \neq q$. In this case the process $(S_n)_{n\geqslant 0}$ is no longer a martingale because its increments are not centered:

$$\mathbb{E}[S_{n+1} - S_n] = 1 \times p + (-1) \times q = p - q \neq 0. \tag{10.4.3}$$

In order to apply the Stopping Time Theorem 10.9, we need to construct a martingale of a different type. Here, we note that the process

$$M_n := \left(\frac{q}{p}\right)^{S_n}, \qquad n \geqslant 0,$$

is a martingale with respect to $(\mathcal{F}_n)_{n\geqslant 0}$.

Step 1. The process $(M_n)_{n\geqslant 0}$ is a martingale.

Indeed, we have

$$\begin{aligned} \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n+1}} \Bigg| \mathcal{F}_n\right] = \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n+1}-S_n} \left(\frac{q}{p}\right)^{S_n} \Bigg| \mathcal{F}_n\right] \\ &= \left(\frac{q}{p}\right)^{S_n} \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n+1}-S_n} \Bigg| \mathcal{F}_n\right] \\ &= \left(\frac{q}{p}\right)^{S_n} \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n+1}-S_n}\right] \\ &= \left(\frac{q}{p}\right)^{S_n} \left(\frac{q}{p}\mathbb{P}(S_{n+1} - S_n = 1) + \left(\frac{q}{p}\right)^{-1} \mathbb{P}(S_{n+1} - S_n = -1)\right) \\ &= \left(\frac{q}{p}\right)^{S_n} \left(p\frac{q}{p} + q\left(\frac{q}{p}\right)^{-1}\right) \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{q}{p}\right)^{S_n}\left(\frac{pq^2+p^2q}{pq}\right)\\
&= \left(\frac{q}{p}\right)^{S_n}(q+p) = \left(\frac{q}{p}\right)^{S_n} = M_n, \quad n \geqslant 0.
\end{aligned}
$$

In particular, the expectation of $(M_n)_{n\geqslant 0}$ is constant over time by Proposition 10.5 since it is a martingale, *i.e.* we have

$$
\left(\frac{q}{p}\right)^k = \mathbb{E}[M_0 \mid S_0 = k] = \mathbb{E}[M_n \mid S_0 = k], \qquad k = 0, 1, \ldots, B, \quad n \geqslant 0.
$$

Step 2. The stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$ is also a martingale, as a consequence of the Stopping Time Theorem 10.9.

Step 3. Since the stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$ remains a martingale by the Stopping Time Theorem 10.9, its expected value $\mathbb{E}[M_{\tau\wedge n} \mid S_0 = k]$ is constant in $n \geqslant 0$ by Proposition 10.5. This gives

$$
\left(\frac{q}{p}\right)^k = \mathbb{E}[M_0 \mid S_0 = k] = \mathbb{E}[M_{\tau\wedge n} \mid S_0 = k].
$$

Next, letting n go to infinity, by dominated convergence we find*

$$
\begin{aligned}
\mathbb{E}[M_\tau \mid S_0 = k] &= \mathbb{E}\left[\lim_{n\to\infty} M_{\tau\wedge n} \,\Big|\, S_0 = k\right]\\
&= \lim_{n\to\infty} \mathbb{E}[M_{\tau\wedge n} \mid S_0 = k]\\
&= \mathbb{E}[M_0 \mid S_0 = k]\\
&= \left(\frac{q}{p}\right)^k,
\end{aligned}
$$

hence

$$
\begin{aligned}
\left(\frac{q}{p}\right)^k &= \mathbb{E}[M_\tau \mid S_0 = k]\\
&= \left(\frac{q}{p}\right)^B \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^B \,\Bigg|\, S_0 = k\right) + \left(\frac{q}{p}\right)^0 \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^0 \,\Bigg|\, S_0 = k\right)\\
&= \left(\frac{q}{p}\right)^B \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^B \,\Bigg|\, S_0 = k\right) + \mathbb{P}(M_\tau = 1 \mid S_0 = k).
\end{aligned}
$$

Solving the system of equations

* Since $\mathbb{P}(\tau < \infty) = 1$, see Exercise 10.1.

$$\begin{cases} \left(\frac{q}{p}\right)^k = \left(\frac{q}{p}\right)^B \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^B \,\Big|\, S_0 = k\right) + \mathbb{P}(M_\tau = 1 \mid S_0 = k) \\ \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^B \,\Big|\, S_0 = k\right) + \mathbb{P}(M_\tau = 1 \mid S_0 = k) = 1, \end{cases}$$

gives

$$\begin{aligned} \mathbb{P}(S_\tau = B \mid S_0 = k) &= \mathbb{P}\left(M_\tau = \left(\frac{q}{p}\right)^B \,\Big|\, S_0 = k\right) \\ &= \frac{(q/p)^k - 1}{(q/p)^B - 1}, \qquad k = 0, 1, \ldots, B, \end{aligned} \tag{10.4.4}$$

and

$$\begin{aligned} \mathbb{P}(S_\tau = 0 \mid S_0 = k) &= \mathbb{P}(M_\tau = 1 \mid S_0 = k) \\ &= 1 - \frac{(q/p)^k - 1}{(q/p)^B - 1}, \\ &= \frac{(q/p)^B - (q/p)^k}{(q/p)^B - 1}, \end{aligned}$$

$k = 0, 1, \ldots, B$, which recovers (2.2.12).

10.5 Mean Game Duration

In this section we show that the mean game durations $\mathbb{E}[\tau \mid S_0 = k]$ computed in Section 2.3 can also be recovered as a second application of the Stopping Time Theorem 10.9.

Unbiased case $p = q = 1/2$

In the unbiased case of a fair game with $p = q = 1/2$, the martingale method can be used by noting that $\left(S_n^2 - n\right)_{n \geqslant 0}$ is also a martingale.

Step 1. The process $\left(S_n^2 - n\right)_{n \geqslant 0}$ is a martingale.

We have

$$\begin{aligned} &\mathbb{E}[S_{n+1}^2 - (n+1) \mid \mathcal{F}_n] = \mathbb{E}[(S_{n+1} - S_n + S_n)^2 - (n+1) \mid \mathcal{F}_n] \\ &= \mathbb{E}[S_n^2 + (S_{n+1} - S_n)^2 + 2S_n(S_{n+1} - S_n) - (n+1) \mid \mathcal{F}_n] \\ &= \mathbb{E}[S_n^2 - n - 1 \mid \mathcal{F}_n] + \mathbb{E}[(S_{n+1} - S_n)^2 \mid \mathcal{F}_n] + 2\mathbb{E}[S_n(S_{n+1} - S_n) \mid \mathcal{F}_n] \end{aligned}$$

$$
\begin{aligned}
&= S_n^2 - n - 1 + \mathbb{E}[(S_{n+1}-S_n)^2 \mid \mathcal{F}_n] + 2S_n\mathbb{E}[S_{n+1}-S_n \mid \mathcal{F}_n] \\
&= S_n^2 - n - 1 + \underbrace{\mathbb{E}[(S_{n+1}-S_n)^2]}_{=1} + 2S_n \underbrace{\mathbb{E}[S_{n+1}-S_n]}_{=0} \\
&= S_n^2 - n, \quad n \geqslant 0,
\end{aligned}
$$

since $\mathbb{E}[S_{n+1}-S_n] = 0$ and $\mathbb{E}[(S_{n+1}-S_n)^2] = 1$.

Step 2. The stopped process $\left(S_{\tau\wedge n}^2 - \tau\wedge n\right)_{n\geqslant 0}$ is also a martingale, as a consequence of the Stopping Time Theorem 10.9.

Step 3. Since the stopped process $\left(S_{\tau\wedge n}^2 - \tau\wedge n\right)_{n\geqslant 0}$ is also a martingale by the Stopping Time Theorem 10.9, its expectation $\mathbb{E}[S_{\tau\wedge n}^2 - \tau\wedge n \mid S_0 = k]$ is constant in $n \geqslant 0$ by Proposition 10.5, hence we have

$$
k^2 = \mathbb{E}[S_0^2 - 0 \mid S_0 = k] = \mathbb{E}\big[S_{\tau\wedge n}^2 - \tau\wedge n \,\big|\, S_0 = k\big],
$$

and since $\mathbb{P}(\tau < \infty) = 1$, after taking the limit as n tends to infinity and using dominated convergence, we find

$$
\begin{aligned}
\mathbb{E}\big[S_\tau^2 - \tau \mid S_0 = k\big] &= \mathbb{E}\Big[\lim_{n\to\infty} S_{\tau\wedge n}^2 - \lim_{n\to\infty} \tau\wedge n \,\Big|\, S_0 = k\Big] \\
&= \mathbb{E}\left[\lim_{n\to\infty} S_{\tau\wedge n}^2 \mid S_0 = k\right] - \mathbb{E}\left[\lim_{n\to\infty} \tau\wedge n \mid S_0 = k\right] \\
&= \lim_{n\to\infty} \mathbb{E}\big[S_{\tau\wedge n}^2 \mid S_0 = k\big] - \lim_{n\to\infty} \mathbb{E}[\tau\wedge n \mid S_0 = k] \\
&= \lim_{n\to\infty} \mathbb{E}\big[S_{\tau\wedge n}^2 - \tau\wedge n \,\big|\, S_0 = k\big] \\
&= k^2,
\end{aligned}
$$

by application of the *dominated* and *monotone* convergence theorems, since $S_{\tau\wedge n}^2 \in [0, B^2]$ for all $n \geqslant 0$ and the mapping $n \mapsto \tau\wedge n$ is non-decreasing in $n \geqslant 0$. This now yields

$$
\begin{aligned}
k^2 &= \mathbb{E}\big[S_\tau^2 - \tau \mid S_0 = k\big] \\
&= \mathbb{E}\big[S_\tau^2 \mid S_0 = k\big] - \mathbb{E}[\tau \mid S_0 = k] \\
&= B^2\mathbb{P}(S_\tau = B \mid S_0 = k) + 0^2 \times \mathbb{P}(S_\tau = 0 \mid S_0 = k) - \mathbb{E}[\tau \mid S_0 = k],
\end{aligned}
$$

i.e.

$$
\begin{aligned}
\mathbb{E}[\tau \mid S_0 = k] &= B^2\mathbb{P}(S_\tau = B \mid S_0 = k) - k^2 \\
&= B^2\frac{k}{B} - k^2 \\
&= k(B-k),
\end{aligned}
$$

$k = 0, 1, \ldots, B$, which recovers (2.3.19).

Biased case $p \neq q$

Finally, we show how to recover the value of the mean game duration, *i.e.* the mean hitting time of the boundaries $\{0, B\}$, in the biased non-symmetric case $p \neq q$.

Step 1. The process $S_n - (p-q)n$ is a martingale.

In this case we note that although $(S_n)_{n\geqslant 0}$ does not have centered increments and is not a martingale, the compensated process

$$S_n - (p-q)n, \qquad n \geqslant 0,$$

is a martingale because, in addition to being independent, its increments are centered random variables:

$$\mathbb{E}[S_{n+1} - S_n - (p-q)] = \mathbb{E}[S_{n+1} - S_n] - (p-q) = 0, \qquad n \geqslant 0,$$

by (10.4.3).

Step 2. The stopped process $(S_{\tau\wedge n} - (p-q)(\tau \wedge n))_{n\geqslant 0}$ is also a martingale, as a consequence of the Stopping Time Theorem 10.9.

Step 3. The expectation $\mathbb{E}[S_{\tau\wedge n} - (p-q)(\tau \wedge n) \mid S_0 = k]$ is constant in $n \geqslant 0$.

Step 4. Since the stopped process $(S_{\tau\wedge n} - (p-q)(\tau \wedge n))_{n\geqslant 0}$ is a martingale by the Stopping Time Theorem 10.9, we have

$$k = \mathbb{E}[S_0 - 0 \mid S_0 = k] = \mathbb{E}[S_{\tau\wedge n} - (p-q)(\tau \wedge n) \mid S_0 = k],$$

and, since $\mathbb{P}(\tau < \infty) = 1$, after taking the limit as n goes to infinity we find

$$\begin{aligned} \mathbb{E}[S_\tau - (p-q)\tau \mid S_0 = k] &= \lim_{n\to\infty} \mathbb{E}[S_{\tau\wedge n} - (p-q)(\tau \wedge n) \mid S_0 = k] \\ &= \mathbb{E}\Big[\lim_{n\to\infty} S_{\tau\wedge n} - (p-q) \lim_{n\to\infty} \tau \wedge n \, \Big| \, S_0 = k \Big] \\ &= k, \end{aligned}$$

which gives

$$\begin{aligned} k &= \mathbb{E}[S_\tau - (p-q)\tau \mid S_0 = k] \\ &= \mathbb{E}[S_\tau \mid S_0 = k] - (p-q)\mathbb{E}[\tau \mid S_0 = k] \\ &= B \times \mathbb{P}(S_\tau = B \mid S_0 = k) + 0 \times \mathbb{P}(S_\tau = 0 \mid S_0 = k) - (p-q)\mathbb{E}[\tau \mid S_0 = k], \end{aligned}$$

i.e.

$$\begin{aligned} (p-q)\mathbb{E}[\tau \mid S_0 = k] &= B \times \mathbb{P}(S_\tau = B \mid S_0 = k) - k \\ &= B \frac{(q/p)^k - 1}{(q/p)^B - 1} - k, \end{aligned}$$

from (10.4.4), hence

$$\mathbb{E}[\tau \mid S_0 = k] = \frac{1}{p-q}\left(B\frac{(q/p)^k - 1}{(q/p)^B - 1} - k\right), \qquad k = 0, 1, \ldots, B,$$

which recovers (2.3.12). In Table 10.1, we summarize the family of martingales used to treat the above problems.

Problem \ Probabilities	Unbiased	Biased
Ruin probability	S_n	$\left(\frac{q}{p}\right)^{S_n}$
Mean game duration	$S_n^2 - n$	$S_n - (p-q)n$

Table 10.1: List of martingales.

The above examples can be summarized in Figure 10.5, which can be generated using the following R code.

```
nsim <- 100;a=-2;b=6;x=1;mu=0;N=21; T<-2.0; t <- 0:(N-1); dt <- T/N; prob=0; time=0;
dev.new(width=16,height=8); for (i in 1:nsim){signal=0;colour="blue";Z <-
    2*(rbinom(N,1,0.5)-0.5);
tau=N+1; X <- c(1,N);X[1]=x;for (j in 2:N){X[j]=X[j-1]+Z[j]+mu*dt
if (X[j]<=a && signal==0) {xtau=a;tau=j-1;signal=-1;colour="purple";time=time+j}
if (X[j]>=b && signal==0)
    {xtau=b;tau=j-1;signal=1;colour="blue";prob=prob+1;time=time+j}}
plot(t, X, xlab = "n", ylab = "", type = "l",lwd=3, ylim = c(a-2,b+2), col = "blue",
    xaxs="i",yaxs="i", xaxt="n", yaxt="n",cex.axis=1.8,cex.lab=1.8,cex.main=2)
if (N>tau) {lines(t, c(X[1:tau],rep(xtau,N-tau))-signal*0.07, xlab = "n", ylab = "", type =
    "l",lwd=3, ylim = c(a-2,b+2), col = "red", xaxs="i",yaxs="i", xaxt="n",
    yaxt="n",cex.axis=1.8,cex.lab=1.8,cex.main=2)}
points(t, X, xlab = "t", ylab = "", type = "p", pch = 16, cex=1.6, ylim = c(-6,12), col = "blue",
    xaxs="i",yaxs="i", xaxt="n", yaxt="n",cex.axis=1.8,cex.lab=1.8,cex.main=2);
    yticks<-c(a,x,b); axis(side=2, mgp=c(1,1,0), at=yticks,labels = c("a=
    ",as.expression(bquote('M'['0']~"= ")),"b= "), las = 2,cex.axis=1.3)
axis(side=2, at = seq(a-2, b+2, by = 1), las = 2,cex.axis=1.3)
axis(side=1, at = seq(0, N, by = 1), las = 2,cex.axis=1.3)
axis(side=1, at = c(tau), labels = expression(paste(tau,"= ")),las = 2,cex.axis=1.3)
abline(h=x,lw=2); abline(h=a,col="black",lwd=3); abline(h=b,col="black",lwd=3)
abline(v=tau,col="purple",lwd=3); abline(v=tau,col="purple",lwd=3);
text(1.17, 5.5, as.expression(bquote('M'[tau~"\u2227 10"]~"= ")),cex=2)
text(2.2, 5.5, X[min(10,tau)+1],cex=2)
text(0.75, 4.5, as.expression(bquote('M'[tau]~"= ")),cex=2); text(1.3, 4.5, X[tau+1],cex=2)
text(0.75, 3.5, as.expression(bquote('M'['5']~"= ")),cex=2); text(1.3, 3.5, X[6],cex=2)
text(0.75, 2.5, as.expression(bquote('M'['2']~"= ")),cex=2); text(1.3, 2.5, X[3],cex=2)
legend(-0.3, 8.35, legend=c(as.expression(bquote('Process M'[n])),
    as.expression(bquote('Stopped process M'[tau~"\u2227 n"]))), col=c("blue", "red"),
    lty=1:1, lwd=3, box.lty=0, cex=2)
grid(N-1,b-a+4,col="black"); grid(N-1,b-a+4,col="black")
readline(prompt = "Pause. Press <Enter> to continue...")}
```

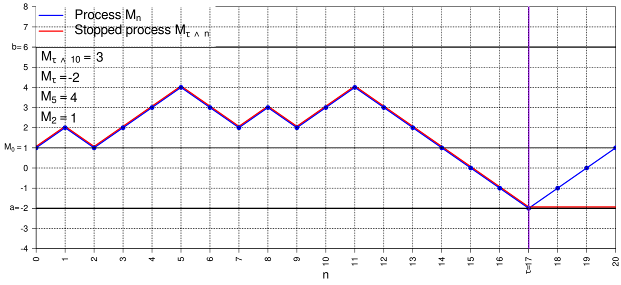


Fig. 10.5: Stopped process.

Exercises

Exercise 10.1 (Steele (2001), page 3). Recover (2.2.30) by showing that for all $k = 0, 1, \ldots, B$ we have $\mathbb{P}(\tau_{0,B} < \infty \mid S_0 = k) = 1$, *i.e.* the stopping time $\tau_{0,B}$ defined in (10.4.1) is finite almost surely.

Exercise 10.2 Doubling down. Consider a sequence $(X_n)_{n\geqslant 1}$ of independent Bernoulli random variables, with

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}, \qquad n \geqslant 1,$$

and the process $(M_n)_{n\geqslant 0}$ defined by $M_0 := 0$ and

$$M_n := \sum_{k=1}^{n} 2^{k-1} X_k, \qquad n \geqslant 1,$$

with in particular

$$\begin{cases} M_1 = X_1, \\ M_2 = X_1 + 2X_2, \\ M_3 = X_1 + 2X_2 + 4X_3, \\ \vdots \end{cases}$$

see Figure 10.6.

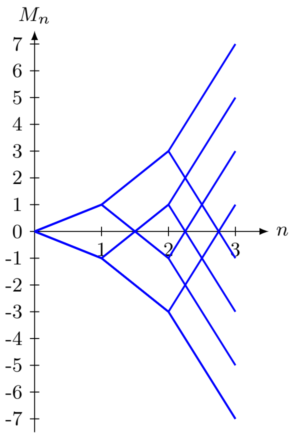


Fig. 10.6: Possible paths of the process $(M_n)_{n\geqslant 0}$.

Note that when $X_1 = X_2 = \cdots = X_{n-1} = -1$ and $X_n = 1$, we have

$$M_n = -\sum_{k=1}^{n-1} 2^{k-1} + 2^{n-1} = -\frac{1-2^{n-1}}{1-2} + 2^{n-1} = 1, \qquad n \geqslant 1,$$

while when $X_1 = X_2 = \cdots = X_{n-1} = X_n = -1$, we have

$$M_n = -\sum_{k=1}^{n} 2^{k-1} = -\frac{1-2^n}{1-2} = 1 - 2^n, \qquad n \geqslant 1.$$

a) Show that the process $(M_n)_{n\geqslant 0}$ is a martingale.
b) Is the random time

$$\tau := \inf\{n \geqslant 1 \ : \ M_n = 1\}$$

a stopping time?
c) Consider the stopped process

$$M_{\tau\wedge n} := M_n \mathbb{1}_{\{n<\tau\}} + \mathbb{1}_{\{\tau\leqslant n\}} = \begin{cases} M_n = 1 - 2^n & \text{if } n < \tau, \\ M_\tau = 1 & \text{if } n \geqslant \tau, \end{cases}$$

$n \geqslant 0$, see Figure 10.7. Give an interpretation of $(M_{n\wedge\tau})_{n\geqslant 0}$ in terms of betting strategy for a gambler starting a game at $M_0 = 0$.

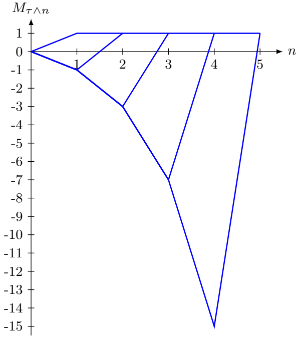


Fig. 10.7: Possible paths of the stopped process $(M_{\tau\wedge n})_{n\geqslant 0}$.

d) Determine the two possible values of $M_{\tau\wedge n}$ and the probability distribution of $M_{\tau\wedge n}$ at any time $n \geqslant 1$.

e) Show, using the result of Question (d), that we have

$$\mathbb{E}[M_{\tau\wedge n}] = 0, \qquad n \geqslant 0.$$

f) Show that the result of Question (e) can be recovered using the Stopping Time Theorem 10.9.

Exercise 10.3 Let $(S_n)_{n\geqslant 0}$ be the random walk defined by $S_0 := 0$ and

$$S_n := \sum_{k=1}^{n} X_k = X_1 + \cdots + X_n, \qquad n \geqslant 1,$$

where $(X_n)_{n\geqslant 1}$ is an i.i.d. Bernoulli sequence of $\{-1,1\}$-valued random variables with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$, $n \geqslant 1$. We consider the random time

$$\tau := \inf\{n \geqslant 1 \ : \ S_n - S_{n-1} = -1\}$$

at which $(S_n)_{n\geqslant 0}$ drops for the first time, and decide to use τ as an exit strategy.

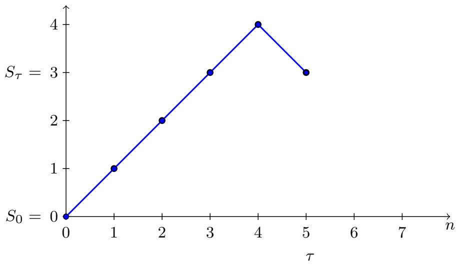


Fig. 10.9: Sample path of the random walk $(S_n)_{n\geqslant 0}$.

a) Show that τ is a stopping time.
b) Find the range of possible values of S_τ.
c) Give the value of $\mathbb{E}[S_\tau]$ according to the stopping time theorem.
d) Find the probability distribution of S_τ, *i.e.* find $\mathbb{P}(S_\tau = k)$ for all $k \in \mathbb{Z}$.
e) Compute

$$\mathbb{E}[S_\tau] = \sum_{k\in\mathbb{Z}} k\mathbb{P}(S_\tau = k),$$

and recover the conclusion of part (c).

Exercise 10.4 Show that, as the discrete-time filtration $(\mathcal{F}_n)_{n\geqslant 0}$ satisfies $\mathcal{F}_{n-1} \subset \mathcal{F}_n$, $n \geqslant 1$, Condition (10.3.1) is equivalent to

$$\{\tau = n\} \in \mathcal{F}_n, \qquad n \geqslant 0. \tag{10.5.1}$$

Exercise 10.5 Give an example of an *a.s.* converging and unbounded sequence $(X_n)_{n\geqslant 0}$ of random variables for which expectation and limit cannot be exchanged, *i.e.*

$$\lim_{n\to\infty} \mathbb{E}[X_n] \neq \mathbb{E}[\lim_{n\to\infty} X_n].$$

Exercise 10.6 Let $(M_n)_{n\geqslant 0}$ be a discrete-time *sub*martingale with respect to a filtration $(\mathcal{F}_n)_{n\geqslant 0}$, with $\mathcal{F}_0 = \{\emptyset, \Omega\}$, *i.e.* we have

$$M_n \leqslant \mathbb{E}[M_{n+1} \mid \mathcal{F}_n], \qquad n \geqslant 0.$$

a) Show that for all $n \geqslant 0$, we have $\mathbb{E}[M_n] \leqslant \mathbb{E}[M_{n+1}]$, *i.e.* *sub*martingales have a *non-decreasing* expected value.
b) Show that independent increment processes whose increments have non-negative expectation are examples of *sub*martingales.
c) (Doob-Meyer decomposition). Show that there exists two processes $(N_n)_{n\geqslant 0}$ and $(A_n)_{n\geqslant 0}$ such that

i) $(N_n)_{n\geqslant 0}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geqslant 0}$,

ii) $(A_n)_{n\geqslant 0}$ is *non-decreasing*, *i.e.* $A_n \leqslant A_{n+1}$, *a.s.*, $n \geqslant 0$,

iii) $(A_n)_{n\geqslant 0}$ is predictable in the sense that A_n is $\mathcal{F}_{n-1}$-measurable, $n \geqslant 1$, and

iv) $M_n = N_n + A_n$, $n \in \mathbb{N}$.

Hint: Let $A_0 := 0$, $A_1 := A_0 + \mathbb{E}[M_1 - M_0 \mid \mathcal{F}_0]$, and

$$A_{n+1} := A_n + \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n], \quad n \geqslant 0,$$

and define $(N_n)_{n\geqslant 0}$ in such a way that it satisfies the four required properties.

d) Show that for all bounded stopping times σ and τ such that $\sigma \leqslant \tau$ *a.s.*, we have

$$\mathbb{E}[M_\sigma] \leqslant \mathbb{E}[M_\tau].$$

Hint: Use the Doob Stopping Time Theorem 10.9 for martingales and (10.3.3).

Exercise 10.7 We say that a function $\phi : \mathbb{R} \to \mathbb{R}$ is *convex* if

$$\phi(px + qy) \leqslant p\phi(x) + q\phi(y), \qquad x, y \in \mathbb{R},$$

for all $p, q \in [0, 1]$ such that $p + q = 1$.

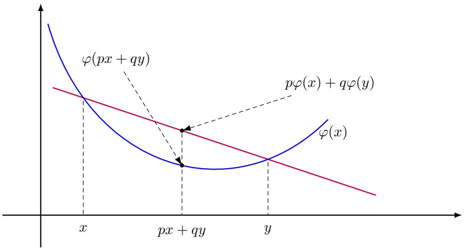


Fig. 10.10: Convex function.

a) Show that for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ we have the inequality

$$\phi\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \leqslant \frac{\phi(x_1) + \phi(x_2) + \cdots + \phi(x_n)}{n}, \qquad (10.5.2)$$

$x_1, \ldots, x_n \in \mathbb{R}$, $n \geqslant 1$.

b) Consider a martingale $(S_n)_{n=0,1,\ldots,N}$ under the probability measure $\mathbb{P}$, with respect to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$, and let ϕ be a convex function. Show an inequality between the quantities

$$\mathbb{E}\left[\phi\left(\frac{S_1+S_2+\cdots+S_N}{N}\right)\right] \quad \text{and} \quad \mathbb{E}[\phi(S_N)].$$

Hint: Use in the following order:

(i) the convexity inequality (10.5.2),
(ii) the martingale property of $(S_k)_{k\geqslant 0}$,
(iii) the conditional Jensen inequality $\phi(\mathbb{E}[F \mid \mathcal{G}]) \leqslant \mathbb{E}[\phi(F) \mid \mathcal{G}]$,
(iv) the tower property of conditional expectations.

c) Given the (convex) function $\phi(x) := (x-K)^+$, show that the price

$$\mathbb{E}\left[\left(\frac{S_1+S_2+\cdots+S_N}{N}-K\right)^+\right]$$

of a discrete-time arithmetic average option with payoff

$$\left(\frac{S_1+S_2+\cdots+S_N}{N}-K\right)^+$$

is upper bounded by the price of the European *call* option with payoff $(S_N-K)^+$ and maturity N.

Exercise 10.8 A stochastic process $(M_n)_{n\geqslant 0}$ is a *sub*martingale if it satisfies

$$M_k \leqslant \mathbb{E}[M_n \mid \mathcal{F}_k], \qquad k=0,1,\ldots,n.$$

a) Show that the expected value $\mathbb{E}[M_n]$ of the *sub*martingale $(M_n)_{n\geqslant 0}$ is non-decreasing in time $n \geqslant 0$.
b) Consider the random walk given by $S_0 := 0$ and

$$S_n := \sum_{k=1}^{n} X_k = X_1+X_2+\cdots+X_n, \qquad n \geqslant 1,$$

where $(X_n)_{n\geqslant 1}$ is an *i.i.d.* Bernoulli sequence of $\{0,1\}$-valued random variables with $\mathbb{P}(X_n=1)=p$, $n\geqslant 1$. Under which condition on $\alpha\in\mathbb{R}$ is the process $(S_n-\alpha n)_{n\geqslant 0}$ a *sub*martingale?

Exercise 10.9 Recall that a discrete-time stochastic process $(M_n)_{n\geqslant 0}$ is a *sub*martingale with respect to a filtration $(\mathcal{F}_n)_{n\geqslant 0}$ if it satisfies

$$M_k \leqslant \mathbb{E}[M_n \mid \mathcal{F}_k], \qquad k=0,1,\ldots,n.$$

a) Show that any convex function $(\phi(M_n))_{n\geqslant 0}$ of a martingale $(M_n)_{n\geqslant 0}$ is itself a *sub*martingale. *Hint*: Use Jensen's inequality.
b) Show that any convex non-decreasing function $\phi(M_n)$ of a *sub*martingale $(M_n)_{n\geqslant 0}$ remains a *sub*martingale.

Problem 10.10

a) Consider $(M_n)_{n\geqslant 0}$ a *nonnegative martingale* generating the filtration $(\mathcal{F}_n)_{n\geqslant 0}$. For any $x>0$, let

$$\tau_x := \inf\{n \geqslant 0 \ : \ M_n \geqslant x\}.$$

Show that the random time τ_x is a stopping time.
b) Show that for all $n \geqslant 0$, we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[M_n]}{x}, \qquad x>0. \tag{10.5.3}$$

Hint: Proceed as in the proof of the classical Markov inequality and use the Doob Stopping Time Theorem 10.9 for the stopping time τ_x.
c) Show that (10.5.3) remains valid for $(M_n)_{n\geqslant 0}$ a nonnegative *sub*martingale.

Hint: Use the Doob Stopping Time Theorem 10.9 for *sub*martingales as in Exercise 10.6-(d).
d) Show that if $(M_n)_{n\geqslant 0}$ a nonnegative *sub*martingale, then for any $n \geqslant 0$ we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[(M_n)^2]}{x^2}, \qquad x>0.$$

e) Show that more generally, we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[(M_n)^p]}{x^p}, \qquad x>0,$$

for all $n \geqslant 0$ and $p \geqslant 1$.
f) Given $(Y_n)_{n\geqslant 1}$ a sequence of centered independent random variables with same mean $\mathbb{E}[Y_n]=0$ and variance $\sigma^2 = \operatorname{Var}[Y_n]$, $n \geqslant 1$, consider the random walk $S_n = Y_1 + Y_2 + \cdots + Y_n$, $n \geqslant 1$, with $S_0 = 0$.

Show that for all $n \geqslant 0$, we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} |S_k| \geqslant x \Big) \leqslant \frac{n\sigma^2}{x^2}, \qquad x>0.$$

g) Show that for any (not necessarily nonnegative) *sub*martingale, we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[(M_n)^+]}{x}, \qquad x>0,$$

where $(z)^+ := \mathrm{Max}(z,0)$, $z \in \mathbb{R}$.

h) A stochastic process $(M_n)_{n\geqslant 0}$ is a *super*martingale* if it satisfies

$$\mathbb{E}[M_n \mid \mathcal{F}_k] \leqslant M_k, \qquad k = 0,1,\ldots,n.$$

Show that for any *nonnegative super*martingale we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x\Big) \leqslant \frac{\mathbb{E}[M_0]}{x}, \qquad x > 0.$$

i) Show that for any *nonnegative submartingale* $(M_n)_{n\geqslant 0}$ and any convex non-decreasing nonnegative function ϕ we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} \phi(M_k) \geqslant x\Big) \leqslant \frac{\mathbb{E}[\phi(M_n)]}{x}, \qquad x > 0.$$

Hint: Consider the stopping time

$$\tau_x^\phi := \inf\{n \geqslant 0 \ : \ \phi(M_n) \geqslant x\},$$

and use the result of Exercise 10.9-(b).

j) Give an example of a nonnegative *super*martingale which is *not* a martingale.

Exercise 10.11 Consider the random walk $(S_n)_{n\geqslant 0}$ on $\{0,1,\ldots,B\}$ with $S_0 := 0$ and independent $\{-1,1\}$-valued increments $(S_{n+1}-S_n)_{n\geqslant 0}$ such that

$$\mathbb{P}(S_{n+1}-S_n = +1) = p \quad \text{and} \quad \mathbb{P}(S_{n+1}-S_n = -1) = q, \quad n \geqslant 0,$$

and the martingale $(M_n)_{n\geqslant 0}$ defined as

$$M_n := \left(\frac{q}{p}\right)^{S_n}, \qquad n \geqslant 0,$$

where $p, q \in (0,1)$ are such that $p+q=1$. Show that for all $n \geqslant 0$ and $r \geqslant 1$, we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x\Big) \leqslant \frac{(p(q/p)^r + q(p/q)^r)^n}{x^r}, \qquad x > 0.$$

* "This obviously inappropriate nomenclature was chosen under the malign influence of the noise level of radio's SUPERman program, a favorite supper-time program of Doob's son during the writing of Doob (1953)", cf. Doob (1984), historical notes, page 808.

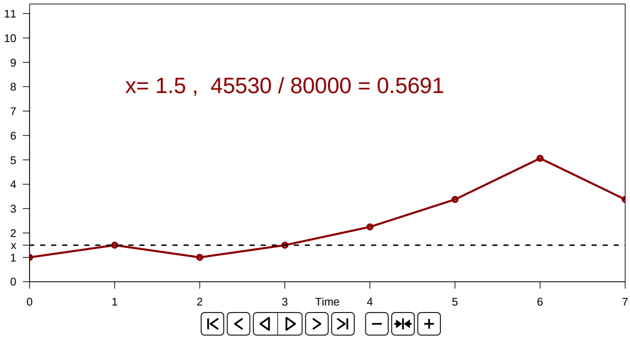


Fig. 10.11: Random walk supremum.*

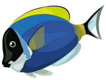
* The animation works in Acrobat Reader on the entire pdf file.

Chapter 11
Spatial Poisson Processes

Spatial Poisson point processes are typically used to model the random scattering of configuration points within a plane or a three-dimensional space $\mathbb{X}$. In case $\mathbb{X} = \mathbb{R}_+$ is the real half line, these random points can be identified with the jump times $(T_k)_{k\geqslant 1}$ of the standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$ introduced in Section 9.1. However, in contrast with the previous chapter, no time ordering is *a priori* imposed here on the index set $\mathbb{X}$. Sections 11.4 and 11.5 contain some more advanced results on moments and deviation inequalities for Poisson stochastic integrals.

11.1 Spatial Poisson (1781-1840) Processes

In this section, we present the construction of spatial Poisson point processes on the space



$$\Omega^{\mathbb{X}} := \left\{ \omega := (x_i)_{i=1}^{N} \subset \mathbb{X},\ N \in \mathbb{N} \cup \{\infty\} \right\}$$

of subsets of $\mathbb{X} \subset \mathbb{R}^d$ called *configurations*, $d \geqslant 1$.

The next figure illustrates a given configuration $\omega \in \Omega^{\mathbb{X}}$.

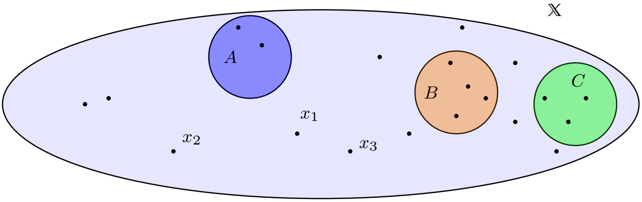


Fig. 11.1: Poisson random samples with $\omega(A) = 2$, $\omega(B) = 4$, $\omega(C) = 3$.

Definition 11.1. *Given a (measurable) subset A of $\mathbb{X}$, we let*

$$\omega(A) = \#\{x \in \omega \ : \ x \in A\} = \sum_{x \in \omega} \mathbb{1}_A(x)$$

denote the number of configuration points in ω that are contained in the set A.

We consider an intensity measure $\sigma(dx)$ on $\mathbb{X}$, possibly given from a non-negative density function $\rho : \mathbb{X} \longrightarrow \mathbb{R}_+$ as $\sigma(dx) = \rho(x)dx$, *i.e.* for any (measurable) subset A of $\mathbb{X}$ we have

$$\begin{aligned} \sigma(A) &= \int_A \sigma(dx) \\ &= \int_A \rho(x)dx \\ &= \int_{\mathbb{X}} \mathbb{1}_A(x)\rho(x)dx. \end{aligned}$$

When $\sigma(\mathbb{X}) < \infty$, the Poisson point process with intensity $\sigma(dx)$ can be constructed in three steps:

1. First, choose the number $\omega(\mathbb{X})$ of points in $\mathbb{X}$ according to a standard Poisson distribution with mean $\sigma(\mathbb{X})$:

$$\mathbb{P}_\sigma^{\mathbb{X}}(\omega(\mathbb{X}) = n) = \mathrm{e}^{-\sigma(\mathbb{X})} \frac{(\sigma(\mathbb{X}))^n}{n!}, \qquad n \geqslant 0.$$

2. Second, scatter $n = \omega(\mathbb{X})$ points $(X_1 \ldots, X_n)$ over $\mathbb{X}$ independently, each of them with the probability distribution $\sigma(dx)/\sigma(\mathbb{X})$, *i.e.*

$$\mathbb{P}_\sigma^{\mathbb{X}}((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n \mid \omega(\mathbb{X}) = n) = \frac{\sigma(A_1)}{\sigma(\mathbb{X})} \cdots \frac{\sigma(A_n)}{\sigma(\mathbb{X})}, \tag{11.1.1}$$

for $A_1, \ldots, A_n$ measurable subsets of $\mathbb{X}$ with finite σ-measure.

```
library(spatstat)
lambda = 10000
bellcurve <- function(x,y,s){return(exp(-s*((x-0.5)**2+(y-0.5)**2)))}
rho <- function(x,y){lambda*bellcurve(x+0.2,y+0.2,70)+lambda*bellcurve(x-0.2,y-0.1,40)}
X <- rpoispp(rho)
plot(X, cols="blue", pch=16, cex=0.7, main = '')
```

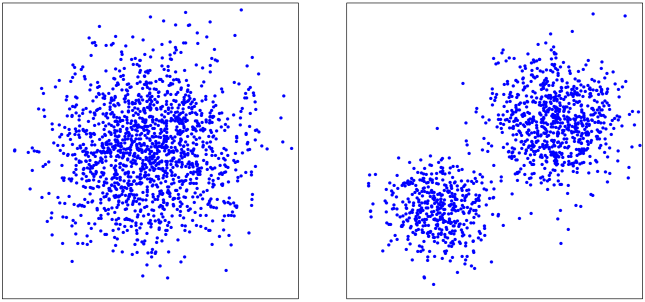
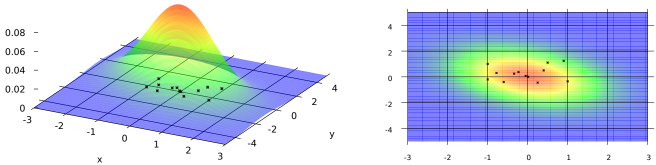


Fig. 11.2: Two Poisson point process samples.

Figure 11.3 presents another Poisson point process sample together with the density of its intensity measure.



(a) Sample points *vs.* intensity $\rho(x)$.

(b) View from above (heat map).

Fig. 11.3: Poisson point process sample on the plane.

The Poisson probability measure $\mathbb{P}_\sigma^{\mathbb{X}}$ with intensity $\sigma(dx) = \rho(x)dx$ on $\mathbb{X}$ satisfying the above points 1 and 2 can be characterized in the next theorem, see Proposition I.6 in Neveu (1977).

Theorem 11.2. *Given $\rho : \mathbb{X} \longrightarrow \mathbb{R}_+$ a nonnegative function, the Poisson probability measure $\mathbb{P}_\sigma^{\mathbb{X}}$ with intensity $\sigma(dx) = \rho(x)dx$ on $\mathbb{X}$ is the only probability measure on $\Omega^{\mathbb{X}}$ satisfying the following two properties:*

i) For any (measurable) subset A of $\mathbb{X}$ such that $\sigma(A) < \infty$, the number $\omega(A)$ of configuration points contained in A is a Poisson random variable with intensity $\sigma(A)$, i.e.

$$\mathbb{P}_\sigma^{\mathbb{X}}(\omega \in \Omega^{\mathbb{X}} \ : \ \omega(A) = n) = \mathrm{e}^{-\sigma(A)} \frac{(\sigma(A))^n}{n!}, \qquad n \geqslant 0.$$

ii) For any sequence $A_1, A_2, \ldots, A_n$ are disjoint measurable subsets of $\mathbb{X}$ with $\sigma(A_k) < \infty$, $k = 1, 2, \ldots, n$, the $\mathbb{N}^n$-valued random vector

$$\omega \longmapsto (\omega(A_1), \ldots, \omega(A_n)), \qquad \omega \in \Omega^{\mathbb{X}},$$

is made of independent random variables for all $n \geqslant 1$.

In the remaining of this chapter, we will assume for simplicity that $\sigma(\mathbb{X}) < \infty$. The Poisson measure $\mathbb{P}_\sigma^{\mathbb{X}}$ can also be defined from

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}[F] = \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n \geqslant 0} \frac{1}{n!} \int_{\mathbb{X}^n} f_n(x_1, x_2, \ldots, x_n) \sigma(dx_1) \cdots \sigma(dx_n) \quad (11.1.2)$$

for F written as

$$F(\omega) = f_0 \mathbb{1}_{\{\omega(\mathbb{X})=0\}} + \sum_{n \geqslant 1} \mathbb{1}_{\{\omega(\mathbb{X})=n\}} f_n(x_1, x_2, \ldots, x_n)$$

where f_n is a symmetric integrable function of $\omega = \{x_1, x_2, \ldots, x_n\}$ when $\omega(\mathbb{X}) = n$, $n \geqslant 1$, cf. *e.g.* Proposition 6.1.3 and § 6.1 in Privault (2009).

In many applications, the intensity function $\rho(x)$ can be constant, *i.e.* $\rho(x) = \lambda > 0$, $x \in \mathbb{X}$, where $\lambda > 0$ is called the intensity parameter, and

$$\sigma(A) = \lambda \int_A dx = \lambda \int_{\mathbb{X}} \mathbb{1}_A(x) dx$$

represents the surface or volume of A in $\mathbb{R}^d$. In this case, (11.1.1) can be used to show that the random points $\{x_1, \ldots, x_n\}$ are uniformly distributed on A^n given that $\{\omega(A) = n\}$.

11.2 Poisson Stochastic Integrals

In the next proposition we consider the Poisson stochastic integral defined as

$$\int_{\mathbb{X}} f(x) \omega(dx) := \sum_{x \in \omega} f(x), \quad (11.2.1)$$

for $f \in L^1(\mathbb{X}, \sigma)$ an integrable function on $\mathbb{X}$, and we compute its first and second order moments and cumulants via its characteristic function.

Proposition 11.3. *Let $f \in L^1(\mathbb{X}, \sigma)$ be an integrable function on $\mathbb{X}$. We have*

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(i \int_{\mathbb{X}} f(x) \omega(dx)\right)\right] = \exp\left(\int_{\mathbb{X}} (\mathrm{e}^{if(x)} - 1) \sigma(dx)\right).$$

Proof. We assume that $\sigma(\mathbb{X}) < \infty$. By (11.1.2) and the definition (11.2.1) of the Poisson stochastic integral, we have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(i\int_{\mathbb{X}} f(x)\omega(dx)\right)\right] \\
&\quad = \mathrm{e}^{-\sigma(\mathbb{X})}\sum_{n\geqslant 0}\frac{1}{n!}\int_{\mathbb{X}}\cdots\int_{\mathbb{X}} \mathrm{e}^{i(f(x_1)+\cdots+f(x_n))}\sigma(dx_1)\cdots\sigma(dx_n). \\
&\quad = \mathrm{e}^{-\sigma(\mathbb{X})}\sum_{n\geqslant 0}\frac{1}{n!}\left(\int_{\mathbb{X}}\mathrm{e}^{if(x)}\sigma(dx)\right)^n \\
&\quad = \exp\left(\int_{\mathbb{X}}(\mathrm{e}^{if(x)}-1)\sigma(dx)\right).
\end{aligned}$$

□

The characteristic function also allows us to compute the expectation of $\int_{\mathbb{X}} f(x)\omega(dx)$ using the relation $i^2 = -1$, as

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\int_{\mathbb{X}} f(x)\omega(dx)\right] &= -i\frac{d}{d\varepsilon}\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(i\varepsilon\int_{\mathbb{X}} f(x)\omega(dx)\right)\right]_{|\varepsilon=0} \\
&= -i\frac{d}{d\varepsilon}\exp\left(\int_{\mathbb{X}}(\mathrm{e}^{i\varepsilon f(x)}-1)\sigma(dx)\right)_{|\varepsilon=0} \\
&= \int_{\mathbb{X}} f(x)\sigma(dx),
\end{aligned}$$

for f an integrable function on $\mathbb{X}$. As a consequence, the *compensated* Poisson stochastic integral

$$\int_{\mathbb{X}} f(x)\omega(dx) - \int_{\mathbb{X}} f(x)\sigma(dx)$$

is a *centered* random variable, *i.e.* we have

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\int_{\mathbb{X}} f(x)\omega(dx) - \int_{\mathbb{X}} f(x)\sigma(dx)\right] = 0.$$

The variance can be similarly computed as

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\left(\int_{\mathbb{X}} f(x)(\omega(dx)-\sigma(dx))\right)^2\right] = \int_{\mathbb{X}} |f(x)|^2\sigma(dx),$$

for all f in the space $L^2(\mathbb{X},\sigma)$ of functions which are square-integrable on $\mathbb{X}$ with respect to $\sigma(dx)$.

More generally, the logarithmic generating function

$$\begin{aligned}
\log \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(\int_{\mathbb{X}} f(x)\omega(dx)\right)\right] &= \int_{\mathbb{X}}(\mathrm{e}^{f(x)}-1)\sigma(dx) \\
&= \sum_{n\geqslant 1}\frac{1}{n!}\int_{\mathbb{X}} f^n(x)\sigma(dx),
\end{aligned}$$

shows that the *cumulants* of $\int_{\mathbb{X}} f(x)\omega(dx)$, see (1.7.11), are given by

$$\kappa_n = \int_{\mathbb{X}} f^n(x)\sigma(dx), \qquad n \geqslant 1. \tag{11.2.2}$$

Example - gamma distribution

When $\mathbb{X} = \mathbb{R}_+$ and $\rho(x) = \lambda e^{-xt}/x$, $\lambda, t > 0$, *i.e.* σ is given by $\sigma(dx) = \rho(x)dx = \lambda e^{-xt}dx/x$, the Poisson stochastic integral $\int_0^\infty x\omega(dx) = \sum_{x\in\omega} x$ has the Laplace transform

$$\begin{aligned}
\mathbb{E}_\sigma\left[\exp\left(-s\int_{\mathbb{X}} x\omega(dx)\right)\right] &= \exp\left(\int_{\mathbb{X}} (e^{-sx}-1)\sigma(dx)\right) \\
&= \exp\left(\lambda\int_{\mathbb{X}} (e^{-sx}-1)e^{-xt}\frac{dx}{x}\right) \\
&= \exp\left(-\lambda\log\left(1+\frac{s}{t}\right)\right) \\
&= \left(1+\frac{s}{t}\right)^{-\lambda} \\
&= \frac{t^\lambda}{\Gamma(\lambda)}\int_0^\infty e^{-sy}y^{\lambda-1}e^{-yt}dy, \qquad s > -t,
\end{aligned}$$

where we used Frullani's identity

$$\log\left(1+\frac{s}{t}\right) = \int_0^\infty (1-e^{-sx})e^{-xt}\frac{dx}{x}, \qquad s, t > 0.$$

This shows that the random variable $\int_0^\infty x\omega(dx) = \sum_{x\in\omega} x$ has the gamma distribution with probability density function

$$y \mapsto \frac{t^\lambda}{\Gamma(\lambda)}y^{\lambda-1}e^{-yt}, \qquad y > 0,$$

shape parameter λ, scaling parameter $1/t$, and mean λ/t.

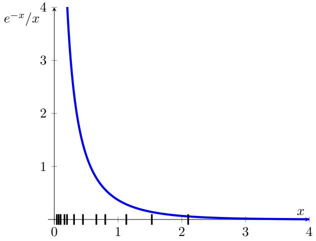


Fig. 11.4: Gamma Lévy density $\rho(x) = \lambda e^{-xt}/x$.

11.3 Transformations of Poisson Measures

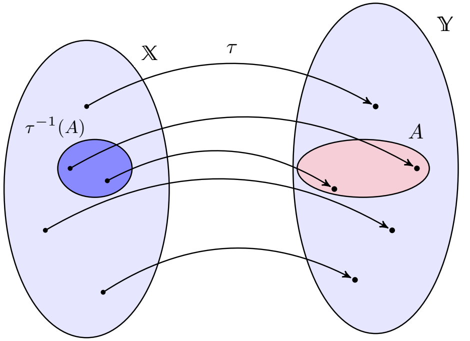
Consider a mapping $\tau : (\mathbb{X}, \sigma) \longrightarrow (\mathbb{Y}, \mu)$, and let

$$\tau_* : \Omega^{\mathbb{X}} \longrightarrow \Omega^{\mathbb{Y}}$$

be the transformed configuration defined by

$$\tau_*\omega = \{\tau(x_1), \tau(x_2), \tau(x_3), \cdots\} := \{\tau(x) \ : \ x \in \omega\}, \qquad \omega \in \Omega^{\mathbb{X}},$$

as illustrated in the following figure.



Proposition 11.4. *Assume that $\tau : \mathbb{X} \to \mathbb{Y}$ is a one-to-one mapping. The random configuration*

$$\begin{aligned} \Omega^{\mathbb{X}} \; : \; &\longrightarrow \Omega^{\mathbb{Y}} \\ \omega &\longmapsto \tau_*(\omega) \end{aligned}$$

has the Poisson distribution $\mathbb{P}^{\mathbb{Y}}_{\mu}$ *with intensity* μ *on* $\mathbb{Y}$*, where* $\mu := \tau_*\sigma$ *is the pushforward of the measure* σ *by* τ*.*

Proof. For any set $A \subset \mathbb{Y}$ of finite μ-measure, we have

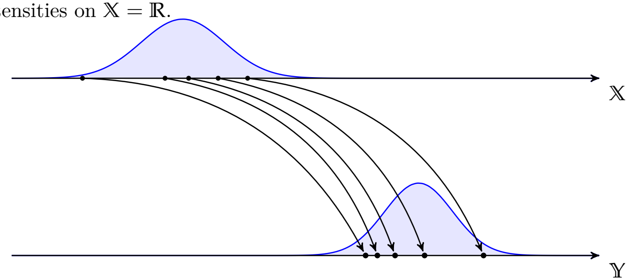
$$\begin{aligned} \mathbb{P}^{\mathbb{X}}_{\sigma}(\{\omega \in \mathbb{X} \ : \ \tau_*\omega(A) = n\}) &= \mathbb{P}^{\mathbb{X}}_{\sigma}(\{\omega \in \mathbb{X} \ : \ \omega(\tau^{-1}(A)) = n\}) \\ &= \mathrm{e}^{-\sigma(\tau^{-1}(A))} \frac{(\sigma(\tau^{-1}(A)))^n}{n!} \\ &= \mathrm{e}^{-\tau_*\sigma(A)} \frac{(\tau_*\sigma(A))^n}{n!} \\ &= \mathrm{e}^{-\mu(A)} \frac{(\mu(A))^n}{n!}. \end{aligned}$$

More generally, we can check that for all families $A_1, A_2, \ldots, A_n$ of disjoint subsets of $\mathbb{Y}$ and $k_1, k_2, \ldots, k_n \in \mathbb{N}$, we have

$$\begin{aligned} &\mathbb{P}^{\mathbb{X}}_{\sigma}\big(\big\{\omega \in \Omega^{\mathbb{X}} \ : \ \tau_*\omega(A_1) = k_1, \ldots, \tau_*\omega(A_n) = k_n\big\}\big) \\ &\quad = \mathbb{P}^{\mathbb{X}}_{\sigma}\big(\big\{\omega \in \Omega^{\mathbb{X}} \ : \ \omega(\tau^{-1}(A_1)) = k_1, \ldots, \tau_*\omega(\tau^{-1}(A_n)) = k_n\big\}\big) \\ &\quad = \prod_{i=1}^n \mathbb{P}^{\mathbb{X}}_{\sigma}(\big\{\omega \in \Omega^{\mathbb{X}} \ : \ \omega(\tau^{-1}(A_i)) = k_i\big\}) \\ &\quad = \prod_{i=1}^n \mathbb{P}^{\mathbb{X}}_{\sigma}\big(\big\{\omega \in \Omega^{\mathbb{X}} \ : \ \tau_*\omega(A_i) = k_i\big\}\big). \end{aligned}$$

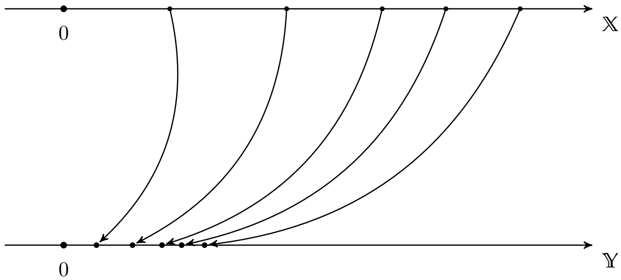
□

The next figure illustrates the transport of measure in the case of Gaussian intensities on $\mathbb{X} = \mathbb{R}$.



For example, in the case of a flat intensity $\rho(x) = \lambda$ on $\mathbb{X} = \mathbb{R}_+$ the intensity of the original Poisson point process becomes doubled under the mapping $\tau(x) = x/2$, since

$$\begin{aligned}\mathbb{P}_\sigma^{\mathbb{X}}(\tau_*\omega([0,t]) = n) &= \mathbb{P}_\sigma^{\mathbb{X}}(\omega(\tau^{-1}([0,t])) = n)\\ &= \mathrm{e}^{-\sigma(\tau^{-1}([0,t]))}\frac{(\sigma(\tau^{-1}([0,t])))^n}{n!}\\ &= \mathrm{e}^{-\sigma([0,2t])}\frac{(\sigma([0,2t]))^n}{n!}\\ &= \mathrm{e}^{-2\lambda t}(2\lambda t)^n/n!.\end{aligned}$$



11.4 Moments of Poisson Stochastic Integrals

As a consequence of (11.2.2) and the relation (1.7.11) between cumulants and moments we find that the n-*th* moment of $\int_{\mathbb{X}} f(x)\omega(dx)$ can be written as

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\left(\int_{\mathbb{X}} f(x)\omega(dx)\right)^n\right]\\ &= \sum_{k=0}^n \frac{1}{k!}\sum_{d_1+\cdots+d_k=n}\frac{n!}{d_1!\cdots d_k!}\int_{\mathbb{X}} f^{d_1}(x)\sigma(dx)\cdots\int_{\mathbb{X}} f^{d_k}(x)\sigma(dx)\\ &= \sum_{k=1}^n \sum_{B_1^n,B_2^n,\ldots,B_k^n}\int_{\mathbb{X}} f^{|B_1^n|}(x)\sigma(dx)\cdots\int_{\mathbb{X}} f^{|B_k^n|}(x)\sigma(dx),\end{aligned}$$

where the sum runs over the partitions $B_1^n, B_2^n, \ldots, B_k^n$ of $\{1, \ldots, n\}$ with cardinality $|B_i^n|$. Similarly, we find that the moments of the centered stochastic integral

$$\int_{\mathbb{X}} f(x)\omega(dx) - \int_{\mathbb{X}} f(x)\sigma(dx) = \int_{\mathbb{X}} f(x)(\omega(dx) - \sigma(dx))$$

satisfy

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\left(\int_{\mathbb{X}} f(x)\omega(dx) - \int_{\mathbb{X}} f(x)\sigma(dx)\right)^n\right]$$
$$= \sum_{k=1}^n \sum_{\substack{B_1^n, B_2^n, \ldots, B_k^n \\ |B_1^n|\geqslant 2, \ldots, |B_k^n|\geqslant 2}} \int_{\mathbb{X}} f^{|B_1^n|}(x)\sigma(dx) \cdots \int_{\mathbb{X}} f^{|B_k^n|}(x)\sigma(dx),$$

where the sum runs over the partitions $B_1^n, B_2^n, \ldots, B_k^n$ of $\{1, \ldots, n\}$ of size $|B_k^n| \geqslant 2$.

By taking $f = \mathbb{1}_A$ we can specialize the above results to the computation of the moments and central moments of the Poisson random variable $\omega(A)$ with parameter $\lambda := \sigma(A)$. For the moments we recover the well-known identity

$$\mathbb{E}\left[(\omega(A))^n\right] = \sum_{k=1}^n \lambda^n S(n,k),$$

where $S(n,k)$ is the Stirling number of the second kind, *i.e.* the number of partitions of $\{1, 2, \ldots, n\}$ made of k subsets, $k = 0, 1, \ldots, n$. Similarly, for the central moments we have

$$\mathbb{E}[(\omega(A) - \sigma(A))^n] = \sum_{k=1}^n \lambda^n S_2(n,k), \tag{11.4.1}$$

where $S_2(n,k)$ is the number of partitions of $\{1, 2, \ldots, n\}$ made of k subsets of size at least 2, $k = 0, 1, \ldots, n$, cf. Privault (2011).

Slivnyak-Mecke identity

The following version of the Slivnyak-Mecke identity Slivnyak (1962), Mecke (1967) allows us to compute the first moment of the first order stochastic integral of a random integrand.

Proposition 11.5. *For $u : \mathbb{X} \times \Omega^{\mathbb{X}} \longrightarrow \mathbb{R}$ a measurable process, we have*

$$\mathbb{E}_\sigma\left[\sum_{x\in\omega} u(x,\omega)\right] = \mathbb{E}_\sigma\left[\int_{\mathbb{X}} u(x, \omega \cup \{x\})\sigma(\mathrm{d}x)\right], \tag{11.4.2}$$

provided that

$$\mathbb{E}_\sigma\left[\int_{\mathbb{X}} |u(x, \omega \cup \{x\})|\sigma(\mathrm{d}x)\right] < \infty.$$

Proof. The proof is done when $\sigma(\mathbb{X}) < \infty$. We consider $u(x,\omega)$ written as

$$u(x,\omega) = \sum_{n\geqslant 0} \mathbb{1}_{\{\omega(\mathbb{X})=n\}} f_n(x; X_1, \ldots, X_n),$$

where for every $x \in \mathbb{X}$, $(x_1, \ldots, x_n) \longmapsto f_n(x; x_1, \ldots, x_n)$ is a symmetric integrable function of $\omega = \{X_1, \ldots, X_n\}$ when $\omega(\mathbb{X}) = n$, for each $n \geqslant 1$. We have

$$\begin{aligned}
&\mathbb{E}_\sigma\left[\sum_{x\in\omega} u(x,\omega)\right] \\
&= \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n\geqslant 1} \frac{1}{n!} \sum_{k=1}^{n} \int_{\mathbb{X}^n} f_n(x_k; x_1, \ldots, x_n) \sigma(\mathrm{d}x_1) \cdots \sigma(\mathrm{d}x_n) \\
&= \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n\geqslant 0} \frac{1}{(n+1)!} \sum_{k=1}^{n+1} \int_{\mathbb{X}^{n+1}} f_{n+1}(x; x_1, \ldots, x_{n+1}) \sigma(\mathrm{d}x) \sigma(\mathrm{d}x_1) \cdots \sigma(\mathrm{d}x_n) \\
&= \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n\geqslant 0} \frac{1}{n!} \int_{\mathbb{X}^n} \int_{\mathbb{X}} f_{n+1}(x; x, x_1, \ldots, x_n) \sigma(\mathrm{d}x) \sigma(\mathrm{d}x_1) \cdots \sigma(\mathrm{d}x_n) \\
&= \mathbb{E}_\sigma\left[\int_{\mathbb{X}} u(x, \omega \cup \{x\}) \sigma(\mathrm{d}x)\right].
\end{aligned}$$

□

Proposition 11.5 can actually be extended to a moment identity, as follows.

Proposition 11.6. *(Privault (2012)). Let $u : \mathbb{X} \times \Omega^{\mathbb{X}} \longrightarrow \mathbb{R}$ be a (measurable) process. We have*

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\left(\int_{\mathbb{X}} u(x,\omega)\omega(dx)\right)^n\right] \\
&= \sum_{k=1}^{n} \sum_{B_1^n, B_2^n, \ldots, B_k^n} \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\int_{\mathbb{X}^k} u^{|B_1^n|}(x_1, \omega \cup \{x_1, x_2, \ldots, x_k\}) \cdots \right. \\
&\qquad \left. \cdots u^{|B_k^n|}(x_k, \omega \cup \{x_1, x_2, \ldots, x_k\}) \sigma(dx_1) \cdots \sigma(dx_k)\right],
\end{aligned}$$

where the sum runs over the partitions $B_1^n, B_2^n, \ldots, B_k^n$ of $\{1, \ldots, n\}$, for any $n \geqslant 1$ such that all terms are integrable.

Proof. We will prove the following slightly more general formula

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[F\left(\int_{\mathbb{X}} u(x,\omega)\omega(dx)\right)^n\right] \\
&= \sum_{k=1}^{n} \sum_{B_1^n, B_2^n, \ldots, B_k^n} \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\int_{\mathbb{X}^k} F(\omega \cup \{x_1, x_2, \ldots, x_k\}) u^{|B_1^n|}(x_1, \omega \cup \{x_1, x_2, \ldots, x_k\}) \cdots \right. \\
&\qquad \left. \cdots u^{|B_k^n|}(x_k, \omega \cup \{x_1, x_2, \ldots, x_k\}) \sigma(dx_1) \cdots \sigma(dx_k)\right],
\end{aligned}$$

by induction on $n \geqslant 1$, for F a sufficiently integrable random variable. Clearly, the formula holds for $n = 0$ and $n = 1$. Assuming that it holds at the rank n

and using Proposition 11.5 we get

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[F\left(\int_{\mathbb{X}} u(x,\omega)\omega(dx)\right)^{n+1}\right] \\
&= \sum_{k=1}^n \sum_{B_1^n,B_2^n,\ldots,B_k^n} \\
&\quad \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\Bigg[\int_{\mathbb{X}^k} u^{|B_1^n|}(x_1,\omega\cup\{x_1,x_2\ldots,x_k\})\cdots u^{|B_k^n|}(x_k,\omega\cup\{x_1,x_2,\ldots,x_k\}) \\
&\quad \times F(\omega\cup\{x_1,x_2,\ldots,x_k\})\sum_{i=1}^k u(x_i,\omega\cup\{x_1,x_2,\ldots,x_k\})\sigma(dx_1)\cdots\sigma(dx_k)\Bigg] \\
&\quad + \sum_{k=1}^n \sum_{B_1^n,B_2^n,\ldots,B_k^n} \\
&\quad \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\Bigg[\int_{\mathbb{X}^k} u^{|B_1^n|}(x_1,\omega\cup\{x_1,x_2,\ldots,x_k\})\cdots u^{|B_k^n|}(x_k,\omega\cup\{x_1,x_2,\ldots,x_k\}) \\
&\quad \times F(\omega\cup\{x_1,x_2,\ldots,x_k\})\int_{\mathbb{X}} u(y,\omega\cup\{x_1,x_2,\ldots,x_k\})\omega(dy)\sigma(dx_1)\cdots\sigma(dx_k)\Bigg] \\
&= \sum_{k=1}^n \sum_{B_1^n,B_2^n\ldots,B_k^n} \\
&\quad \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\Bigg[\int_{\mathbb{X}^k} u^{|B_1^n|}(x_1,\omega\cup\{x_1,x_2,\ldots,x_k\})\cdots u^{|B_k^n|}(x_k,\omega\cup\{x_1,x_2,\ldots,x_k\}) \\
&\quad \times F(\omega\cup\{x_1,x_2,\ldots,x_k\})\sum_{i=1}^k u(x_i,\omega\cup\{x_1,x_2,\ldots,x_k\})\sigma(dx_1)\cdots\sigma(dx_k) \\
&+ \sum_{k=1}^n \sum_{B_1^n,B_2^n,\ldots,B_k^n} \\
&\quad \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\Bigg[\int_{\mathbb{X}^{k+1}} u^{|B_1^n|}(x_1,\omega\cup\{x_1,x_2,\ldots,x_k,y\})\cdots u^{|B_k^n|}(x_k,\omega\cup\{x_1,x_2,\ldots,x_k,y\}) \\
&\quad \times F(\omega\cup\{x_1,x_2,\ldots,x_k,y\})u(y,\omega\cup\{x_1,x_2,\ldots,x_k,y\})\sigma(dy)\sigma(dx_1)\cdots\sigma(dx_k)] \\
&= \sum_{k=1}^{n+1} \sum_{B_1^{n+1},B_2^{n+1},\ldots,B_k^{n+1}} \\
&\quad \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\Bigg[\int_{\mathbb{X}^k} F(\omega\cup\{x_1,x_2,\ldots,x_k\})u^{|B_1^{n+1}|}(x_1,\omega\cup\{x_1,x_2,\ldots,x_k\})\times\cdots \\
&\quad \cdots\times u^{|B_k^{n+1}|}(x_k,\omega\cup\{x_1,x_2,\ldots,x_k\})\sigma(dx_1)\cdots\sigma(dx_k)\Bigg],
\end{aligned}
$$

where the sum runs over the partitions $B_1^{n+1}, B_2^{n+1}, \ldots, B_k^{n+1}$ of $\{1, \ldots, n+1\}$. □

11.5 Deviation Inequalities

The next proposition is a particular case of results proved for general Poisson functionals in Houdré and Privault (2002), Wu (2000), cf. also Corollary 6.9.3 of Privault (2009). Such results are useful for the accurate estimation of loss (or outage) probabilities in a wireless network whose users are randomly located within the space $\mathbb{X}$, cf. *e.g.* Decreusefond et al. (2009).

Proposition 11.7. *Let $f : \mathbb{X} \longrightarrow \mathbb{R}$ be a (measurable) function such that*

i) $f(y) \leqslant K$, *$\sigma(dy)$-a.e., for some $K > 0$, and*

ii) $\int_{\mathbb{X}} |f(y)|^2 \sigma(dy) \leqslant \alpha^2$, *for some $\alpha > 0$.*

Then we have

$$\mathbb{P}_\sigma^{\mathbb{X}} \left(\int_{\mathbb{X}} f(y)\omega(dy) - \int_{\mathbb{X}} f(y)\sigma(dy) \geqslant x \right) \leqslant \mathrm{e}^{x/K} \left(1 + \frac{xK}{\alpha^2} \right)^{-\frac{x}{K} - \frac{\alpha^2}{K^2}}, \tag{11.5.1}$$

for all $x > 0$.

Proof. We let

$$\int_{\mathbb{X}} f(y)\tilde{\omega}(dy) := \int_{\mathbb{X}} f(y)(\omega(dy) - \sigma(dy))$$

denote the compensated Poisson stochastic integral of f, and let

$$M(s) := \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) \right], \qquad s \geqslant 0,$$

denote the moment generating function of $\int_{\mathbb{X}} f(y)\tilde{\omega}(dy)$. By Proposition 11.5 we have

$$\begin{aligned}
M'(s) &= \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\int_{\mathbb{X}} f(x)\tilde{\omega}(dx) \exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) \right] \\
&= \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\int_{\mathbb{X}} \exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) f(x)\tilde{\omega}(dx) \right. \\
&\qquad \left. - \int_{\mathbb{X}} \exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) f(x)\sigma(dx) \right] \\
&= \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\int_{\mathbb{X}} \exp\left(sf(x) + s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) f(x)\sigma(dx) \right. \\
&\qquad \left. - \int_{\mathbb{X}} \exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) f(x)\sigma(dx) \right] \\
&= \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\int_{\mathbb{X}} f(x)(\mathrm{e}^{sf(x)} - 1)\sigma(dx) \exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) \right] \\
&= \int_{\mathbb{X}} f(x)(\mathrm{e}^{sf(x)} - 1)\sigma(dx) \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) \right] \\
&= s \int_{\mathbb{X}} |f(x)|^2 \frac{\mathrm{e}^{sf(x)} - 1}{sf(x)} \sigma(dx) \mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}} \left[\exp\left(s \int_{\mathbb{X}} f(y)\tilde{\omega}(dy) \right) \right]
\end{aligned}$$

$$\begin{aligned}&\leqslant \frac{\mathrm{e}^{sK}-1}{K}\int_{\mathbb{X}}|f(x)|^2\sigma(dx)\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(s\int_{\mathbb{X}}f(y)\tilde{\omega}(dy)\right)\right]\\&=\alpha^2\frac{\mathrm{e}^{sK}-1}{K}\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(s\int_{\mathbb{X}}f(y)\tilde{\omega}(dy)\right)\right]=\alpha^2\frac{\mathrm{e}^{sK}-1}{K}M(s),\end{aligned}$$

which shows that

$$\frac{M'(s)}{M(s)}\leqslant h(s):=\alpha^2\frac{\mathrm{e}^{sK}-1}{K},\qquad s\geqslant 0,$$

hence

$$M(t)\leqslant\exp\left(\int_0^t h(s)ds\right)=\exp\left(\alpha^2\int_0^t\frac{\mathrm{e}^{sK}-1}{K}ds\right),\qquad t\geqslant 0.$$

Consequently we have

$$\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}[F\mathrm{e}^{tF}]\leqslant h(t)\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}[\mathrm{e}^{tF}],\qquad t\geqslant 0.$$

Next, the Markov inequality shows that

$$\begin{aligned}&\mathbb{P}_\sigma^{\mathbb{X}}\left(\int_{\mathbb{X}}f(y)(\omega(dy)-\sigma(dy))\geqslant x\right)=\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\mathbb{1}_{\{\int_{\mathbb{X}}f(y)(\omega(dy)-\sigma(dy))\geqslant x\}}\right]\\&\quad\leqslant\mathrm{e}^{-tx}\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\mathbb{1}_{\{\int_{\mathbb{X}}f(y)(\omega(dy)-\sigma(dy))\geqslant x\}}\exp\left(t\int_{\mathbb{X}}f(y)\omega(dy)\right)\right]\\&\quad\leqslant\mathrm{e}^{-tx}\mathbb{E}_{\mathbb{P}_\sigma^{\mathbb{X}}}\left[\exp\left(t\int_{\mathbb{X}}f(y)\omega(dy)\right)\right]\\&\quad\leqslant\exp\left(-tx+\int_0^t h(s)ds\right)\\&\quad\leqslant\exp\left(-tx+\alpha^2\int_0^t\frac{\mathrm{e}^{sK}-1}{K}ds\right)\\&\quad=\exp\left(-tx+\frac{\alpha^2}{K^2}(\mathrm{e}^{tK}-tK-1)\right).\end{aligned}$$

Minimizing the above term in t with $t=K^{-1}\log(1+Kx/\alpha^2)$ shows that

$$\mathbb{P}_\sigma^{\mathbb{X}}\left(\int_{\mathbb{X}}f(y)(\omega(dy)-\sigma(dy))\geqslant x\right)\leqslant\exp\left(\frac{x}{K}-\left(\frac{x}{K}+\frac{\alpha^2}{K^2}\right)\log\left(1+\frac{xK}{\alpha^2}\right)\right),$$

which yields (11.5.1). □

The bound (11.5.1) also shows that

$$\begin{aligned}&\mathbb{P}_\sigma^{\mathbb{X}}\left(\int_{\mathbb{X}}f(y)\omega(dy)-\int_{\mathbb{X}}f(y)\sigma(dy)\geqslant x\right)\\&\qquad\leqslant\exp\left(-\frac{x}{2K}\log\left(1+\frac{xK}{\alpha^2}\right)\right)=\left(1+\frac{xK}{\alpha^2}\right)^{-x/2K},\end{aligned}$$

for all $x > 0$.

In case the function $f : \mathbb{X} \longrightarrow \mathbb{R}$ is such that

i) $f(y) \leqslant 0$, $\sigma(dy)$-a.e., for some $K \in \mathbb{R}$, and

ii) $\int_{\mathbb{X}} |f(y)|^2 \sigma(dy) \leqslant \alpha^2$,

the above argument can be applied with $K = 0$, and also shows that

$$\mathbb{P}_\sigma^{\mathbb{X}} \left(\int_{\mathbb{X}} f(y)\omega(dy) - \int_{\mathbb{X}} f(y)\sigma(dy) \geqslant x \right) \leqslant \exp\left(-\frac{x^2}{2\alpha^2} \right),$$

for all $x > 0$.

Exercises

Exercise 11.1 Consider the standard Poisson process $(N_t)_{t\in\mathbb{R}_+}$ on $\mathbb{R}_+$ with intensity $\lambda = 2$ and jump times $(T_k)_{k\geqslant 1}$. Compute

$$\mathbb{E}[T_1 + T_2 + T_3 \mid N_2 = 2].$$

Exercise 11.2 Consider the spatial Poisson point process on $\mathbb{R}^2$ with intensity $\lambda = 0.5$ per square meter. What is the probability that there are 10 events within a circle of radius 3 meters.

Exercise 11.3 Some living organisms are distributed in space according to a Poisson point process of intensity $\theta = 0.6$ units per mm^3. Compute the probability that more than two living organisms are found within a 10 mm^3 volume.

Exercise 11.4 Defects are present over a piece of fabric according to a Poisson point process with intensity of one defect per piece of fabric. Both halves of the piece is checked separately. What is the probability that both inspections record at least one defect?

Exercise 11.5 Let $\lambda > 0$ and suppose that N points are independently and uniformly distributed over the interval $[0, N]$. Determine the probability distribution for the number of points in the interval $[0, \lambda]$ as $N \to \infty$.

Exercise 11.6 Suppose that $X(A)$ is a spatial Poisson point process of discrete items scattered on the plane $\mathbb{R}^2$ with intensity $\lambda = 0.5$ per square meter. We let

$$D((x,y),r) = \{(u,v) \in \mathbb{R}^2 \ : \ (x-u)^2 + (y-v)^2 \leqslant r^2\}$$

denote the disc with radius r centered at (x,y) in $\mathbb{R}^2$. No evaluation of numerical expressions is required in this exercise.

a) What is the probability that 10 items are found within the disk $D((0,0),3)$ with radius 3 meters centered at the origin?
b) What is the probability that 5 items are found within the disk $D((0,0),3)$ *and* 3 items are found within the disk $D((x,y),3)$ with $(x,y)=(7,0)$?
c) What is the probability that 8 items are found anywhere within

   $D((0,0),3) \bigcup D((x,y),3)$ with $(x,y)=(7,0)$?
d) Given that 5 items are found within the disk $D((0,0),1)$, what is the probability that 3 of them are located within the disk $D((1/2,0),1/2)$ centered at $(1/2,0)$ with radius $1/2$?

Exercise 11.7 Let S_n be a Poisson random variable with parameter λn for all $n \geqslant 1$, with $\lambda > 0$. Show that the moments of order p of $(S_n - \lambda n)/\sqrt{n}$ satisfy the bound

$$\operatorname*{Sup}_{n \geqslant 1} \mathbb{E}\left[\left| \frac{S_n - \lambda n}{\sqrt{n}} \right|^p \right] < C_p$$

where $C_p > 0$ is a finite constant for all $p \geqslant 1$. *Hint*: Use Relation (11.4.1).

Exercise 11.8 Let $(N_t)_{t \in \mathbb{R}_+}$ denote a standard Poisson process on $\mathbb{R}_+$. Given a bounded function $f \in L^1(\mathbb{R}_+)$ we let

$$\int_0^\infty f(y)(dN_y - dy)$$

denote the compensated Poisson stochastic integral of f, and let

$$M(s) := \mathbb{E}\left[\exp\left(s \int_0^\infty f(y)(dN_y - dy) \right) \right] = \exp\left(\int_0^\infty (\mathrm{e}^{sf(y)} - sf(y) - 1) dy \right),$$

$s \geqslant 0$, denote the moment generating function of $\int_0^\infty f(y)(dN_y - dy)$.

a) Show that we have

$$\frac{M'(s)}{M(s)} \leqslant h(s) := \alpha^2 \frac{e^{sK} - 1}{K}, \qquad s \geqslant 0,$$

   provided that $f(t) \leqslant K$, dt-a.e., for some $K > 0$ and provided in addition that $\int_0^\infty |f(y)|^2 dy \leqslant \alpha^2$, for some $\alpha > 0$.
b) Show that

$$M(t) \leqslant \exp\left(\int_0^t h(s) ds \right) = \exp\left(\alpha^2 \int_0^t \frac{e^{sK} - 1}{K} ds \right), \qquad t \geqslant 0.$$

c) Show, using Markov's inequality, that

$$\mathbb{P}\left(\int_0^\infty f(y)(dN_y - dy) \geqslant x\right) \leqslant \mathrm{e}^{-tx}\mathbb{E}\left[\exp\left(t\int_0^\infty f(y)dN_y\right)\right],$$

and that

$$\mathbb{P}\left(\int_0^\infty f(y)(dN_y - dy) \geqslant x\right) \leqslant \exp\left(-tx + \alpha^2\int_0^t \frac{e^{sK}-1}{K}ds\right).$$

d) By minimization in t, show that

$$\mathbb{P}\left(\int_0^\infty f(y)dN_y - \int_0^\infty f(y)dy \geqslant x\right) \leqslant \mathrm{e}^{x/K}\left(1+\frac{xK}{\alpha^2}\right)^{-x/K-\alpha^2/K^2},$$

for all $x > 0$, and that

$$\mathbb{P}\left(\int_0^\infty f(y)dN_y - \int_0^\infty f(y)dy \geqslant x\right) \leqslant \left(1+\frac{xK}{\alpha^2}\right)^{-x/2K},$$

for all $x > 0$.

Exercise 11.9 Let Φ be a Poisson point process with finite intensity measure $\sigma(dx)$ on $\mathbb{X} := \mathbb{R}^d \times \mathbb{R}_+$, given by $\sigma(dx) = \lambda dy \rho(r)dr$, $x = (y, r) \in \mathbb{R}^d \times \mathbb{R}_+$, where $\lambda > 0$ and $\rho(r)$ is a probability density function on $\mathbb{R}_+$. We also consider the Probability Generating Functional (PGFl) $\mathcal{G}_\Phi(f) := \mathbb{E}\left[\prod_{x\in\Phi} f(x)\right]$.

(a) Show that $\mathcal{G}_\Phi(f) = \exp\left(\int_{\mathbb{X}}(f(x)-1)\sigma(dx)\right)$, where $f - 1 \in L^1(\mathbb{X}, \sigma)$.
(b) Using the PGFl $\mathcal{G}_\Phi$, recover the probability $\mathbb{P}(\Phi \cap A = \emptyset)$ that *no* process points can be found within a given subset A of $\mathbb{X}$.
(c) Consider the Boolean model Ξ on $\mathbb{R}^d$ made of the union of balls constructed by associating every point (y, r) of Φ to a ball of radius r centered at $y \in \mathbb{R}^d$.

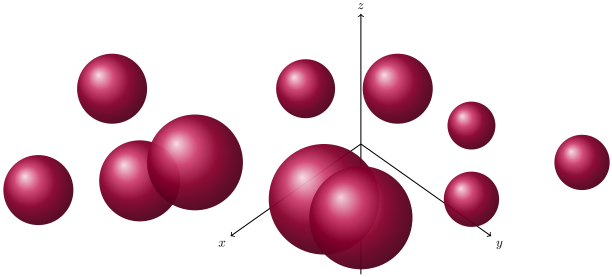


Fig. 11.5: Boolean model in dimension three.

Using the volume $v_d = \pi^{d/2}/\Gamma(1+d/2)$ of the unit ball in $\mathbb{R}^d$, find the probability that the union of balls contains the point 0.

(d) Find the probability that this Boolean model covers the whole space $\mathbb{R}^d$.

Exercise 11.10 Consider a Poisson point process ω with intensity $\sigma(dy)e^{-r}dr$ on $[0,1]^d \times [0,\infty)$, given by $\omega := \{(Y_k, R_k)\}_k$. Each point $(x,r) \in \omega$ models a location $x \in [0,1]^d$ along with a radius $r \in [0,\infty)$ corresponding to the radius of the ball centered around it. The spherical Boolean model Ξ with ω as its driving Poisson point process can now be constructed as

$$\Xi = \bigcup_{(x,r)\in\omega} B(x,r),$$

which consists in the random subset of points in $[0,1]^d$ which are covered by at least one Euclidean ball centered around the points of the Poisson point process ω.

a) Give the probability of not observing any ball of radius smaller than $1/2$.
b) Give the mean number of balls which have radius less than $1/2$.

Chapter 12
Reliability Theory

This chapter consists in a short review of survival probabilities based on failure rate and reliability functions, in connection with Poisson processes having a time-dependent intensity.

12.1 Survival Probabilities

Let $\tau : \Omega \longrightarrow \mathbb{R}_+$ denote (random) the lifetime of an entity, and let $\mathbb{P}(\tau \geqslant t)$ denote its probability of surviving at least t years, $t > 0$. The probability of surviving up to a (deterministic) time T, given that the entity has already survived up to time t, is given by

$$\begin{aligned}\mathbb{P}(\tau > T \mid \tau > t) &= \frac{\mathbb{P}(\tau > T \text{ and } \tau > t)}{\mathbb{P}(\tau > t)} \\ &= \frac{\mathbb{P}(\tau > T)}{\mathbb{P}(\tau > t)}, \qquad 0 \leqslant t \leqslant T.\end{aligned}$$

Let now

$$\lambda(t) := \lim_{h \searrow 0} \frac{\mathbb{P}(\tau < t + h \mid \tau > t)}{h}, \qquad t \geqslant 0,$$

denote the *failure rate function* associated to τ. Letting $A = \{\tau < t + h\}$ and $B = \{\tau > t\}$ we note that $(\Omega \setminus A) \subset B$, hence $A \cap B = B \setminus A^c$, and

$$\begin{aligned}\lambda(t) &= \lim_{h \searrow 0} \frac{\mathbb{P}(\tau < t+h \mid \tau > t)}{h} \\ &= \frac{1}{\mathbb{P}(\tau > t)} \lim_{h \searrow 0} \frac{\mathbb{P}(\tau < t+h \text{ and } \tau > t)}{h} \\ &= \frac{1}{\mathbb{P}(\tau > t)} \lim_{h \searrow 0} \frac{\mathbb{P}(\tau > t) - \mathbb{P}(\tau > t+h)}{h} \\ &= -\frac{d}{dt} \log \mathbb{P}(\tau > t) \\ &= -\frac{1}{\mathbb{P}(\tau > t)} \frac{d}{dt} \mathbb{P}(\tau > t) \qquad (12.1.1) \\ &= -\frac{1}{R(t)} \frac{d}{dt} R(t),\end{aligned}$$

where the *reliability function* $R(t)$ is defined by

$$R(t) := \mathbb{P}(\tau > t), \qquad t \geqslant 0.$$

This yields

$$R'(t) = -\lambda(t) R(t),$$

with $R(0) = \mathbb{P}(\tau > 0) = 1$, which has for solution

$$R(t) = \mathbb{P}(\tau > t) = R(0) \exp\left(-\int_0^t \lambda(u) du\right) = \exp\left(-\int_0^t \lambda(u) du\right), \tag{12.1.2}$$

$t \geqslant 0$. Hence we have

$$\mathbb{P}(\tau > T \mid \tau > t) = \frac{R(T)}{R(t)} = \exp\left(-\int_t^T \lambda(u) du\right), \qquad t \in [0, T]. \tag{12.1.3}$$

In case the failure rate function $\lambda(t) = c$ is constant we recover the memoryless property of the exponential distribution with parameter $c > 0$, cf. (9.2.3).

Relation (12.1.2) can be recovered informally as

$$\mathbb{P}(\tau > T) = \prod_{0<t<T} \mathbb{P}(\tau > t + dt \mid \tau > t) = \prod_{0<t<T} \exp\left(-\lambda(t) dt\right),$$

which yields

$$\mathbb{P}(\tau > t) = \exp\left(-\int_0^t \lambda(s) ds\right), \qquad t \geqslant 0,$$

in the limit.

12.2 Poisson Process with Time-Dependent Intensity

Recall that the random variable τ has the exponential distribution with parameter $\lambda > 0$ if

$$\mathbb{P}(\tau > t) = \mathrm{e}^{-\lambda t}, \qquad t \geqslant 0,$$

cf. (1.5.3). Given $(\tau_n)_{n\geqslant 0}$ a sequence of *i.i.d.* exponentially distributed random variables, letting

$$T_n = \tau_0 + \cdots + \tau_{n-1}, \qquad n \geqslant 1,$$

and

$$N_t = \sum_{n\geqslant 1} \mathbb{1}_{[T_n,\infty)}(t), \qquad t \geqslant 0,$$

defines the standard Poisson process with intensity $\lambda > 0$ of Section 9.1 and we have

$$\mathbb{P}(N_t - N_s = k) = \mathrm{e}^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}, \qquad k \geqslant 0.$$

The intensity of the Poisson process can in fact made time-dependent. For example under the time change

$$X_t = N_{\int_0^t \lambda(s)ds}$$

where $(\lambda(u))_{u\in\mathbb{R}_+}$ is a deterministic function of time, we have

$$\mathbb{P}(X_t - X_s = k) = \frac{\left(\int_s^t \lambda(u)du\right)^k}{k!} \exp\left(-\int_s^t \lambda(u)du\right), \qquad k \geqslant 0.$$

In this case we have

$$\mathbb{P}(X_{t+h} - X_t = 0) = \mathrm{e}^{-\lambda(t)h} + o(h) = 1 - \lambda(t)h + o(h), \qquad h \searrow 0, \tag{12.2.1}$$

and

$$\mathbb{P}(X_{t+h} - X_t = 1) = 1 - \mathrm{e}^{-\lambda(t)h} + o(h) \simeq \lambda(t)h, \qquad h \searrow 0, \tag{12.2.2}$$

which can also viewed as a pure birth process with time-dependent intensity. Letting τ_0 denote the first jump time of $(X_t)_{t\in\mathbb{R}_+}$, we have

$$R(t) = \mathbb{P}(\tau_0 > t) = \mathbb{P}(X_t = 0) = \exp\left(-\int_0^t \lambda(u)du\right), \qquad t \geqslant 0, \tag{12.2.3}$$

hence by (12.1.3) we find

$$\mathbb{P}(X_{t+h} = 0 \mid X_t = 0) = \mathbb{P}(\tau_0 > t + h \mid \tau_0 > t)$$

$$= \mathrm{e}^{-\lambda(t)h} + o(h) = 1 - \lambda(t)h + o(h), \qquad h \searrow 0,$$

and

$$\mathbb{P}(X_{t+h} \geqslant 1 \mid X_t = 0) = \mathbb{P}(\tau_0 < t+h \mid \tau_0 > t) = 1 - \mathbb{P}(\tau_0 > t+h \mid \tau_0 > t)$$
$$= 1 - \mathrm{e}^{-\lambda h} \simeq \lambda(t)h + o(h), \qquad h \searrow 0,$$

which coincide respectively with $\mathbb{P}(X_{t+h} - X_t = 0)$ and $\mathbb{P}(X_{t+h} - X_t = 1)$ in (12.2.1) and (12.2.2) above, as $(X_t)_{t\in\mathbb{R}_+}$ has independent increments.

Cox processes

The intensity process $\lambda(s)$ can also be made random. In this case, $(X_t)_{t\in\mathbb{R}_+}$ is called a *Cox process* and it may not have independent increments. For example, assume that $(\lambda_u)_{u\in\mathbb{R}_+}$ is a two-state Markov chain on $\{0, \lambda\}$, with transitions

$$\mathbb{P}(\lambda_{t+h} = \lambda \mid \lambda_t = 0) = \alpha h, \qquad h \searrow 0,$$

and

$$\mathbb{P}(\lambda_{t+h} = 0 \mid \lambda_t = \lambda) = \beta h, \qquad h \searrow 0.$$

In this case the probability distribution of N_t can be explicitly computed, cf. Chapter VI-7 in Karlin and Taylor (1998).

Renewal processes

A *renewal process* is a counting process $(N_t)_{t\in\mathbb{R}_+}$ given by

$$N_t = \sum_{k\geqslant 1} k \mathbb{1}_{[T_k, T_{k+1})}(t) = \sum_{k\geqslant 1} \mathbb{1}_{[T_k,\infty)}(t), \qquad t \geqslant 0,$$

in which $\tau_k = T_{k+1} - T_k$, $k \in \mathbb{N}$, is a sequence of independent identically distributed random variables. In particular, Poisson processes are renewal processes.

12.3 Mean Time to Failure

The mean time to failure is given, from (12.1.1), by

$$\mathbb{E}[\tau] = \int_0^\infty t \frac{d}{dt}\mathbb{P}(\tau < t)dt = -\int_0^\infty t \frac{d}{dt}\mathbb{P}(\tau > t)dt$$
$$= -\int_0^\infty tR'(t)dt = \int_0^\infty R(t)dt, \tag{12.3.1}$$

provided that $\lim_{t\searrow\infty} tR(t) = 0$. For example when τ has the distribution function (12.2.3) we get

$$\mathbb{E}[\tau] = \int_0^\infty R(t)dt = \int_0^\infty \exp\left(-\int_0^t \lambda(u)du\right) dt.$$

In case the function $\lambda(t) = \lambda > 0$ is constant we recover the mean value

$$\mathbb{E}[\tau] = \int_0^\infty \mathrm{e}^{-\lambda t} dt = \frac{1}{\lambda}$$

of the exponential distribution with parameter $\lambda > 0$.

Exercise

Exercise 12.1 Assume that the random time τ has the *Weibull* distribution with probability density

$$f_\beta(x) = \beta \mathbb{1}_{[0,\infty)} x^{\beta-1} \mathrm{e}^{-x^\beta}, \qquad x \in \mathbb{R},$$

where $\beta > 0$ is a called the shape parameter.

a) Compute the distribution function F_β of the Weibull distribution.
b) Compute the reliability function $R(t) = \mathbb{P}(\tau > t)$.
c) Compute the failure rate function $\lambda(t)$.
d) Compute the mean time to failure.

Some Useful Identities

Here we present a summary of algebraic identities that are used in this text.

Indicator functions

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ \\ 0 & \text{if } x \notin A. \end{cases} \qquad \mathbb{1}_{[a,b]}(x) = \begin{cases} 1 & \text{if } a \leqslant x \leqslant b, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Binomial coefficients

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}, \qquad k = 0, 1, \ldots, n.$$

Exponential series

$$e^x = \sum_{n \geqslant 0} \frac{x^n}{n!}, \qquad x \in \mathbb{R}. \tag{A.1}$$

Geometric sum

$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}, \qquad r \neq 1. \tag{A.2}$$

Geometric series

$$\sum_{k\geqslant 0} r^k = \frac{1}{1-r}, \qquad -1 < r < 1. \tag{A.3}$$

Differentiation of geometric series

$$\sum_{k\geqslant 1} kr^{k-1} = \frac{\partial}{\partial r}\sum_{k\geqslant 0} r^k = \frac{\partial}{\partial r}\frac{1}{1-r} = \frac{1}{(1-r)^2}, \qquad -1 < r < 1. \tag{A.4}$$

Binomial identities

$$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n.$$

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n. \tag{A.5}$$

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}.$$

$$\begin{aligned}\sum_{k=0}^{n} k\binom{n}{k} a^k b^{n-k} &= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} a^k b^{n-k} \\ &= \sum_{k=0}^{n-1} \frac{n!}{(n-1-k)!k!} a^{k+1} b^{n-1-k} \\ &= n\sum_{k=0}^{n-1} \binom{n-1}{k} a^{k+1} b^{n-1-k} \\ &= na(a+b)^{n-1}, \qquad n \geqslant 1,\end{aligned}$$

$$\begin{aligned}\sum_{k=0}^{n} k\binom{n}{k} a^k b^{n-k} &= a\frac{\partial}{\partial a}\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \\ &= a\frac{\partial}{\partial a}(a+b)^n \\ &= na(a+b)^{n-1}, \qquad n \geqslant 1.\end{aligned}$$

Sums of integers

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}. \tag{A.6}$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}. \tag{A.7}$$

Taylor expansion

$$(1+x)^\alpha = \sum_{k\geqslant 0} \frac{x^k}{k!}\alpha(\alpha-1)\times\cdots\times(\alpha-(k-1)). \tag{A.8}$$

Differential equation

$$\text{The solution of } f'(t) = cf(t) \text{ is given by } f(t) = f(0)\mathrm{e}^{ct}, \quad t \geqslant 0. \tag{A.9}$$

Solutions to the Exercises

Chapter 1 - Probability Background

Exercise 1.1 Assuming that (1.2.2) holds, by the complement rule and using the relation

$$\sum_{\emptyset \neq I \subset J} (-1)^{|I|+1} = 1 - \sum_{I \subset J} (-1)^{|I|} = 1$$

we have

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j \in J} A_j\right) &= \mathbb{P}\left(\left(\bigcup_{j \in J} A_j^c\right)^c\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{j \in J} A_j^c\right) \\
&= 1 - \sum_{I \subset J} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i^c\right) \\
&= 1 - \sum_{\emptyset \neq I \subset J} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i^c\right) \\
&= \sum_{I \subset J} (-1)^{|I|+1} \left(1 - \mathbb{P}\left(\bigcap_{i \in I} A_i^c\right)\right) \\
&= \sum_{I \subset J} (-1)^{|I|+1} \mathbb{P}\left(\bigcup_{i \in I} A_i\right),
\end{aligned}$$

which proves (1.2.3).

Exercise 1.2

a) We have

$$\mathbb{P}(X < \infty) = \sum_{k\geqslant 0} \mathbb{P}(X = k) = q \sum_{k\geqslant 0} p^k = \frac{q}{1-p},$$

hence

$$\mathbb{P}(X = \infty) = 1 - \mathbb{P}(X < \infty) = 1 - \frac{q}{1-p} = \frac{1-p-q}{1-p}.$$

When $0 \leqslant q < 1-p$ we have $\mathbb{P}(X = \infty) > 0$, hence $\mathbb{E}[X] = \infty$.

b) We have $\mathbb{P}(Y < \infty) = \mathbb{P}(X < \infty) = 1$ and

$$\mathbb{E}[Y] = \sum_{k\geqslant 0} r^k \mathbb{P}(Y = r^k) = \sum_{k\geqslant 0} r^k \mathbb{P}(X = k) = (1-p) \sum_{k\geqslant 0} (pr)^k,$$

hence $\mathbb{E}[Y] = (1-p)/(1-pr) < \infty$ when $r < 1/p$, and $\mathbb{E}[Y] = +\infty$ when $r \geqslant 1/p$.

Exercise 1.3 We write

$$Z = \sum_{k=1}^{N} X_k$$

where $\mathbb{P}(N = n) = 1/6$, $n = 1, 2, \ldots, 6$, and X_k is a Bernoulli random variable with parameter $1/2$, $k = 1, 2, \ldots, 6$.

a) We have

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[\mathbb{E}[Z \mid N]] = \sum_{n=1}^{6} \mathbb{E}[Z \mid N = n]\mathbb{P}(N = n) \\
&= \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k=1}^{N} X_k \;\middle|\; N = n\right] \mathbb{P}(N = n) = \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k=1}^{n} X_k\right] \mathbb{P}(N = n) \\
&= \sum_{n=1}^{6} \sum_{k=1}^{n} \mathbb{E}[X_k]\mathbb{P}(N = n) = \frac{1}{6} \sum_{n=1}^{6} \sum_{k=1}^{n} \frac{1}{2} = \frac{1}{2 \times 6} \sum_{n=1}^{6} n \\
&= \frac{1}{2 \times 6} \times \frac{6 \times (6+1)}{2} = \frac{7}{4}, \qquad \text{(S.1)}
\end{aligned}$$

where we applied (A.6). Concerning the variance, we have

$$\begin{aligned}
\mathbb{E}[Z^2] &= \mathbb{E}[\mathbb{E}[Z^2 \mid N]] \\
&= \sum_{n=1}^{6} \mathbb{E}[Z^2 \mid N = n]\mathbb{P}(N = n)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=1}^{6} \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2 \;\middle|\; N=n\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k=1}^{n} X_k \sum_{l=1}^{n} X_l \;\middle|\; N=n\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k,l=1}^{n} X_k X_l \;\middle|\; N=n\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k=1}^{n} X_k^2 + \sum_{1\leqslant k\neq l\leqslant n} X_k X_l\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \mathbb{E}\left[\sum_{k=1}^{n} X_k^2\right] \mathbb{P}(N=n) + \sum_{n=1}^{6} \mathbb{E}\left[\sum_{1\leqslant k\neq l\leqslant n} X_k X_l\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2\right] \mathbb{P}(N=n) + \sum_{n=1}^{6} \sum_{1\leqslant k\neq l\leqslant n} \mathbb{E}\left[X_k X_l\right] \mathbb{P}(N=n) \\
&= \sum_{n=1}^{6} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2\right] \mathbb{P}(N=n) + \sum_{n=1}^{6} \sum_{1\leqslant k\neq l\leqslant n} \mathbb{E}[X_k]\mathbb{E}[X_l]\mathbb{P}(N=n) \\
&= \frac{1}{2\times 6} \sum_{n=1}^{6} n + \frac{1}{6\times 2^2} \sum_{n=1}^{6} n(n-1) \\
&= \frac{1}{2\times 6} \sum_{n=1}^{6} n + \frac{1}{6\times 2^2} \sum_{n=1}^{6} n^2 - \frac{1}{6\times 2^2} \sum_{n=1}^{6} n \\
&= \frac{1}{6\times 2^2} \sum_{n=1}^{6} n + \frac{1}{6\times 2^2} \sum_{n=1}^{6} n^2 \\
&= \frac{1}{6\times 2^2} \frac{6(6+1)}{2} + \frac{1}{6\times 2^2} \frac{6(6+1)(2\times 6+1)}{6} = \frac{14}{3},
\end{aligned} \tag{S.2}
$$

where we used (A.7) and (S.1), hence

$$
\mathrm{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{14}{3} - \frac{49}{16} = \frac{77}{48}. \tag{S.3}
$$

Using (1.6.10) and $\mathrm{Var}[X_k] = p(1-p)$ we could also write

$$
\mathrm{Var}[Z \mid N=n] = \mathrm{Var}\left[\sum_{k=1}^{N} X_k \;\middle|\; N=n\right]
$$

$$= \mathrm{Var}\left[\sum_{k=1}^{n} X_k\right]$$
$$= \sum_{k=1}^{n} \mathrm{Var}[X_k]$$
$$= np(1-p),$$

which implies

$$\mathbb{E}[Z^2 \mid N = n] = \mathrm{Var}[Z \mid N = n] + (\mathbb{E}[Z \mid N = n])^2 = np(1-p) + n^2p^2,$$

hence

$$\begin{aligned}
\mathbb{E}\left[Z^2\right] &= \sum_{k=1}^{6} \mathbb{E}[Z^2 \mid N = n]\mathbb{P}(N = n) \\
&= \frac{1}{6}\sum_{k=1}^{6} \mathbb{E}[Z^2 \mid N = n] \\
&= \frac{1}{6}\sum_{k=1}^{6} (np - np^2 + n^2p^2),
\end{aligned}$$

and recovers (S.2).

b) Applying (1.3.1) with $B = \{Z = k\}$, we have

$$\begin{aligned}
\mathbb{P}(Z = l) &= \mathbb{P}\left(\sum_{k=1}^{N} X_k = l\right) \\
&= \sum_{n=1}^{6} \mathbb{P}\left(\sum_{k=1}^{N} X_k = l \;\middle|\; N = n\right)\mathbb{P}(N = n) \\
&= \sum_{n=1}^{6} \mathbb{P}\left(\sum_{k=1}^{n} X_k = l\right)\mathbb{P}(N = n).
\end{aligned}$$

Next, we notice that the deterministic summation $\displaystyle\sum_{k=1}^{n} X_k$ has a binomial distribution with parameter $(n, 1/2)$, *i.e.*

$$\mathbb{P}\left(\sum_{k=1}^{n} X_k = l\right) = \begin{cases} \displaystyle\binom{n}{l}\left(\frac{1}{2}\right)^l\left(\frac{1}{2}\right)^{n-l} & \text{if } l = 0, 1, \ldots, n, \\ 0 & \text{if } l > n, \end{cases}$$

which yields

$$\begin{aligned}\mathbb{P}(Z=0) &= \sum_{n=1}^{6} \mathbb{P}\left(\sum_{k=1}^{n} X_k = 0\right)\mathbb{P}(N=n)\\ &= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^n\\ &= \frac{1}{6\times 2}\sum_{n=0}^{5}\left(\frac{1}{2}\right)^n\\ &= \frac{1-(1/2)^6}{6},\end{aligned}$$

and, for $l = 1, 2, \ldots, 6$,

$$\begin{aligned}\mathbb{P}(Z=l) &= \sum_{n=1}^{6} \mathbb{P}\left(\sum_{k=1}^{n} X_k = l\right)\mathbb{P}(N=n)\\ &= \frac{1}{6}\sum_{n=l}^{6}\left(\frac{1}{2}\right)^l\left(\frac{1}{2}\right)^{n-l}\binom{n}{l}\\ &= \frac{1}{6}\sum_{n=l}^{6}\binom{n}{l}\left(\frac{1}{2}\right)^n, \qquad l = 1, 2, \ldots, 6,\end{aligned}$$

hence

$$\mathbb{P}(Z=l) = \frac{1}{6}\sum_{n=\mathrm{Max}(1,l)}^{6}\binom{n}{l}\left(\frac{1}{2}\right)^n, \qquad l = 0, 1, \ldots, 6.$$

Note that we have

$$\begin{aligned}\sum_{l=0}^{6}\mathbb{P}(Z=l) &= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^n + \frac{1}{6}\sum_{l=1}^{6}\sum_{n=l}^{6}\binom{n}{l}\left(\frac{1}{2}\right)^n\\ &= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^n + \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^n\sum_{l=1}^{n}\binom{n}{l}\\ &= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^n\sum_{l=0}^{n}\binom{n}{l}\\ &= 1,\end{aligned}$$

where we used the identity $\displaystyle\sum_{l=0}^{n}\binom{n}{l} = (1+1)^n = 2^n$.

c) We have

$$
\begin{aligned}
\mathbb{E}[Z] &= \sum_{l=0}^{6} l\mathbb{P}(Z=l) = \frac{1}{6}\sum_{l=0}^{6} l \sum_{n=l}^{6}\left(\frac{1}{2}\right)^{n}\binom{n}{l} \\
&= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n} l\binom{n}{l} = \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n}\frac{n!}{(n-l)!(l-1)!} \\
&= \frac{1}{6}\sum_{n=1}^{6} n\left(\frac{1}{2}\right)^{n}\sum_{l=0}^{n-1}\frac{(n-1)!}{(n-1-l)!l!} = \frac{1}{6}\sum_{n=1}^{6} n\left(\frac{1}{2}\right)^{n}\sum_{l=0}^{n-1}\binom{n-1}{l} \\
&= \frac{1}{6\times 2}\sum_{n=1}^{6} n = \frac{1}{6\times 2}\frac{6\times(6+1)}{2} = \frac{7}{4},
\end{aligned}
$$

which recovers (S.1), where we applied (A.5). We also have

$$
\begin{aligned}
&\mathbb{E}[Z^2] = \sum_{l=0}^{6} l^2\mathbb{P}(Z=l) \\
&= \frac{1}{6}\sum_{l=0}^{6} l^2 \sum_{n=l}^{6}\left(\frac{1}{2}\right)^{n}\binom{n}{l} = \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n} l^2\binom{n}{l} \\
&= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n}\frac{(l-1+1)n!}{(n-l)!(l-1)!} \\
&= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n}\frac{(l-1)n!}{(n-l)!(l-1)!} + \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n}\frac{n!}{(n-l)!(l-1)!} \\
&= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=2}^{n}\frac{n!}{(n-l)!(l-2)!} + \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=1}^{n}\frac{n!}{(n-l)!(l-1)!} \\
&= \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n} n(n-1)\sum_{l=0}^{n-2}\frac{(n-2)!}{(n-2-l)!l!} + \frac{1}{6}\sum_{n=1}^{6}\left(\frac{1}{2}\right)^{n}\sum_{l=0}^{n-1}\frac{n!}{(n-1-l)!l!} \\
&= \frac{1}{6\times 2^2}\sum_{n=1}^{6} n(n-1) + \frac{1}{6\times 2}\sum_{n=1}^{6} n \\
&= \frac{1}{6\times 2^2}\sum_{n=1}^{6} n^2 - \frac{1}{6\times 2^2}\sum_{n=1}^{6} n + \frac{1}{6\times 2}\sum_{n=1}^{6} n \\
&= \frac{1}{6\times 2^2}\sum_{n=1}^{6} n^2 + \frac{1}{6\times 2^2}\sum_{n=1}^{6} n \\
&= \frac{1}{6\times 2^2}\frac{6(6+1)(2\times 6+1)}{6} + \frac{1}{6\times 2^2}\frac{6(6+1)}{2} = \frac{14}{3},
\end{aligned}
$$

which recovers (S.3).

Exercise 1.4

a) We assume that the sequence of Bernoulli trials is represented by a family $(X_k)_{k\geqslant 1}$ of independent Bernoulli random variables with distribution

$$\mathbb{P}(X_k = 1) = p, \qquad \mathbb{P}(X_k = 0) = 1 - p, \qquad\qquad k \geqslant 1.$$

We have

$$Z = X_1 + X_2 + \cdots + X_N = \sum_{k=1}^{N} X_k,$$

and, since $\mathbb{E}[X_k] = p$,

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}\left[\sum_{k=1}^{N} X_k\right] = \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{n} X_k\right] \mathbb{P}(N = n)\\ &= \sum_{n\geqslant 0} \left(\sum_{k=1}^{n} \mathbb{E}[X_k]\right) \mathbb{P}(N = n) = p \sum_{n\geqslant 0} n\mathbb{P}(N = n) = p\mathbb{E}[N].\end{aligned}$$

Note that N need not have the Poisson distribution for the above equality to hold.

Next, the expectation of the Poisson random variable N with parameter $\lambda > 0$ is given as in (1.6.4) by

$$\begin{aligned}\mathbb{E}[N] &= \sum_{n\geqslant 0} n\mathbb{P}(N = n) = \mathrm{e}^{-\lambda} \sum_{n\geqslant 0} n\frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} \frac{\lambda^n}{(n-1)!}\\ &= \lambda\mathrm{e}^{-\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!} = \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda, \qquad (S.4)\end{aligned}$$

where we used the exponential series (A.1), hence

$$\mathbb{E}[Z] = p\lambda.$$

Concerning the variance we have, since $\mathbb{E}[X_k^2] = p$,

$$\begin{aligned}\mathbb{E}[Z^2] &= \mathbb{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2\right]\\ &= \sum_{n\geqslant 0} \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] \mathbb{P}(N = n)\\ &= \sum_{n\geqslant 0} \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)\left(\sum_{l=1}^{n} X_l\right)^2\right] \mathbb{P}(N = n)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k,l=1}^{n} X_k X_l\right] \mathbb{P}(N=n) \\
&= \sum_{n\geqslant 0} \mathbb{E}\left[\left(\sum_{k=1}^{n} (X_k)^2 + \sum_{1\leqslant k\neq l\leqslant n} X_k X_l\right)\right] \mathbb{P}(N=n) \\
&= \sum_{n\geqslant 0} \left(\mathbb{E}\left[\sum_{k=1}^{n} (X_k)^2\right] + \sum_{1\leqslant k\neq l\leqslant n} \mathbb{E}[X_k]\mathbb{E}[X_l]\right) \mathbb{P}(N=n) \\
&= \sum_{n\geqslant 0} \left(np + n(n-1)p^2\right) \mathbb{P}(N=n) \\
&= p(1-p) \sum_{n\geqslant 0} n\mathbb{P}(N=n) + p^2 \sum_{n\geqslant 0} n^2 \mathbb{P}(N=n) \\
&= p(1-p)\mathbb{E}[N] + p^2\mathbb{E}[N^2].
\end{aligned}
$$

Again, the above equality holds without requiring that N has the Poisson distribution.

Next, we have

$$
\begin{aligned}
&\mathbb{E}[N^2] = \sum_{n\geqslant 0} n^2 \mathbb{P}(N=n) = \mathrm{e}^{-\lambda} \sum_{n\geqslant 0} n^2 \frac{\lambda^n}{n!} \\
&= \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} n \frac{\lambda^n}{(n-1)!} = \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} (n-1) \frac{\lambda^n}{(n-1)!} + \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} \frac{\lambda^n}{(n-1)!} \\
&= \mathrm{e}^{-\lambda} \sum_{n\geqslant 2} \frac{\lambda^n}{(n-2)!} + \mathrm{e}^{-\lambda} \sum_{n\geqslant 1} \frac{\lambda^n}{(n-1)!} = \lambda^2 \mathrm{e}^{-\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!} + \lambda \mathrm{e}^{-\lambda} \sum_{n\geqslant 0} \frac{\lambda^n}{n!} \\
&= \lambda + \lambda^2,
\end{aligned}
$$

hence

$$
\operatorname{Var}[N] = \mathbb{E}[N^2] - (\mathbb{E}[N])^2 = \lambda, \tag{S.5}
$$

and

$$
\begin{aligned}
\operatorname{Var}[Z] &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 \\
&= p(1-p)\mathbb{E}[N] + p^2\mathbb{E}[N^2] - (p\mathbb{E}[N])^2 \\
&= p(1-p)\mathbb{E}[N] + p^2 \operatorname{Var}[N] \\
&= \lambda p(1-p) + \lambda p^2 \\
&= p\lambda.
\end{aligned}
$$

b) For $l \in \mathbb{N}$, using (1.3.1) with $B = \{Z = l\}$ and the fact that $\sum_{k=1}^{n} X_k$ has a binomial distribution with parameter (n, p), we have

$$\begin{aligned}
\mathbb{P}(Z=l) &= \sum_{n\geqslant 0}\mathbb{P}\left(\sum_{k=1}^{N}X_k=l \;\Big|\; N=n\right)\mathbb{P}(N=n)\\
&= \sum_{n\geqslant 0}\mathbb{P}\left(\sum_{k=1}^{n}X_k=l\right)\mathbb{P}(N=n)\\
&= \mathrm{e}^{-\lambda}\sum_{n\geqslant l}\binom{n}{l}p^l(1-p)^{n-l}\frac{\lambda^n}{n!} = \frac{p^l}{l!}\mathrm{e}^{-\lambda}\sum_{n\geqslant l}\frac{1}{(n-l)!}(1-p)^{n-l}\lambda^n\\
&= \frac{(\lambda p)^l}{l!}\mathrm{e}^{-\lambda}\sum_{n\geqslant 0}(1-p)^n\frac{\lambda^n}{n!} = \frac{(\lambda p)^l}{l!}\mathrm{e}^{-\lambda}\mathrm{e}^{(1-p)\lambda} = \frac{(\lambda p)^l}{l!}\mathrm{e}^{-p\lambda},
\end{aligned}$$

hence Z has a Poisson distribution with parameter $p\lambda$. This result is known as *thinning* of the Poisson distribution with parameter $\lambda > 0$.

c) From Question (b), Z is a Poisson random variable with parameter $p\lambda$, hence from (S.4) and (S.5) we have $\mathbb{E}[Z] = \mathrm{Var}[Z] = p\lambda$.

Exercise 1.5 Using the relation $\mathbb{P}(X < Y) = \lambda/(\lambda+\mu)$ that follows from (1.5.9), we can compute the following conditional expectation in the same way as in (1.6.6) of Lemma 1.10:

$$\begin{aligned}
&\mathbb{E}[\min(X,Y) \mid X<Y] = \mathbb{E}[X \mid X<Y] = \frac{1}{\mathbb{P}(X<Y)}\mathbb{E}\left[X\mathbb{1}_{\{X<Y\}}\right]\\
&= \mu(\lambda+\mu)\int_0^\infty \mathrm{e}^{-\mu y}\int_0^y x\mathrm{e}^{-\lambda x}dxdy\\
&= (\lambda+\mu)\frac{\mu}{\lambda^2}\int_0^\infty \mathrm{e}^{-\mu y}(1-(1+\lambda y)\mathrm{e}^{-\lambda y})dy\\
&= (\lambda+\mu)\frac{\mu}{\lambda^2}\left(\int_0^\infty \mathrm{e}^{-\mu y}dy - \int_0^\infty \mathrm{e}^{-\mu y}\mathrm{e}^{-\lambda y}dy - \lambda\int_0^\infty \mathrm{e}^{-\mu y}y\mathrm{e}^{-\lambda y})dy\right)\\
&= (\lambda+\mu)\frac{\mu}{\lambda^2}\left(\frac{1}{\mu} - \frac{1}{\lambda+\mu} - \frac{\lambda}{(\lambda+\mu)^2}\right)\\
&= (\lambda+\mu)\frac{\mu}{\lambda^2}\left(\frac{(\lambda+\mu)^2}{\mu(\lambda+\mu)^2} - \frac{\mu(\lambda+\mu)}{\mu(\lambda+\mu)^2} - \frac{\lambda\mu}{\mu(\lambda+\mu)^2}\right)\\
&= (\lambda+\mu)\frac{\mu}{\lambda^2}\left(\frac{(\lambda+\mu)^2-\mu(\lambda+\mu)-\lambda\mu}{\mu(\lambda+\mu)^2}\right)\\
&= \frac{1}{\lambda+\mu} = \mathbb{E}[\min(X,Y)].
\end{aligned}$$

Note that (1.8.10) could also be obtained as in (1.6.7) or from point (iv) page 47, since the random variable $\min(X,Y)$ is actually independent of the event $\{X<Y\}$. Indeed, for any $a > 0$ we have

$$\begin{aligned}
\mathbb{P}(\min(X,Y) > a \text{ and } X<Y) &= \lambda\mu\int_a^\infty \mathrm{e}^{-\mu y}\int_a^y \mathrm{e}^{-\lambda x}dxdy\\
&= \mu\int_a^\infty \mathrm{e}^{-\mu y}(\mathrm{e}^{-\lambda a}-\mathrm{e}^{-\lambda y})dy
\end{aligned}$$

$$= \mu e^{-\lambda a} \int_a^\infty e^{-\mu y} dy - \mu \int_a^\infty e^{-(\lambda+\mu) y} dy$$
$$= \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)a}$$
$$= \mathbb{P}(X < Y)\mathbb{P}(\min(X,Y) > a).$$

Exercise 1.6 Since U is uniformly distributed given L over the interval $[0, L]$, we have

$$\varphi_{U|L=y}(x) = \frac{1}{y}\mathbb{1}_{[0,y]}(x), \qquad x \in \mathbb{R}, \quad y > 0,$$

hence by the definition (1.5.7) of the conditional density $\varphi_{U|L=y}(x)$ we have

$$\begin{aligned}\varphi_{(U,L)}(x,y) &= \varphi_{U|L=y}(x)\varphi_L(y) \\ &= \frac{1}{y}\mathbb{1}_{[0,y]}(x) y e^{-y} \mathbb{1}_{[0,\infty)}(y) \\ &= \mathbb{1}_{[0,y]}(x)\mathbb{1}_{[0,\infty)}(y) e^{-y}. \end{aligned} \qquad \text{(S.6)}$$

Next, we want to determine the density function $(x,y) \longmapsto \varphi_{(U,L-U)}(x,y)$. For this, for all bounded functions $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$, we rewrite the expectation

$$\mathbb{E}[h(U, L-U)] = \int_{-\infty}^\infty \int_{-\infty}^\infty h(x,z)\varphi_{(U,L-U)}(x,z) dx dz \qquad \text{(S.7)}$$

as

$$\begin{aligned}\mathbb{E}[h(U, L-U)] &= \int_{-\infty}^\infty \left(\int_{-\infty}^\infty h(x, y-x)\varphi_{(U,L)}(x,y) dy\right) dx \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty h(x,z)\varphi_{(U,L)}(x, x+z) dx dz, \end{aligned} \qquad \text{(S.8)}$$

using the change of variable $(x,z) = (x, y-x)$. Identifying (S.7) with (S.8) yields the relation

$$\varphi_{(U,L-U)}(x,z) = \varphi_{(U,L)}(x, x+z),$$

hence from (S.6) we get

$$\begin{aligned}\varphi_{(U,L-U)}(x,z) &= \varphi_{(U,L)}(x, x+z) \\ &= \mathbb{1}_{[0,x+z]}(x)\mathbb{1}_{[0,\infty)}(x+z) e^{-x-z} \\ &= \mathbb{1}_{[0,\infty)}(x)\mathbb{1}_{[0,\infty)}(z) e^{-x-z}, \end{aligned}$$

since

$$\begin{aligned}\{0 \leqslant x \leqslant x+z \ \ \& \ \ x+z \geqslant 0\} &\Longleftrightarrow \{0 \leqslant x \leqslant x+z\} \\ &\Longleftrightarrow \{x \geqslant 0 \ \& \ z \geqslant 0\}. \end{aligned}$$

We refer to Theorem 12.7 page 92 of Jacod and Protter (2000) for the general version of the above change of variable formula.

Exercise 1.7

a) Assuming that X and Y are independent Poisson random variables with parameters λ and μ, we have

$$\begin{aligned}\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n} \mathbb{P}(X=k \text{ and } \ X+Y=n) \\ &= \sum_{k=0}^{n} \mathbb{P}(X=k \text{ and } Y=n-k) \\ &= \sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k) = \mathrm{e}^{-\lambda-\mu}\sum_{k=0}^{n}\frac{\lambda^k}{k!}\frac{\mu^{n-k}}{(n-k)!} \\ &= \mathrm{e}^{-\lambda-\mu}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k} = \mathrm{e}^{-\lambda-\mu}\frac{(\lambda+\mu)^n}{n!}, \qquad \text{(S.9)}\end{aligned}$$

hence $X+Y$ has a Poisson distribution with parameter $\lambda+\mu$. This result can be recovered from the probability generating function (1.7.4) of the Poisson distribution and Relation (1.7.9), as

$$G_{X+Y}(s) = G_X(s)G_Y(s) = \mathrm{e}^{\lambda(s-1)}\mathrm{e}^{\mu(s-1)} = \mathrm{e}^{(\lambda+\mu)(s-1)}, \quad -1 \leqslant s \leqslant 1.$$

b) We have

$$\begin{aligned}&\mathbb{P}(X=k \mid X+Y=n) = \frac{\mathbb{P}(X=k \text{ and } X+Y=n)}{\mathbb{P}(X+Y=n)} \\ &= \frac{\mathbb{P}(X=k \text{ and } Y=n-k)}{\mathbb{P}(X+Y=n)} = \frac{\mathbb{P}(X=k)\mathbb{P}(Y=n-k)}{\mathbb{P}(X+Y=n)} \\ &= \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\mathrm{e}^{-\mu}\frac{\mu^{n-k}}{(n-k)!}\left(\mathrm{e}^{-\lambda-\mu}\frac{(\lambda+\mu)^n}{n!}\right)^{-1} \\ &= \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}, \qquad \text{(S.10)}\end{aligned}$$

hence, given $X+Y=n$, the random variable X has a binomial distribution with parameters n and $\lambda/(\lambda+\mu)$.

c) In this case, using the exponential probability density

$$\varphi_\Lambda(x) = \theta\mathbb{1}_{[0,\infty)}(x)\mathrm{e}^{-\theta x}, \qquad x \in \mathbb{R},$$

satisfying

$$\mathbb{P}(\Lambda \leqslant x) = \int_{-\infty}^{x} \varphi_\Lambda(y)dy = \int_0^x \varphi_\Lambda(y)dy, \qquad x \in \mathbb{R},$$

and

$$d\mathbb{P}(\Lambda \leqslant x) = \varphi_\Lambda(x)dx,$$

we have

$$\begin{aligned}
\mathbb{P}(X = k) &= \int_0^\infty \mathbb{P}(X = k \mid \Lambda = x) d\mathbb{P}(\Lambda \leqslant x) \\
&= \int_0^\infty \mathbb{P}(X = k \mid \Lambda = x)\varphi_\Lambda(x)dx \\
&= \theta \int_0^\infty \frac{x^k}{k!} \mathrm{e}^{-(\theta+1)x} dx \\
&= \frac{\theta}{(\theta+1)^{k+1}} \int_0^\infty \frac{y^k}{k!} \mathrm{e}^{-y} dy \\
&= \frac{\theta}{(\theta+1)^{k+1}} \\
&= \left(1 - \frac{1}{\theta+1}\right)\left(\frac{1}{\theta+1}\right)^k,
\end{aligned}$$

i.e. X and Y are independent random variables with same geometric distribution with parameter $p = 1/(\theta+1)$. Therefore $X + Y$ has a negative binomial distribution with parameter $(r, p) = (2, 1/(\theta+1))$, cf. (1.5.12), *i.e.*,

$$\mathbb{P}(X + Y = k) = (k+1)\left(1 - \frac{1}{\theta+1}\right)^2 \left(\frac{1}{\theta+1}\right)^k, \qquad k \in \mathbb{N}.$$

Hence we have

$$\begin{aligned}
\mathbb{P}(X = k \mid X + Y = n) &= \frac{\mathbb{P}(X = k \text{ and } X + Y = n)}{\mathbb{P}(X + Y = n)} \\
&= \frac{\mathbb{P}(X = k \text{ and } Y = n - k)}{\mathbb{P}(X + Y = n)} \\
&= \frac{\mathbb{P}(X = k)\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} \\
&= \left(1 - \frac{1}{\theta+1}\right)\left(\frac{1}{\theta+1}\right)^k \left(1 - \frac{1}{\theta+1}\right)\left(\frac{1}{\theta+1}\right)^{n-k} \\
&\quad \times \frac{1}{n+1}\left(1 - \frac{1}{\theta+1}\right)^{-2}\left(\frac{1}{\theta+1}\right)^{-n} \\
&= \frac{1}{n+1}, \qquad k = 0, 1, \ldots, n,
\end{aligned}$$

which shows that the distribution of X given $X+Y=n$ is the discrete uniform distribution on $\{0,1,\ldots,n\}$.

d) In case X and Y have the same parameter, *i.e.* $\lambda=\mu$, we have

$$\mathbb{P}(X=k \mid X+Y=n)=\binom{n}{k}\frac{1}{2^n}, \quad k=0,1,\ldots,n,$$

which becomes independent of λ. Hence, when λ is represented by a random variable Λ with probability density $x\longmapsto\varphi_\Lambda(x)$ on $\mathbb{R}_+$, from (S.10) we get

$$\begin{aligned}\mathbb{P}(X=k \mid X+Y=n)&=\int_0^\infty \mathbb{P}(X=k \mid X+Y=n \text{ and } \Lambda=x)d\mathbb{P}(\Lambda\leqslant x)\\
&=\int_0^\infty \mathbb{P}(X=k \mid X+Y=n \text{ and } \Lambda=x)\varphi_\Lambda(x)dx\\
&=\int_0^\infty \mathbb{P}(X=k \mid X+Y=n)\varphi_\Lambda(x)dx\\
&=\mathbb{P}(X=k \mid X+Y=n)\int_0^\infty \varphi_\Lambda(x)dx\\
&=\mathbb{P}(X=k \mid X+Y=n)=\binom{n}{k}\frac{1}{2^n}, \qquad k=0,1,\ldots,n.\end{aligned}$$

This relation holds regardless of the expression of the probability density φ_Λ, and in particular when Λ has an exponential distribution with parameter $\theta>0$.

Exercise 1.8 Let C_1 denote the color of the first drawn pen, and let C_2 denote the color of the second drawn pen. We have

$$\mathbb{P}(C_1=R)=\mathbb{P}(C_1=G)=\frac{1}{2},$$

and

$$\mathbb{P}(C_2=R \text{ and } C_1=R)=\frac{2}{3}, \qquad \mathbb{P}(C_2=R \text{ and } C_1=G)=\frac{1}{3}.$$

On the other hand, we have

$$\begin{aligned}&\mathbb{P}(C_2=R)=\mathbb{P}(C_2=R \text{ and } C_1=R)+\mathbb{P}(C_2=R \text{ and } C_1=G)\\
&=\mathbb{P}(C_2=R \mid C_1=R)\mathbb{P}(C_1=R)+\mathbb{P}(C_2=R \mid C_1=G)\mathbb{P}(C_1=G)\\
&=\frac{2}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{2}=\frac{1}{2},\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}(C_2=G)=\mathbb{P}(C_2=G \text{ and } C_1=R)+\mathbb{P}(C_2=G \text{ and } C_1=G)\\
&=\mathbb{P}(C_2=G \mid C_1=R)\mathbb{P}(C_1=R)+\mathbb{P}(C_2=G \mid C_1=G)\mathbb{P}(C_1=G)\end{aligned}$$

$$= \frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2} = \frac{1}{2}.$$

Finally, the probability we wish to compute is

$$\begin{aligned}
\mathbb{P}(C_1 = R \mid C_2 = R) &= \frac{\mathbb{P}(C_1 = R \text{ and } C_2 = R)}{\mathbb{P}(C_2 = R)} \\
&= \mathbb{P}(C_2 = R \mid C_1 = R)\frac{\mathbb{P}(C_1 = R)}{\mathbb{P}(C_2 = R)} = \frac{2}{3} \times \frac{1/2}{1/2} = \frac{2}{3}.
\end{aligned}$$

Interestingly, we note that although the probabilities of finding a red or green pen at the second stage remain the same as at the first stage ($\mathbb{P}(C_2 = G) = \mathbb{P}(C_1 = G) = 1/2$), the result obtained at the second stage does provide some information on the outcome at the first stage.

Exercise 1.9

a) The probability that the system operates is given by

$$\mathbb{P}(X \geqslant 2) = \binom{3}{2} p^2(1-p) + p^3 = 3p^2 - 2p^3,$$

where X is a binomial random variable with parameter $(3, p)$.

b) The probability that the system operates is given by

$$\begin{aligned}
\int_0^1 \mathbb{P}(X \geqslant 2 \mid p = x) d\mathbb{P}(p \leqslant x) &= \int_0^1 \mathbb{P}(X \geqslant 2 \mid p = x) dx \\
&= \int_0^1 (3x^2 - 2x^3) dx \\
&= 3\int_0^1 x^2 dx - 2\int_0^1 x^3 dx \\
&= \frac{1}{2},
\end{aligned}$$

similarly to Exercise 1.7-(b).

Chapter 2 - Gambling Problems

Exercise 2.1

a) By first step analysis, we have

$$f(k) = (1-2r)f(k) + rf(k+1) + rf(k-1),$$

which yields the equation

$$f(k) = \frac{1}{2}f(k+1) + \frac{1}{2}f(k-1), \qquad 1 \leqslant k \leqslant S-1, \tag{S.11}$$

with the boundary conditions $f(0) = 1$ and $f(S) = 0$, which is identical to Equation (2.2.19).

We refer to this equation as the *homogeneous equation*.

b) According to the result of (2.2.20) in Section 2.2 we know that the general solution of (S.11) has the form

$$f(k) = C_1 + C_2 k, \qquad k = 0, 1, \ldots, S$$

and after taking into account the boundary conditions we find

$$f(k) = \frac{S-k}{S}, \qquad k = 0, 1, \ldots, S.$$

Intuitively, the game is fair to both players for all values of r, so that the probability of ruin should be the same for all $r \in (0, 1/2)$, which is indeed the case.

c) By first step analysis, we have

$$\begin{aligned} h(k) &= (1-2r)(1+h(k)) + r(1+h(k+1)) + r(1+h(k-1)) \\ &= 1 + (1-2r)h(k) + rh(k+1) + rh(k-1), \end{aligned}$$

hence the equation

$$h(k) = \frac{1}{2r} + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1), \qquad 1 \leqslant k \leqslant S-1,$$

with the boundary conditions $h(0) = 0$ and $h(S) = 0$, which is identical to (2.3.7a) by changing $h(k)$ into $2r \times h(k)$ in (2.3.7a). We note that r should be strictly positive, otherwise if $r = 0$ we will have $h(k) = \infty$ for all $k = 1, 2, \ldots, S-1$ and $h(0) = h(S) = 0$.

d) After trying a solution of the form $h(k) = Ck^2$ we find

$$Ck^2 = \frac{1}{2r} + \frac{1}{2}C(k+1)^2 + \frac{1}{2}C(k-1)^2,$$

hence C should be equal to $C = -1/(2r)$, hence $k \longmapsto -k^2/(2r)$ is a particular solution.

e) Given the hint, the general solution has the form

$$h(k) = C_1 + C_2 k - \frac{k^2}{2r}, \qquad k = 0, 1, \ldots, S,$$

which gives

$$h(k) = \frac{k(S-k)}{2r}, \qquad k = 0, 1, \ldots, S, \tag{S.12}$$

after taking into account the boundary conditions.

Remark. The inclusion of draws changes the standard discrete time step length 1 into $1/(2r)$, which can be interpreted as the mean $\mathbb{E}[\tau] = 1/(2r)$ of the geometrically distributed random variable τ representing the time k spent by remaining at the same state, with

$$\mathbb{P}(\tau = k) = 2r(1-2r)^{k-1}, \qquad k \geqslant 1.$$

f) Starting from any state $\textcircled{k} \in \{1, 2, \ldots, S-1\}$, the mean duration goes to infinity when r goes to zero.

When r goes to 0 the probability $1-2r$ of a draw increases, therefore the game should take longer. Hence the above answer is compatible with intuition.

Exercise 2.2

a) Taking $a = m$ and $b = S$ in (2.3.20), we find the probability

$$\frac{1-(p/q)^{S-k}}{1-(p/q)^{S-m}} = \frac{(q/p)^S - (q/p)^k}{(q/p)^S - (q/p)^m}. \tag{S.13}$$

b) Taking $a = 0$ and $b = m$ in (2.3.20), we find the probability

$$1 - \frac{1-(p/q)^{m-k}}{1-(p/q)^m} = \frac{1-(q/p)^k}{1-(q/p)^m}. \tag{S.14}$$

c) Taking $k = m+1$ in (S.13), we find that the probability of returning to $\textcircled{m}$ after starting from $\boxed{m+1}$ is

$$\frac{(q/p)^S - (q/p)^{m+1}}{(q/p)^S - (q/p)^m}.$$

Similarly, taking $k = m-1$ in (S.14), we find that the probability of returning to $\textcircled{m}$ after starting from $\boxed{m-1}$ is

$$\frac{1-(q/p)^{m-1}}{1-(q/p)^m}.$$

Therefore, by first step analysis we find that the probability of coming back to state $\textcircled{m}$ in finite time after starting from $X_0 = m$ is

$$\begin{aligned}
&p\frac{1-(p/q)^{S-(m+1)}}{1-(p/q)^{S-m}} + q\frac{1-(q/p)^{m-1}}{1-(q/p)^m} \\
&= \frac{p(1-(p/q)^{S-(m+1)})(1-(q/p)^m) + q(1-(q/p)^{m-1})(1-(p/q)^{S-m})}{(1-(p/q)^{S-m})(1-(q/p)^m)}
\end{aligned}$$

$$= \frac{(p - q(p/q)^{S-m})(1-(q/p)^m) + (q - p(q/p)^m)(1-(p/q)^{S-m})}{(1-(p/q)^{S-m})(1-(q/p)^m)}$$
$$= \frac{1 - 2p(q/p)^m - 2q(p/q)^{S-m} + (p/q)^{S-2m}}{(1-(p/q)^{S-m})(1-(q/p)^m)}$$
$$= \frac{1 - 2p(q/p)^m - 2q(p/q)^{S-m} + (p/q)^{S-2m}}{1-(q/p)^m - (p/q)^{S-m} + (p/q)^{S-2m}}, \qquad 1 \leqslant m \leqslant S-1.$$

When $S = 2$ and $m = 1$ we find that the above probability is zero, as expected.

d) Using (2.3.12) we can similarly compute the mean time to either come back to m or reach any of the two boundaries $\{0, S\}$, whichever comes first, using first step analysis, as follows:

$$1 + \frac{p}{q-p}\left(1 - (S-m)\frac{1-q/p}{1-(q/p)^{S-m}}\right) + \frac{q}{q-p}\left(m - 1 - m\frac{1-(q/p)^{m-1}}{1-(q/p)^m}\right)$$
$$= -\frac{p(S-m)}{q-p}\frac{1-q/p}{1-(q/p)^{S-m}} + \frac{mq}{q-p}\frac{(q/p)^m - (q/p)^{m-1}}{(q/p)^m - 1}, \quad 1 \leqslant m \leqslant S-1.$$

When $S = 2$ and $m = 1$ we find that the above expression equals 1 as expected.

e) When $p = q = 1/2$ we would find

$$\frac{1}{2}\left(\frac{S-m-1}{S-m}\right) + \frac{1}{2}\left(\frac{m-1}{m}\right) = 1 - \frac{S}{2m(S-m)},$$

for the probability of coming back to state $\textcircled{m}$ in finite time after starting from $X_0 = m$, which gives $1 - 2/S$ when $S = 2m$, $m \geqslant 1$. Similarly, the mean time to either come back to m or to reach any of the two boundaries $\{0, S\}$, whichever comes first, is

$$1 + \frac{1}{2}(S-m-1) + \frac{1}{2}(m-1) = \frac{S}{2},$$

which does not depend on $m = 1, 2, \ldots, S-1$.

Exercise 2.3

a) Starting from state $\textcircled{k}$ we can reach $\boxed{k+1}$ in one time step by going up with probability p, or we can go down to $\boxed{k-1}$ with probability q, in which case we have to reach $\boxed{k+1}$ without hitting state $\textcircled{0}$ in two stages:

i) first, reach $\boxed{k}$ from $\boxed{k-1}$, with probability $\mathbb{P}(\tau_k < \tau_0 \mid X_0 = k-1)$,

ii) then, reach $\boxed{k+1}$ from $\boxed{k}$ with probability $\mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k)$.

This gives

$$\begin{aligned} p_k &= \mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k) \\ &= p + q\mathbb{P}(\tau_k < \tau_0 \mid X_0 = k-1)\mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k) \\ &= p + qp_{k-1}p_k, \qquad k = 1, 2, \ldots, S-1. \end{aligned}$$

b) We need to check that (2.3.22) is satisfied when p_k is given by

$$p_k = \frac{1-(q/p)^k}{1-(q/p)^{k+1}}, \qquad k = 0, 1, \ldots, S-1.$$

Starting from the right-hand side of (2.3.22), this means checking the following:

$$\begin{aligned} p + q\frac{1-(q/p)^{k-1}}{1-(q/p)^k}\frac{1-(q/p)^k}{1-(q/p)^{k+1}} &= p + q\frac{1-(q/p)^{k-1}}{1-(q/p)^{k+1}} \\ &= \frac{p(1-(q/p)^{k+1}) + q(1-(q/p)^{k-1})}{1-(q/p)^{k+1}} \\ &= \frac{p+q-p(q/p)^{k+1} - q(q/p)^{k-1}}{1-(q/p)^{k+1}} \\ &= \frac{p+q-q(q/p)^k - p(q/p)^k}{1-(q/p)^{k+1}} \\ &= \frac{1-(q/p)^k}{1-(q/p)^{k+1}}. \end{aligned}$$

c) We have

$$\begin{aligned} &\mathbb{P}(\tau_S < \tau_0 \mid X_0 = k) \\ &= \mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k) \\ &\quad \times \mathbb{P}(\tau_{k+2} < \tau_0 \mid X_0 = k+1) \times \cdots \times \mathbb{P}(\tau_S < \tau_0 \mid X_0 = S-1) \\ &= \prod_{l=k}^{S-1} p_l = \prod_{l=k}^{S-1} \frac{1-(q/p)^l}{1-(q/p)^{l+1}} = \frac{1-(q/p)^k}{1-(q/p)^S}, \qquad k = 0, 1, \ldots, S, \end{aligned}$$

which recovers (2.2.28).

d) In the symmetric case $p = q = 1/2$, the same reasoning as in Question (a) above shows that p_k should satisfy the equation

$$p_k = \frac{1}{2} + \frac{1}{2}p_{k-1}p_k, \qquad k = 1, 2, \ldots, S-1,$$

which rewrites as

$$p_k(2 - p_{k-1}) = 1, \qquad k = 1, 2, \ldots, S-1.$$

We check that the above equation is satisfied by

$$p_k := \frac{k}{k+1}, \qquad k = 1, 2, \ldots, S-1,$$

hence we find again

$$\begin{aligned}
&\mathbb{P}(\tau_S < \tau_0 \mid X_0 = k) \\
&= \mathbb{P}(\tau_{k+1} < \tau_0 \mid X_0 = k) \\
&\quad \times \mathbb{P}(\tau_{k+2} < \tau_0 \mid X_0 = k+1) \times \cdots \times \mathbb{P}(\tau_S < \tau_0 \mid X_0 = S-1) \\
&= \prod_{l=k}^{S-1} p_l \\
&= p_k p_{k+1} p_{k+2} \cdots p_{N-1} \\
&= \prod_{l=k}^{S-1} \frac{l}{l+1} \\
&= \frac{k}{S}, \qquad k = 0, 1, \ldots, S,
\end{aligned}$$

showing that

$$\mathbb{P}(\tau_0 < \tau_S \mid X_0 = k) = 1 - \frac{k}{S}, \qquad k = 0, 1, \ldots, S,$$

which recovers (2.2.13).

Exercise 2.4

a) We have

$$f(0) = \mathbb{P}(R_A \mid X_0 = 0) = q + pq.$$

b) We have

$$f(2) = \mathbb{P}(R_A \mid X_0 = 2) = 0,$$

and

$$\begin{aligned}
f(1) &= \mathbb{P}(R_A \mid X_0 = 1) \\
&= p\mathbb{P}(R_A \mid X_0 = 2) + q\mathbb{P}(R_A \mid X_0 = 0) \\
&= q^2 + pq^2.
\end{aligned}$$

Exercise 2.5

a) Letting $\varepsilon = r - 1$, *i.e.* $r = 1 + \varepsilon$, we have

$$h_S(k) = \frac{1}{q-p}\left(k - S\frac{1-r^k}{1-r^S}\right) = \frac{1}{p\varepsilon}\left(k - S\frac{1-(1+\varepsilon)^k}{1-(1+\varepsilon)^S}\right)$$

$$= \frac{1}{p\varepsilon}\left(k - S\frac{\sum_{i=1}^{k}\binom{k}{i}\varepsilon^i}{\sum_{i=1}^{S}\binom{S}{i}\varepsilon^i}\right) = \frac{1}{p\varepsilon}\left(\frac{k\sum_{i=1}^{S}\binom{S}{i}\varepsilon^i - S\sum_{i=1}^{k}\binom{k}{i}\varepsilon^i}{\sum_{i=1}^{S}\binom{S}{i}\varepsilon^i}\right)$$

$$= \frac{1}{p}\left(\frac{kS(S-1)/2 - Sk(k-1)/2 + k\sum_{i=3}^{S}\binom{S}{i}\varepsilon^{i-2} - S\sum_{i=3}^{k}\binom{k}{i}\varepsilon^{i-2}}{S + \varepsilon\sum_{i=2}^{S}\binom{S}{i}\varepsilon^{i-2}}\right)$$

hence

$$\lim_{r\to 1}\frac{1}{q-p}\left(k - S\frac{1-r^k}{1-r^S}\right) = \lim_{\varepsilon\to 0}\frac{1}{p\varepsilon}\left(k - S\frac{1-(1+\varepsilon)^k}{1-(1+\varepsilon)^S}\right) = k(S-k), \tag{S.15}$$

which recovers (2.3.19).

b) Letting $\varepsilon = 1 - r$, we have

$$f_S(k) = \frac{r^S - r^k}{r^S - 1} = \frac{(1+\varepsilon)^S - (1+\varepsilon)^k}{(1+\varepsilon)^S - 1}$$

$$= \frac{\sum_{i=0}^{S}\binom{S}{i}\varepsilon^i - \sum_{i=0}^{k}\binom{k}{i}\varepsilon^i}{\sum_{i=0}^{S}\binom{S}{i}\varepsilon^i - 1} = \frac{S - k + \varepsilon^2\sum_{i=2}^{k}\binom{k}{i}\varepsilon^{i-2} - \varepsilon^2\sum_{i=2}^{S}\binom{S}{i}\varepsilon^{i-2}}{S + \epsilon^2\sum_{i=2}^{S}\binom{S}{i}\varepsilon^{i-2}}$$

hence

$$\lim_{r\to 1}\frac{r^S - r^k}{r^S - 1} = \lim_{\varepsilon\to 0}\frac{(1+\varepsilon)^S - (1+\varepsilon)^k}{(1+\varepsilon)^S - 1} = \frac{S-k}{S},$$

which recovers (2.2.22).

Exercise 2.6 By first step analysis, we now have

$$f(k) = (1-\alpha-\beta)f(k) + \alpha f(k+1) + \beta f(k-1),$$

which yields

$$f(k) = \frac{\alpha}{\alpha+\beta}f(k+1) + \frac{\beta}{\alpha+\beta}f(k-1), \qquad 1 \leqslant k \leqslant S-1,$$

with $f(0) = 1$ and $f(S) = 0$. Therefore this non-symmetric problem with draw can be reduced to a non-symmetric gambling process without draw and probabilities $p = \alpha/(\alpha+\beta)$ and $q = \beta/(\alpha+\beta)$, and solution

$$f(k) = \frac{(\beta/\alpha)^k - (\beta/\alpha)^S}{1 - (\beta/\alpha)^S}, \qquad k = 0, 1, \ldots, S,$$

obtained by (2.2.12), with $\alpha \neq \beta$. For the mean game duration, by first step analysis we find the equation

$$\begin{aligned} h(k) &= (1 - \alpha - \beta)(1 + h(k)) + \alpha(1 + h(k+1)) + \beta(1 + h(k-1)) \\ &= 1 + (1 - \alpha - \beta)h(k) + \alpha h(k+1) + \beta h(k-1), \end{aligned}$$

i.e.

$$h(k) = \frac{1}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} h(k+1) + \frac{\beta}{\alpha + \beta} h(k-1), \qquad 1 \leqslant k \leqslant S - 1,$$

or

$$(\alpha + \beta)h(k) = 1 + p(\alpha + \beta)h(k+1) + q(\alpha + \beta)h(k-1),$$

with $p = \alpha/(\alpha + \beta)$ and $q = \beta/(\alpha + \beta)$ which, using (2.3.12), shows that

$$h(k) = \frac{1}{\beta - \alpha} \left(k - S \frac{1 - (\beta/\alpha)^k}{1 - (\beta/\alpha)^S} \right), \qquad k = 0, 1, \ldots, S, \tag{S.16}$$

when $\alpha \neq \beta$. As in Question (e) of Exercise 2.1, the inclusion of a draw changes the discrete time step length from 1 to $1/(\alpha + \beta)$, which is the mean $\mathbb{E}[\tau]$ of a geometrically distributed random variable τ with distribution

$$\mathbb{P}(\tau = k) = (\alpha + \beta)(1 - \alpha - \beta)^{k-1}, \qquad k \geqslant 1.$$

When α tends to β, the result of Exercise 2.1 can be recovered as on page 434. Indeed, the limit of (S.16) as $\alpha \to \beta$ is given as in (S.15) by

$$h(k) = \frac{k(S - k)}{2\beta}, \qquad k = 0, 1, \ldots, S,$$

which is consistent with (S.12).

Problem 2.7

a) We have

$$g(k) = pg(k+1) + qg(k-1), \qquad k = 1, 2, \ldots, S - 1, \tag{S.17}$$

with

$$g(0) = pg(1) + qg(0) \tag{S.18}$$

for $k = 0$, and the boundary condition $g(S) = 1$.

b) We observe that the constant function $g(k) = C$ is solution of both (S.17) and (S.18) and the boundary condition $g(S) = 1$ yields $C = 1$, hence

$$g(k) = \mathbb{P}(W \mid X_0 = k) = 1$$

for all $k = 0, 1, \ldots, S$.

c) We have

$$h(k) = 1 + ph(k+1) + qh(k-1), \qquad k = 1, 2, \ldots, S-1, \tag{S.19}$$

with

$$h(0) = 1 + ph(1) + qh(0)$$

for $k = 0$, and the boundary condition $h(S) = 0$.

d) Case $p \neq q$. The solution of the homogeneous equation

$$h(k) = ph(k+1) + qh(k-1), \qquad k = 1, 2, \ldots, S-1,$$

has the form

$$h(k) = C_1 + C_2(q/p)^k, \qquad k = 1, 2, \ldots, S-1,$$

and we can check that $k \mapsto k/(p-q)$ is a particular solution. Hence the general solution of (S.19) has the form

$$h(k) = \frac{k}{q-p} + C_1 + C_2(q/p)^k, \qquad k = 0, 1, \ldots, S,$$

with

$$\begin{cases} 0 = h(S) = \dfrac{S}{q-p} + C_1 + C_2(q/p)^S, \\ \\ ph(0) = p(C_1 + C_2) = 1 + ph(1) = 1 + p\left(\dfrac{1}{q-p} + C_1 + C_2\dfrac{q}{p}\right), \end{cases}$$

which yields

$$\begin{cases} C_1 = q\dfrac{(q/p)^S}{(p-q)^2} - \dfrac{S}{q-p}, \\ \\ C_2 = -\dfrac{q}{(p-q)^2}, \end{cases}$$

and

$$h(k) = \mathbb{E}[T_S \mid X_0 = k] = \frac{S-k}{p-q} + \frac{q}{(p-q)^2}((q/p)^S - (q/p)^k),$$

$k = 0, 1, \ldots, S$.

Case $p = q = 1/2$. The solution of the homogeneous equation is given by

$$h(k) = C_1 + C_2 k, \qquad k = 1, 2, \ldots, S-1,$$

and the general solution to (S.19) has the form

$$h(k) = -k^2 + C_1 + C_2 k, \qquad k = 1, 2, \ldots, S,$$

with

$$\begin{cases} 0 = h(S) = -S^2 + C_1 + C_2 S, \\ \dfrac{h(0)}{2} = \dfrac{C_1}{2} = 1 + \dfrac{h(1)}{2} = 1 + \dfrac{C_1 + C_2 - 1}{2}, \end{cases}$$

hence

$$\begin{cases} C_1 = S(S+1), \\ C_2 = -1, \end{cases}$$

which yields

$$h(k) = \mathbb{E}[T_S \mid X_0 = k] = (S + k + 1)(S - k), \quad k = 0, 1, \ldots, S.$$

e) When $p \neq q$ we have

$$p_k := \mathbb{P}(T_S < T_0 \mid X_0 = k) = \frac{1 - (q/p)^k}{1 - (q/p)^S}, \qquad k = 0, 1, \ldots, S,$$

and when $p = q = 1/2$ we find

$$p_k = \frac{k}{S}, \qquad k = 0, 1, \ldots, S.$$

f) The equality holds because, given that we start from state $\boxed{k+1}$ at time 1, whether $T_S < T_0$ or $T_S > T_0$ does not depend on the past of the process before time 1. In addition it does not matter whether we start from state $\boxed{k+1}$ at time 1 or at time 0.

g) We have

$$\begin{aligned} \mathbb{P}(X_1 = k+1 \mid X_0 = k \text{ and } T_S < T_0) &= \frac{\mathbb{P}(X_1 = k+1,\ X_0 = k,\ T_S < T_0)}{\mathbb{P}(X_0 = k \text{ and } T_S < T_0)} \\ &= \frac{\mathbb{P}(T_S < T_0 \mid X_1 = k+1 \text{ and } X_0 = k)\mathbb{P}(X_1 = k+1 \text{ and } X_0 = k)}{\mathbb{P}(T_S < T_0 \text{ and } X_0 = k)} \\ &= p\frac{\mathbb{P}(T_S < T_0 \mid X_0 = k+1)}{\mathbb{P}(T_S < T_0 \mid X_0 = k)} = p\frac{p_{k+1}}{p_k}, \end{aligned}$$

$k = 0, 1, \ldots, S-1$. By the result of Question (e), when $p \neq q$ we find

$$\mathbb{P}(X_1 = k+1 \mid X_0 = k \text{ and } T_S < T_0) = p\frac{1 - (q/p)^{k+1}}{1 - (q/p)^k},$$

$k = 1, 2, \ldots, S-1$, and in case $p = q = 1/2$ we get

$$\mathbb{P}(X_1 = k+1 \mid X_0 = k \text{ and } T_S < T_0) = \frac{k+1}{2k},$$

$k = 1, 2, \ldots, S-1$. Note that this probability is higher than $p = 1/2$.

h) Similarly, we have

$$\begin{aligned}
&\mathbb{P}(X_1 = k-1 \mid X_0 = k \text{ and } T_0 < T_S) \\
&\quad = \frac{\mathbb{P}(X_1 = k-1,\ X_0 = k \text{ and } T_0 < T_S)}{\mathbb{P}(X_0 = k \text{ and } T_0 < T_S)} \\
&\quad = \frac{\mathbb{P}(T_0 < T_S \mid X_1 = k-1 \text{ and } X_0 = k)\mathbb{P}(X_1 = k-1 \text{ and } X_0 = k)}{\mathbb{P}(T_0 < T_S \text{ and } X_0 = k)} \\
&\quad = q\frac{\mathbb{P}(T_0 < T_S \mid X_0 = k-1)}{\mathbb{P}(T_0 < T_S \mid X_0 = k)} \\
&\quad = q\frac{1-p_{k-1}}{1-p_k},
\end{aligned}$$

$k = 1, 2, \ldots, S-1$. When $p \neq q$ this yields

$$\mathbb{P}(X_1 = k-1 \mid X_0 = k \text{ and } T_0 < T_S) = q\frac{(q/p)^{k-1} - (q/p)^S}{(q/p)^k - (q/p)^S},$$

$k = 1, 2, \ldots, S-1$, and when $p = q = 1/2$ we find

$$\mathbb{P}(X_1 = k-1 \mid X_0 = k \text{ and } T_0 < T_S) = \frac{S+1-k}{2(S-k)},$$

$k = 1, 2, \ldots, S-1$. Note that this probability is higher than $q = 1/2$.

i) We find

$$h(k) = 1 + p\frac{p_{k+1}}{p_k}h(k+1) + \left(1 - p\frac{p_{k+1}}{p_k}\right)h(k-1), \tag{S.20}$$

$k = 1, 2, \ldots, S-1$, or, due to the first step equation $p_k = pp_{k+1} + qp_{k-1}$,

$$p_k h(k) = p_k + pp_{k+1}h(k+1) + qp_{k-1}h(k-1), \quad k = 1, 2, \ldots, S-1,$$

with the boundary condition $h(S) = 0$. When $p = q = 1/2$ we have $p_k = k/S$ by Question (e), hence (S.20) becomes

$$h(k) = 1 + \frac{k+1}{2k}h(k+1) + \frac{k-1}{2k}h(k-1),$$

$k = 1, 2, \ldots, S-1$.

j) We have to solve the equation

$$kh(k) = k + \frac{1}{2}(k+1)h(k+1) + \frac{1}{2}(k-1)h(k-1), \qquad k = 1, 2, \ldots, S-1,$$

with the boundary condition $h(S) = 0$. Letting $g(k) := kh(k)$ we check that $g(k)$ satisfies

$$g(k) = k + \frac{1}{2}g(k+1) + \frac{1}{2}g(k-1), \qquad k = 1, 2, \ldots, S-1, \tag{S.21}$$

with the boundary conditions $g(0) = 0$ and $g(S) = 0$. We check that $g(k) = Ck^3$ is a particular solution when $C = -1/3$, hence the solution of (S.21) has the form

$$g(k) = -\frac{1}{3}k^3 + C_1 + C_2 k,$$

by the homogeneous solution given in Section 2.3, where C_1 and C_2 are determined by the boundary conditions

$$0 = g(0) = C_1$$

and

$$0 = g(S) = -\frac{1}{3}S^3 + C_1 + C_2 S,$$

i.e. $C_1 = 0$ and $C_2 = S^2/3$. Consequently, we have

$$g(k) = \frac{k}{3}(S^2 - k^2), \qquad k = 0, 1, \ldots, S,$$

hence we have

$$h(k) = \mathbb{E}[T_S \mid X_0 = k,\ T_S < T_0] = \frac{S^2 - k^2}{3}, \qquad k = 1, 2, \ldots, S.$$

The conditional expectation $h(0)$ is actually undefined because the event $\{X_0 = 0,\ T_S < T_0\}$ has probability 0.

Problem 2.8

a) We have $g(0) = 0$ and $g(S) = g(S+1) = 1$.
b) We have

$$g(k) = pg(k+2) + rg(k) + 2pg(k-1), \qquad 1 \leqslant k \leqslant S-1,$$

and $g(0) = 0$ and $g(S) = g(S+1) = 1$.
c) Trying a solution of the form $g(k) = C\lambda^k$ shows that λ must satisfy

$$\lambda^k = p\lambda^{k+2} + r\lambda^k + 2p\lambda^{k-1}, \qquad 1 \leqslant k \leqslant S-1,$$

i.e.

$$\lambda = p\lambda^3 + r\lambda + 2p, \qquad 1 \leqslant k \leqslant S-1,$$

hence, using the relation $3p + r = 1$ we find

$$\lambda^3 - 3\lambda + 2 = (\lambda^2 - 2\lambda + 1)(\lambda + 2) = (\lambda - 1)^2(\lambda + 2) = 0, \qquad k = 1, 2, \ldots, S-1,$$

with solutions $\lambda \in \{1, -2\}$. Hence we can now try a solution of the form

$$g(k) = C_1 + C_2 k + C_3(-2)^k$$

with the boundary conditions

$$\begin{cases} g(0) = 0 = C_1 + C_3, \\ g(S) = 1 = C_1 + C_2 S + (-2)^S C_3, \\ g(S+1) = 1 = C_1 + C_2(S+1) - 2 \times (-2)^S C_3, \end{cases}$$

i.e.

$$\begin{cases} C_1 = \dfrac{1}{1 - (3S+1)(-2)^S} \\ C_2 = -\dfrac{3(-2)^S}{1 - (3S+1)(-2)^S} \\ C_3 = -\dfrac{1}{1 - (3S+1)(-2)^S}, \end{cases}$$

which yields

$$\begin{aligned} g(k) &= \frac{1}{1 - (3S+1)(-2)^S} - \frac{3k(-2)^S}{1 - (3S+1)(-2)^S} - \frac{(-2)^k}{1 - (3S+1)(-2)^S} \\ &= \frac{1 - 3k(-2)^S - (-2)^k}{1 - (3S+1)(-2)^S}, \qquad k = 0, 1, \ldots, S+1. \end{aligned} \tag{S.22}$$

In the graph of Figure S.1 the ruin probability (S.22) is plotted as a function of k for $p = 1/3$ and $r = 0$.

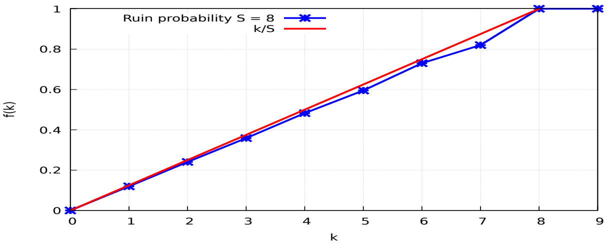


Fig. S.1: Ruin probability $f_8(k)$ as a function of $X_0 = k \in [0, 9]$.

d) Intuitively, this problem corresponds to a fair game for all values of p since the expected gain $\mathbb{E}[X_n - X_{n-1}] = 0$ vanishes* for all $n \geqslant 1$, so that the probability of ruin $f(k)$ should be the same for all $p \in (0, 1/2)$, which is indeed the case in (S.22).
e) We have $h(0) = h(S) = h(S+1) = 0$.
f) We find

$$h(k) = 1 + ph(k+2) + rh(k) + 2ph(k-1), \qquad 1 \leqslant k \leqslant S-1.$$

g) After checking that $k \mapsto Ck$ cannot be a solution we search for a particular solution of the form $h(k) = Ck^2$, which yields

$$Ck^2 = 1 + pC(k+2)^2 + rCk^2 + 2pC(k-1)^2, \qquad 1 \leqslant k \leqslant S-1,$$

i.e.

$$0 = 1 + pC(2k+4) + 2pC(-2k+1), \qquad 1 \leqslant k \leqslant S-1,$$

or $C = -1/(6p)$.
h) The general solution takes the form

$$h(k) = -\frac{k^2}{6p} + C_1 + C_2 k + C_3(-2)^k$$

with the boundary conditions

$$\begin{cases} h(0) = 0 = C_1 + C_3, \\ \\ h(S) = 0 = -\dfrac{S^2}{6p} + C_1 + C_2 S + (-2)^S C_3, \\ \\ h(S+1) = 0 = -\dfrac{(S+1)^2}{6p} + C_1 + C_2(S+1) - 2 \times (-2)^S C_3, \end{cases}$$

i.e.

$$\begin{cases} C_1 = -\dfrac{S(S+1)}{6p(1 - 3S(-2)^S - (-2)^S)} \\ \\ C_2 = \dfrac{S}{6p} + \dfrac{(S+1)(1-(-2)^S)}{6p(1 - 3S(-2)^S - (-2)^S)} \\ \\ C_3 = \dfrac{S(S+1)}{6p(1 - 3S(-2)^S - (-2)^S)}, \end{cases}$$

which yields

* We are in presence of a fair game in which probabilities of going up and down are not equal, while increments are by two units and decrements are by one unit.

$$h(k) = -\frac{k^2}{6p} - \frac{S(S+1)}{6p(1-(3S+1)(-2)^S)} + k\left(\frac{S}{6p} + \frac{(S+1)(1-(-2)^S)}{6p(1-(3S+1)(-2)^S)}\right)$$
$$+ \frac{S(S+1)(-2)^k}{6p(1-(3S+1)(-2)^S)}$$
$$= \frac{k(S-k)}{6p} + \frac{-S(S+1) + k(S+1)(1-(-2)^S) + S(S+1)(-2)^k}{6p(1-(3S+1)(-2)^S)},$$
$$= \frac{k(S-k)}{6p} + \frac{k(S+1)(1-(-2)^S) - S(S+1)(1-(-2)^k)}{6p(1-(3S+1)(-2)^S)}$$
$$= \frac{k(S-k)}{6p} + \frac{k(S+1) - S(S+1)}{6p(1-(3S+1)(-2)^S)} + \frac{-k(S+1)(-2)^S + S(S+1)(-2)^k}{6p(1-(3S+1)(-2)^S)}$$
$$= \frac{k(S-k)}{6p} + (S+1)\frac{S(-2)^k - k(-2)^S - (S-k)}{6p(1-(3S+1)(-2)^S)}, \qquad \text{(S.23)}$$

$k = 0, 1, \ldots, S+1$. In the graph of Figure S.2 the mean game duration (S.23) is plotted as a function of k with $p = 1/3$.

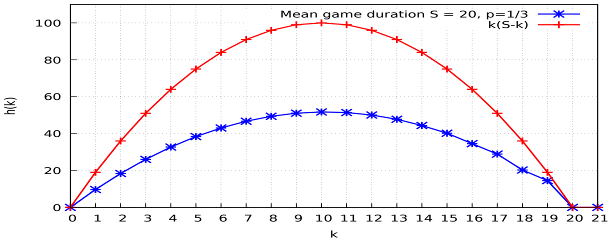


Fig. S.2: Mean game duration $h_{20}(k)$ as a function of $X_0 = k \in [0, 20]$.

i) Starting from a state $k \in \{1, 2, \ldots, S-1\}$, the mean duration goes to infinity when p goes to zero. Indeed, when p goes to 0 the probability $1-3p$ of a draw increases and the game should take longer. Hence the above answer is compatible with intuition.
j) When S tends to infinity, the ruin probability starting from $k \geqslant 1$ tends to 0 and the mean game duration tends to infinity.

Problem 2.9

a) We have $h(0) = h(1) = 0$.
b) Here we need to increment the value of τ by one unit in the first step analysis argument, which yields

$$h(k) = p\mathbb{E}\big[(1+\tau)^2 \,\big|\, X_0 = k+1\big] + q\mathbb{E}\big[(1+\tau)^2 \,\big|\, X_0 = k-1\big]$$

$$
\begin{aligned}
&= p\mathbb{E}\big[1+2\tau+\tau^2 \,\big|\, X_0=k+1\big] + q\mathbb{E}\big[1+2\tau+\tau^2 \,\big|\, X_0=k-1\big] \\
&= 1 + p\mathbb{E}\big[2\tau+\tau^2 \,\big|\, X_0=k+1\big] + q\mathbb{E}\big[2\tau+\tau^2 \,\big|\, X_0=k-1\big] \\
&= 1 + 2p\mathbb{E}\big[\tau \,\big|\, X_0=k+1\big] + 2q\mathbb{E}\big[\tau \,\big|\, X_0=k-1\big] \\
&\quad + p\mathbb{E}\big[\tau^2 \,\big|\, X_0=k+1\big] + q\mathbb{E}\big[\tau^2 \,\big|\, X_0=k-1\big] \\
&= 1 + 2p\mathbb{E}\big[\tau \,\big|\, X_0=k+1\big] + 2q\mathbb{E}\big[\tau \,\big|\, X_0=k-1\big] + ph(k+1) + qh(k-1),
\end{aligned}
$$

$1 \leqslant k \leqslant S-1$, where we used the relation $p+q=1$.

c) We have

$$
\begin{aligned}
h(k) &= 1 + \mathbb{E}\big[\tau \,\big|\, X_0=k+1\big] + \mathbb{E}\big[\tau \,\big|\, X_0=k-1\big] + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1) \\
&= 1 + (S-k-1)(k+1) + (S-k+1)(k-1) + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1) \\
&= 1 + (S-k-1)k + (S-k-1) + (S-k+1)k - (S-k+1) \\
&\quad + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1) \\
&= -1 + 2(S-k)k + \frac{1}{2}h(k+1) + \frac{1}{2}h(k-1),
\end{aligned}
$$

$1 \leqslant k \leqslant S-1$.

d) We have

$$
h(k) = C_1 + C_2 k + \frac{2k^2}{3} - \frac{k^3}{3}(2S-k)
$$

with

$$
0 = h(0) = C_1
$$

and

$$
0 = h(S) = C_2 S + \frac{2S^2}{3} - \frac{S^4}{3}
$$

hence

$$
C_2 = -\frac{2S}{3} + \frac{S^3}{3} = \frac{S}{3}\left(S^2-2\right)
$$

and

$$
h(k) = k\frac{S}{3}\left(S^2-2\right) + \frac{2k^2}{3} - \frac{k^3}{3}(2S-k).
$$

e) We have

$$
\begin{aligned}
v(k) &= h(k) - k^2(S-k)^2 \\
&= k\frac{S}{3}\left(S^2-2\right) + \frac{2k^2}{3} - \frac{k^3}{3}(2S-k) - k^2(S-k)^2 \\
&= k\left(\frac{S}{3}\left(S^2-2\right) + \frac{2k}{3} - S\frac{k^2}{3} - \frac{k^2}{3}(S-k) - k(S-k)^2\right) \\
&= \frac{k}{3}\left(S\left(S^2-2\right) + 2k - Sk^2 - k^2(S-k) - 3k(S-k)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{k}{3}\left((S-k)\left(S^2-2\right)+Sk(S-k)-k^2(S-k)-3k(S-k)^2\right)\\
&= \frac{k}{3}\left((S-k)\left(S^2-2+Sk\right)-k^2(S-k)-3k(S-k)^2\right)\\
&= \frac{k(S-k)}{3}\left(k^2+(S-k)^2-2\right).
\end{aligned}
$$

Remark. When $k = S/2$ we find

$$v(k) = \frac{S^2}{12}\left(S^2/2-2\right)$$

and the standard deviation

$$\sigma(k) \simeq \frac{S^2}{\sqrt{24}} \simeq \frac{S^2}{5}.$$

if S is large.

In the graph of Figure S.3 the standard deviation of the game duration τ is plotted as a function of the initial state $k = 0, 1, \ldots, S$.

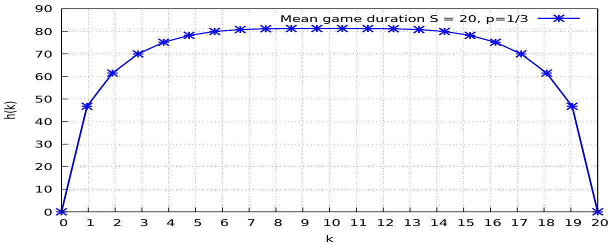


Fig. S.3: Standard deviation of τ as a function of $X_0 = k \in [0, 20]$.

f) We have $v(1) = 0$, which is consistent with the fact that the game duration is constant equal to one when $k = 1$ and $S = 2$.

Chapter 3 - Random Walks

Exercise 3.1

a) We find $\binom{4}{3} = \binom{4}{1} = 4$ paths, as follows.

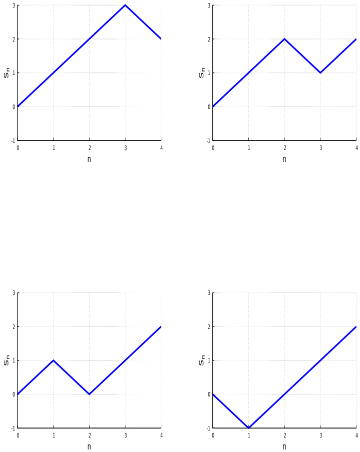


Fig. S.4: Four paths leading from 0 to 2 in four time steps.

b) In each of the

$$\binom{4}{3} = \binom{4}{1} = \frac{4!}{3!} = 4$$

paths there are 3 steps up (with probability p) and 1 step down (with probability $q = 1 - p$), hence the result.

c) We consider two cases depending on the parity of n and k.

i) In case n and k are even, or written as $n = 2n'$ and $k = 2k'$, Relation (3.3.3) shows that

$$\begin{aligned}
\mathbb{P}(S_n = k \mid S_0 = 0) &= \mathbb{P}(S_{2n'} = 2k' \mid S_0 = 0) \\
&= \binom{2n'}{n' + k'} p^{n'+k'} q^{n'-k'} \\
&= \binom{n}{(n+k)/2} p^{(n+k)/2} q^{(n-k)/2}, \qquad -n \leqslant k \leqslant n.
\end{aligned}$$

ii) In case n and k are odd, or written as $n = 2n' + 1$ and $k = 2k' + 1$, (3.3.4) shows that

$$\begin{aligned}
\mathbb{P}(S_n = k \mid S_0 = 0) &= \mathbb{P}(S_{2n'+1} = 2k' + 1 \mid S_0 = 0) \\
&= \binom{2n' + 1}{n' + k' + 1} p^{n'+k'+1} q^{n'-k'}
\end{aligned}$$

$$= \binom{n}{(n+k)/2} p^{(n+k)/2} q^{(n-k)/2}, \qquad -n \leqslant k \leqslant n.$$

d) By a first step analysis started at state ⓪ we have, letting $p_{n,k} := \mathbb{P}(S_n = k)$,

$$\begin{aligned} p_{n+1,k} &= \mathbb{P}(S_{n+1} = k) \\ &= \mathbb{P}(S_{n+1} = k \mid S_0 = 0) \\ &= \mathbb{P}(S_{n+1} = k \mid S_1 = 1)\mathbb{P}(S_1 = 1 \mid S_0 = 0) \\ &\quad + \mathbb{P}(S_{n+1} = k \mid S_1 = -1)\mathbb{P}(S_1 = -1 \mid S_0 = 0) \\ &= p\mathbb{P}(S_{n+1} = k \mid S_1 = 1) + q\mathbb{P}(S_{n+1} = k \mid S_1 = -1) \\ &= p\mathbb{P}(S_{n+1} = k-1 \mid S_1 = 0) + q\mathbb{P}(S_{n+1} = k+1 \mid S_1 = 0) \\ &= p\mathbb{P}(S_n = k-1 \mid S_0 = 0) + q\mathbb{P}(S_n = k+1 \mid S_0 = 0) \\ &= pp_{n,k-1} + qp_{n,k+1}, \end{aligned}$$

which yields

$$p_{n+1,k} = pp_{n,k-1} + qp_{n,k+1},$$

for all $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. The same equation could be obtained by a *backstep* analysis, by looking at the two possible ways to reach state k at time $n+1$ after starting from either $k-1$ or $k+1$ at time n.

e) We consider two cases depending on the parity of $n+1+k$.

i) If $n+1+k$ is odd the equation is clearly satisfied as both the right-hand side and left-hand side of (3.4.25) are equal to 0.

ii) If $n+1+k$ is even we have

$$\begin{aligned} pp_{n,k-1} + qp_{n,k+1} &= p\binom{n}{(n-1+k)/2} p^{(n-1+k)/2}(1-p)^{(n+1-k)/2} \\ &\quad + q\binom{n}{(n+1+k)/2} p^{(n+1+k)/2}(1-p)^{(n-1-k)/2} \\ &= \binom{n}{(n-1+k)/2} p^{(n+1+k)/2} q^{(n+1-k)/2} \\ &\quad + \binom{n}{(n+1+k)/2} p^{(n+1+k)/2} q^{(n+1-k)/2} \\ &= p^{(n+1+k)/2} q^{(n+1-k)/2} \left(\binom{n}{(n-1+k)/2} + \binom{n}{(n+1+k)/2} \right) \\ &= p^{(n+1+k)/2} q^{(n+1-k)/2} \left(\frac{n!}{((n+1-k)/2)!((n-1+k)/2)!} \right. \\ &\quad \left. + \frac{n!}{((n-1-k)/2)!((n+1+k)/2)!} \right) \\ &= p^{(n+1+k)/2} q^{(n+1-k)/2} \left(\frac{n!(n+1+k)/2}{((n+1-k)/2)!((n+1+k)/2)!} \right. \end{aligned}$$

$$
\begin{aligned}
&+\frac{n!(n+1-k)/2}{((n_1-k)/2)!((n+1+k)/2)!}\Bigg) \\
&= p^{(n+1+k)/2} q^{(n+1-k)/2} \frac{n!(n+1)}{((n+1-k)/2)!((n+1+k)/2)!} \\
&= p^{(n+1+k)/2} q^{(n+1-k)/2} \binom{n+1}{(n+1+k)/2} \\
&= p_{n+1,k},
\end{aligned}
$$

which shows that $p_{n,k}$ satisfies Equation (3.4.25). In addition we clearly have

$$
p_{0,0} = \mathbb{P}(S_0 = 0) = 1 \quad \text{and} \quad p_{0,k} = \mathbb{P}(S_0 = k) = 0, \qquad k \neq 0.
$$

Exercise 3.2

a) By the spatial homogeneity of the random walk, we have

$$
\mathbb{E}[T_{i-1} \mid S_0 = i] = \mathbb{E}[T_0 \mid S_0 = 1], \qquad i = 1, 2, \ldots, k,
$$

hence by adding inductively the average travel times from $\textcircled{i}$ to $\boxed{i-1}$ we find

$$
\mathbb{E}[T_0 \mid S_0 = k] = \sum_{i=1}^{k} \mathbb{E}[T_{i-1} \mid S_0 = i] = k\mathbb{E}[T_0 \mid S_0 = 1].
$$

Next, by first step analysis we have

$$
\begin{aligned}
\mathbb{E}[T_0 \mid S_0 = 1] &= q \times 1 + p \times (1 + \mathbb{E}[T_0 \mid S_0 = 2]) \\
&= q \times 1 + p \times (1 + 2\mathbb{E}[T_0 \mid S_0 = 1]) \\
&= 1 + 2p\mathbb{E}[T_0 \mid S_0 = 1],
\end{aligned}
$$

hence

$$
\mathbb{E}[T_0 \mid S_0 = 1] = \frac{1}{1-2p} = \frac{1}{q-p}, \qquad q > p.
$$

When $p \geqslant q$ we find $\mathbb{E}[T_0 \mid S_0 = 1] = +\infty$.

b) By the spatial homogeneity of the random walk, we have

$$
\mathbb{P}(T_{i-1} < \infty \mid S_0 = i) = \mathbb{P}(T_0 < \infty \mid S_0 = 1), \qquad i = 1, 2, \ldots, k,
$$

hence by multiplying inductively the hitting probabilities of state $\boxed{\text{i-1}}$ starting from $\textcircled{i}$ we find

$$
\mathbb{P}(T_0 < \infty \mid S_0 = k) = \prod_{i=1}^{k} \mathbb{P}(T_{i-1} < \infty \mid S_0 = i) = (\mathbb{P}(T_0 < \infty \mid S_0 = 1))^k .
$$

c) By first step analysis, we have

$$\mathbb{P}(T_0 < \infty \mid S_0 = 1) = q + p(\mathbb{P}(T_0 < \infty \mid S_0 = 1))^2,$$

which yields the equation

$$p\alpha^2 - \alpha + q = 0$$

for $\alpha := \mathbb{P}(T_0 < \infty \mid S_0 = 1)$, and possible solutions $\alpha \in \{1, q/p\}$. When $p > q$ we find

$$\mathbb{P}(T_0 < \infty \mid S_0 = 1) = \frac{q}{p},$$

when $p \leqslant q$ we have $\mathbb{P}(T_0 < \infty \mid S_0 = 1) = 1$. We conclude that in general,

$$\mathbb{P}(T_0 < \infty \mid S_0 = k) = \min\left(1, \left(\frac{q}{p}\right)^k\right), \qquad k \in \mathbb{N},$$

which coincides with (2.2.14) and (3.4.16).

Exercise 3.3 Recall that by (3.4.27) we have $G_{T_0^r}(s) = 1 - \sqrt{1-4pqs^2}$, $s \in (-1,1)$.

a) We have $\mathbb{P}(T_0^r = 0) = G_{T_0^r}(0) = 0$ and

$$\mathbb{P}(T_0^r < \infty) = G_{T_0^r}(1) = 1 - \sqrt{1-4pq} = 2\min(p,q),$$

by (3.4.14).

b) In general we have $\mathbb{P}(T_0^r = n) = \frac{1}{n!} G_{T_0^r}^{(n)}(0)$, $n \geqslant 0$, with

$$\begin{cases} G'_{T_0^r}(s) = 4pqs(1-4pqs^2)^{-1/2}, \\ \\ G''_{T_0^r}(s) = 4pq(1-4pqs^2)^{-3/2}, \\ \\ G'''_{T_0^r}(s) = 48p^2q^2s(1-4pqs^2)^{-5/2}, \\ \\ G^{(4)}_{T_0^r}(s) = 48p^2q^2(1+16pqs^2)(1-4pqs^2)^{-7/2}, \end{cases}$$

hence

$$\begin{cases} \mathbb{P}(T_0^r = 1) = G'_{T_0^r}(0) = 0, \\ \\ \mathbb{P}(T_0^r = 2) = G''_{T_0^r}(0) = 2pq, \\ \\ \mathbb{P}(T_0^r = 3) = \dfrac{1}{3!} G'''_{T_0^r}(0) = 0, \\ \\ \mathbb{P}(T_0^r = 4) = \dfrac{1}{4!} G^{(4)}_{T_0^r}(0) = 2p^2q^2. \end{cases}$$

c) We have

$$\mathbb{E}[T_0^r \mathbb{1}_{\{T_0^r < \infty\}}] = G'_{T_0^r}(1) = \frac{4pq}{\sqrt{1-4pq}} = \frac{4pq}{|p-q|},$$

hence by (1.6.6) in Lemma 1.10 we find

$$\mathbb{E}[T_0^r \mid T_0^r < \infty] = \frac{1}{\mathbb{P}(T_0^r < \infty)} \mathbb{E}[T_0^r \mathbb{1}_{\{T_0^r < \infty\}}] = 2\frac{\operatorname{Max}(p,q)}{|p-q|},$$

see (3.4.19).

Exercise 3.4

a) We have the boundary conditions $G_0(s) = 1$, $G_{\pm\infty}(s) = 0$, $s \in (-1,1)$, and

$$\begin{aligned} G_i(s) &= \mathbb{E}\big[s^{T_0} \,\big|\, X_0 = i\big] \\ &= p\mathbb{E}\big[s^{1+T_0} \,\big|\, X_0 = i+1\big] + q\mathbb{E}\big[s^{1+T_0} \,\big|\, X_0 = i-1\big] \\ &= ps\mathbb{E}\big[s^{T_0} \,\big|\, X_0 = i+1\big] + qs\mathbb{E}\big[s^{T_0} \,\big|\, X_0 = i-1\big], \qquad i \geqslant 1, \end{aligned}$$

hence we have

$$G_i(s) = psG_{i+1}(s) + qsG_{i-1}(s), \quad i \in \mathbb{Z} \setminus \{0\}, \quad s \in [0,1). \qquad \text{(S.24)}$$

b) We note that (S.24) has the same form as Equation (2.2.6), up to a parametrization by $s \in [0,1)$. Hence, substituting a possible solution of the form $G_i(s) = (\alpha(s))^i$ into (S.24) we find

$$(\alpha(s))^i = ps(\alpha(s))^{i+1} + qs(\alpha(s))^{i-1},$$

which yields the characteristic equation

$$ps(\alpha(s))^2 - \alpha(s) + qs = 0, \qquad s \in [0,1),$$

with unknown $\alpha(s)$ and solutions

$$\alpha_+(s) = \frac{1+\sqrt{1-4pqs^2}}{2ps} = \frac{2qs}{1-\sqrt{1-4pqs^2}} > \frac{1}{2ps}, \qquad |s| < 1/(2\sqrt{pq}),$$

and

$$\alpha_-(s) = \frac{1-\sqrt{1-4pqs^2}}{2ps} = \frac{2qs}{1+\sqrt{1-4pqs^2}} < 2qs, \qquad |s| < 1/(2\sqrt{pq}),$$

hence the general solution of (S.24) is given by

$$G_i(s) = C_+(\alpha_+(s))^i + C_-(\alpha_-(s))^i = C_+\left(\frac{1+\sqrt{1-4pqs^2}}{2ps}\right)^i + C_-\left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^i, \quad i \in \mathbb{Z}.$$

Hence the boundary conditions $G_{\pm\infty}(s) = 0$, or the boundedness condition $|G_i(s)| \leqslant 1$ for sufficiently small $s \in (-1, 1)$, show that $C_+ = 0$ and $C_- = 1$ when $i \geqslant 0$, *i.e.*

$$G_i(s) = \left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^i = \left(\frac{2qs}{1+\sqrt{1-4pqs^2}}\right)^i, \qquad i \geqslant 0, \quad \text{(S.25)}$$

and $C_+ = 1$ and $C_- = 0$ when $i \leqslant 0$, *i.e.*

$$G_i(s) = \left(\frac{1+\sqrt{1-4pqs^2}}{2ps}\right)^i = \left(\frac{2qs}{1-\sqrt{1-4pqs^2}}\right)^i, \qquad i \leqslant 0,$$

while by symmetry we also have

$$G_i(s) = \left(\frac{2qs}{1+\sqrt{1-4pqs^2}}\right)^i = \left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^i, \qquad i \geqslant 0, \quad \text{(S.26)}$$

and

$$G_i(s) = \left(\frac{2qs}{1-\sqrt{1-4pqs^2}}\right)^i = \left(\frac{1+\sqrt{1-4pqs^2}}{2ps}\right)^i, \qquad i \leqslant 0. \quad \text{(S.27)}$$

c) Relations (S.25) and (S.26) show from (3.4.14) that

$$\mathbb{P}(T_0 < \infty \mid X_0 = i) = G_i(1)$$

$$= \left(\frac{2q}{1+\sqrt{1-4pq}}\right)^i$$
$$= \left(\frac{2q}{1+|p-q|}\right)^i$$
$$= \min\left(1, \left(\frac{q}{p}\right)^i\right), \qquad i \geqslant 0,$$

which recovers (2.2.14) and (3.4.16).

d) When $p \geqslant q$ and $i \leqslant 0$, we have

$$G_i'(s) = \frac{is^{i-1}}{(2q)^{-i}}(1-\sqrt{1-4pqs^2})^{-i} - \frac{4pqi}{(2q)^{-i}}\frac{(1-\sqrt{1-4pqs^2})^{-i-1}}{\sqrt{1-4pqs^2}},$$

and when $s = 1$ this yields

$$\begin{aligned}
G_i'(1) &= \frac{i}{(2q)^{-i}}(1-\sqrt{1-4pq})^{-i} - \frac{4pqi}{(2q)^{-i}}\frac{(1-\sqrt{1-4pq})^{-i-1}}{\sqrt{1-4pq}} \\
&= \frac{i}{(2q)^{-i}}\,(1-(p-q))^{-i} - \frac{4pqi}{(2q)^{-i}}\frac{(1-(p-q))^{-i-1}}{p-q} \\
&= i - \frac{2pi}{p-q} \\
&= -\frac{i}{p-q}, \qquad i \leqslant 0.
\end{aligned}$$

e) By first step analysis starting from state ⓪ we have, when $q \geqslant p$,

$$\begin{aligned}
\mathbb{E}\big[s^{T_0^r} \mid X_0 = 0\big] &= ps\mathbb{E}\big[s^{T_0} \mid X_0 = 1\big] + qs\mathbb{E}\big[s^{T_0} \mid X_0 = -1\big] \\
&= psG_1(s) + qsG_{-1}(s) \\
&= \frac{1-\sqrt{1-4pqs^2}}{2} + \frac{2pqs^2}{1+\sqrt{1-4pqs^2}} \\
&= \frac{(1-\sqrt{1-4pqs^2})(1+\sqrt{1-4pqs^2}) + 4pqs^2}{2(1+\sqrt{1-4pqs^2})} \\
&= \frac{4pqs^2}{1+\sqrt{1-4pqs^2}} \\
&= 1-\sqrt{1-4pqs^2}, \qquad s \in [0,1),
\end{aligned}$$

which recovers (3.4.9), while when $q \leqslant p$ we find similarly

$$\mathbb{E}\big[s^{T_0^r} \mid X_0 = 0\big] = psG_1(s) + qsG_{-1}(s)$$

$$= \frac{2qps^2}{1+\sqrt{1-4pqs^2}} + \frac{1-\sqrt{1-4pqs^2}}{2}$$
$$= 1-\sqrt{1-4pqs^2}, \qquad s \in [0,1).$$

Exercise 3.5 By (3.4.21) we have

$$\begin{aligned}
\mathbb{E}\big[T_0^r \mathbb{1}_{\{T_0^r<\infty\}} \mid S_0=0\big] &= \sum_{n\geqslant 1} n\mathbb{P}\big(T_0^r = n \mid S_0=0\big) \\
&= \sum_{k\geqslant 1} 2k\mathbb{P}\big(T_0^r = 2k \mid S_0=0\big) \\
&= \sum_{k\geqslant 1} \frac{2k}{2k-1}\binom{2k}{k}(1/4)^k \qquad \text{(S.28)} \\
&= \infty,
\end{aligned}$$

where the general term of the above series satisfies the equivalence

$$\frac{2k}{2k-1}\binom{2k}{k}(1/4)^k \simeq_{k\to\infty} \frac{(2k)!}{4^k(k!)^2} \simeq_{k\to\infty} \sqrt{\frac{1}{2\pi k}}$$

by Stirling's approximation $k! \simeq (k/e)^k\sqrt{2\pi k}$ as k tends to ∞, from which we conclude to the divergence of the series (S.28).

Exercise 3.6

a) We have

$$\begin{aligned}
\mathbb{E}[M_n] &= \mathbb{E}\left[\sum_{k=1}^n 2^{k-1}X_k\right] \\
&= \sum_{k=1}^n 2^{k-1}\mathbb{E}[X_k] \\
&= (p-q)\sum_{k=0}^{n-1} 2^k \\
&= (p-q)\frac{1-2^n}{1-2} \\
&= (p-q)(2^n-1), \qquad n \geqslant 0.
\end{aligned}$$

b) The two possible values of $M_{\tau\wedge n}$ are 1 and

$$-\sum_{k=1}^n 2^{k-1} = -\frac{1-2^n}{1-2} = 1-2^n, \qquad n \geqslant 1.$$

We have

$$\mathbb{P}(M_{n\wedge\tau} = 1 - 2^n) = q^n \quad \text{and} \quad \mathbb{P}(M_{n\wedge\tau} = 1) = 1 - q^n, \qquad n \geqslant 1.$$

c) The stopped process $(M_{\min(n,\tau)})_{n\in\mathbb{N}}$ represents the wealth of a gambler whose strategy of the gambler is to double the stakes each time he loses, and to quit the game as soon as his gains reach \$1.
d) We have

$$\begin{aligned}\mathbb{E}[M_{n\wedge\tau}] &= (1-2^n)\mathbb{P}(M_{n\wedge\tau} = 1-2^n) + \mathbb{P}(M_{n\wedge\tau} = 1)\\ &= (1-2^n)q^n + 1 - q^n\\ &= 1 - (2q)^n, \qquad n \geqslant 1.\end{aligned}$$

In particular, we find $\mathbb{E}[M_{n\wedge\tau}] = 0$ in the fair game with $p = q = 1/2$.

Exercise 3.7

a) We have $\mathbb{P}(T^{(m)} < m) = 0$, $\mathbb{P}(T^{(m)} = m) = p^m$, $\mathbb{P}(T^{(m)} = m+1) = qp^m$, and $\mathbb{P}(T^{(m)} = m+2) = q^2p^m + pqp^m = qp^m$.
b) The idea is to start by start by flipping a coin and to observe the number k of consecutive "1" until we get the first "0".

   1) If $k = m$ then the game ends, and this happens with probability $\mathbb{P}(T^{(m)} = m) = p^m$.
   2) If $k < m$, the sequence of "1" is broken and we need to start again at time $k+1$. This happens with probability $p^k q$ and we need to factor in the power s^{k+1} where $k+1$ is the number of time steps until we reach the first "0", and restart the counter $T^{(m)}$.

   Therefore we have

$$\begin{aligned}G_{T^{(m)}}(s) &= s^m\mathbb{P}(T^{(m)} = m) + \sum_{k=0}^{m-1} qp^k\mathbb{E}\big[s^{k+1+T^{(m)}}\big]\\ &= p^m s^m + \sum_{k=0}^{m-1} p^k q s^{k+1}\mathbb{E}\big[s^{T^{(m)}}\big]\\ &= p^m s^m + \sum_{k=0}^{m-1} p^k q s^{k+1} G_{T^{(m)}}(s), \qquad s \in (-1,1).\end{aligned}$$

c) We have

$$G_{T^{(m)}}(s)\left(1 - \sum_{k=0}^{m-1} p^k q s^{k+1}\right) = p^m s^m,$$

and

$$\begin{aligned}G_{T^{(m)}}(s) &= \frac{p^m s^m}{1-\sum_{k=0}^{m-1} p^k q s^{k+1}} \\ &= \frac{p^m s^m}{1-\dfrac{qs(1-(ps)^m}{1-ps}} \\ &= \frac{p^m s^m(1-ps)}{1-ps-qs(1-(ps)^m)} \\ &= \frac{p^m s^m(1-ps)}{1-s+qp^m s^{m+1}}, \qquad s \in (-1,1].\end{aligned}$$

We note that

$$\mathbb{P}\big(T^{(m)} < \infty\big) = G_{T^{(m)}}(1) = (1-p)\frac{p^m}{qp^m} = 1.$$

d) Differentiating in (3.4.28) with respect to s, we find

$$G'_{T^{(m)}}(s) = \sum_{k=0}^{m-1} p^k q(k+1)s^k G_{T^{(m)}}(s) + \sum_{k=0}^{m-1} p^k q s^{k+1} G'_{T^{(m)}}(s) + mp^m s^{m-1},$$

$s \in (-1,1)$, hence, since $G_{T^{(m)}}(1) = 1$ because $\mathbb{P}\big(T^{(m)} < \infty\big) = 1$, we have

$$\begin{aligned}G'_{T^{(m)}}(1) &= (1-p)\sum_{k=0}^{m-1} p^k(k+1) + mp^m + G'_{T^{(m)}}(1)\sum_{k=0}^{m-1} p^k q \\ &= \frac{1-p^m}{1-p} + qG'_{T^{(m)}}(1)\sum_{k=0}^{m-1} p^k \\ &= \frac{1-p^m}{1-p} + qG'_{T^{(m)}}(1)\frac{1-p^m}{1-p} \\ &= \frac{1-p^m}{1-p} + G'_{T^{(m)}}(1)(1-p^m),\end{aligned}$$

which yields

$$\mathbb{E}\big[T^{(m)}\big] = G'_{T^{(m)}}(1) = \frac{1-p^m}{(1-p)p^m} = \frac{1-(1/p)^m}{p(1-1/p)} = \sum_{k=1}^{m} \frac{1}{p^k}.$$

For example, for an unbiased coin with $p = 1/2$ the mean time until the first winning streak of length m is

$$\mathbb{E}\big[T^{(m)}\big] = \sum_{k=1}^{m} \frac{1}{(1/2)^k} = \sum_{k=1}^{m} 2^k = 2\frac{1-2^m}{1-2} = 2(2^m-1).$$

Exercise 3.8

a) The sequence $(Z_n)_{n\geqslant 0}$ is a Markov chain since every new transition is determined by the current state, and its transition matrix P is given by

$$P = \begin{bmatrix} q & p & 0 & \cdots\cdots & 0 & 0 \\ q & 0 & p & \cdots\cdots & 0 & 0 \\ q & 0 & 0 & \cdots\cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ q & 0 & \cdots & \ddots & p & 0 & 0 \\ q & 0 & \cdots\cdots & & 0 & p & 0 \\ q & 0 & \cdots\cdots & & 0 & 0 & p \end{bmatrix},$$

b) By first step analysis, the mean hitting times $\mathbb{E}\big[T^{(m)} \mid Z_0 = l\big]$, $l = 0, 1, \ldots, m$, satisfy the equations

$$\begin{cases} \mathbb{E}[T_m \mid Z_0 = 0] = 1 + (1-p)\mathbb{E}[T_m \mid Z_0 = 0] + p\mathbb{E}[T_m \mid Z_0 = 1] \\ \mathbb{E}[T_m \mid Z_0 = 1] = 1 + (1-p)\mathbb{E}[T_m \mid Z_0 = 0] + p\mathbb{E}[T_m \mid Z_0 = 2] \\ \qquad \vdots \\ \mathbb{E}[T_m \mid Z_0 = m-1] = 1 + (1-p)\mathbb{E}[T_m \mid Z_0 = 0] + p\mathbb{E}[T_m \mid Z_0 = m] \\ \mathbb{E}[T_m \mid Z_0 = m] = 0, \end{cases}$$

i.e.

$$\begin{cases} \mathbb{E}[T_m \mid Z_0 = 0] = \dfrac{1}{p} + \mathbb{E}[T_m \mid Z_0 = 1] \\ p\mathbb{E}[T_m \mid Z_0 = 1] = p\mathbb{E}[T_m \mid Z_0 = 2] + \mathbb{E}[T_m \mid Z_0 = 0] - \mathbb{E}[T_m \mid Z_0 = 1] \\ \qquad \vdots \\ p\mathbb{E}[T_m \mid Z_0 = m-1] = p\mathbb{E}[T_m \mid Z_0 = m] \\ \qquad\qquad +\mathbb{E}[T_m \mid Z_0 = m-2] - \mathbb{E}[T_m \mid Z_0 = m-1] \\ \mathbb{E}[T_m \mid Z_0 = m] = 0, \end{cases}$$

or

$$\begin{cases} \mathbb{E}\big[T^{(m)} \mid Z_0 = 0\big] = \dfrac{1}{p} + \mathbb{E}\big[T^{(m)} \mid Z_0 = 1\big] \\ \mathbb{E}\big[T^{(m)} \mid Z_0 = 1\big] = \dfrac{1}{p} + \mathbb{E}\big[T^{(m)} \mid Z_0 = 2\big] \\ \qquad \vdots \\ \mathbb{E}\big[T^{(m)} \mid Z_0 = m-1\big] = \dfrac{1}{p} + \mathbb{E}\big[T^{(m)} \mid Z_0 = m\big], \\ \mathbb{E}\big[T^{(m)} \mid Z_0 = m\big] = 0, \end{cases}$$

with solution

$$\begin{aligned}\mathbb{E}\big[T^{(m)} \mid Z_0 = k\big] &= \sum_{l=k+1}^{m} \frac{1}{p^l} \\ &= \frac{1}{p^{k+1}} \sum_{l=0}^{m-k-1} \frac{1}{p^l} \\ &= \frac{1-(1/p)^{m-k}}{(1-1/p)p^{k+1}} \\ &= \frac{1-p^{m-k}}{(1-p)p^m}, \qquad k = 0, 1, \ldots, m.\end{aligned}$$

c) We have

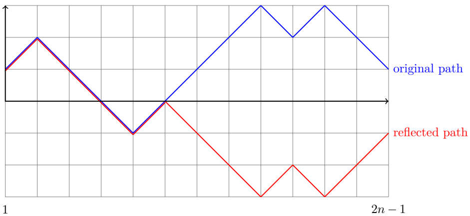
$$\begin{aligned}\mathbb{E}\big[T^{(m)}\big] &= \mathbb{E}\big[T^{(m)} \mid Z_0 = 0\big] \\ &= \sum_{l=1}^{m} \frac{1}{p^l} \\ &= \frac{1}{p} \sum_{l=0}^{m-1} \frac{1}{p^l} \\ &= \frac{1-p^m}{(1-p)p}.\end{aligned}$$

Exercise 3.9

a) Applying (3.4.29) with $m = n-1$ and $k = 0$, we find that this number is $\binom{2n-2}{n-1}$.

b) Applying (3.4.29) with $m = n-1$ and $k = -1$, we find that this number is $\binom{2n-2}{n-2} = \binom{2n-2}{n}$.

c) On the graph below, the blue path joining $S_1 = 1$ to $S_{2n-1} = 1$ by crossing ⓪ is associated to a unique red path joining $S_1 = 1$ to $S_{2n-1} = -1$.

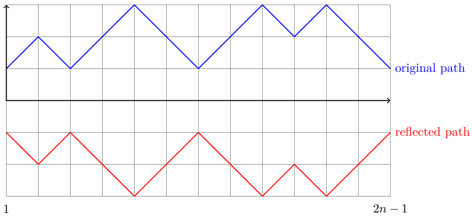


d) This number is $\binom{2n-2}{n-2} = \binom{2n-2}{n}$.

e) This number is

$$\binom{2n-2}{n-1} - \binom{2n-2}{n-2} = \frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{(n-2)!n!}$$
$$= \frac{(n^2 - n(n-1))(2n-2)!}{n!n!}$$
$$= \frac{(2n-2)!}{(n-1)!n!}.$$

f) According to the following graph, to each of the $\frac{(2n-2)!}{(n-1)!n!}$ blue paths joining $S_1 = 1$ to $S_{2n-2} = 1$ *without* crossing ⓪ between time 1 and time $2n-1$ we can associate a red path joining $S_1 = -1$ to $S_{2n-2} = -1$ *without* crossing ⓪.



Adding the number of paths joining $S_1 = 1$ to $S_{2n-2} = 1$ *without* crossing ⓪ from time 1 to time $2n-1$ to the number of paths joining $S_1 = -1$ to $S_{2n-2} = -1$ *without* crossing ⓪ between time 1 and time $2n-1$, we get the total to the number of paths joining $S_0 = 0$ to $S_{2n} = 0$ *without* crossing ⓪ from time 0 to time $2n$, as follows:

$$2\frac{(2n-2)!}{(n-1)!n!} = \frac{2n(2n-2)!}{n!n!} = \frac{1}{2n-1}\binom{2n}{n}.$$

We note that the convolution equation (3.4.3) is also satisfied in the following form:

$$\frac{1}{2n-1}\binom{2n}{n} = \sum_{k=0}^{n-1} \frac{1}{2n-2k-1}\binom{2n-2k}{n-k}\binom{2k}{k}$$
$$= \sum_{k=0}^{n-1} \frac{1}{2k-1}\binom{2k}{k}\binom{2n-2k}{n-k}, \qquad n \geqslant 1,$$

see *e.g.* *here*. We note that the number of paths joining $S_0 = 0$ to $S_{2n+2} = 0$ *without* crossing ⓪, from time 0 to time $2n+2$ satisfies

$$\frac{1}{2n+1}\binom{2n+2}{n+1} = \frac{2}{n+1}\binom{2n}{n},$$

where $\frac{1}{n+1}\binom{2n}{n}$ is the *Catalan number* counting the number of Dyck lattice paths joining $S_0 = 0$ to $S_n = n$ without passing above the diagonal between time 0 and time n.

Exercise 3.10

a) By independence of the sequence $(X_k)_{1 \leqslant k \leqslant n}$, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] &= \prod_{k=1}^n \mathbb{E}[\mathrm{e}^{tX_k}] \\ &= (q + p\mathrm{e}^t)^n, \quad n \geqslant 0, \quad t \in \mathbb{R}.\end{aligned}$$

b) By the classical Markov or Chernoff bound argument, we have

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - p) \geqslant z\right) &= \mathbb{P}\left(\exp\left(t\sum_{k=1}^n X_k\right) \geqslant \mathrm{e}^{ntz+npt}\right) \\ &= \mathrm{e}^{-ntz-npt}\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] \\ &= \mathrm{e}^{-ntz-npt}(q + p\mathrm{e}^t)^n \\ &= \mathrm{e}^{-n(t(p+z)-\log(q+p\mathrm{e}^t))}, \qquad t > 0.\end{aligned}$$

c) By differentiating $t \mapsto xt - \log(q + p\mathrm{e}^t)$ with respect to $t > 0$, we find that the maximizing value $t(x)$ is given by

$$t(x) = \log\frac{qx}{(1-x)p}, \qquad x \in (0,1).$$

d) We have

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - p) \geqslant z\right) \leqslant \mathrm{e}^{-n((p+z)t(x)-\log(q+p\mathrm{e}^{t(x)}))} \\ &= \exp\left(-n\left((p+z)\log\frac{(p+z)q}{(q-z)p} - \log\frac{q}{q-z}\right)\right), \qquad 0 \leqslant z < q.\end{aligned}$$

e) Applying Taylor's formula with remainder

$$f(t) = f(0) + tf'(0) + \frac{t^2}{2} f''(\theta t)$$

to the function $f(t) := \log(q + pe^t)$ with $f(0) = 0$, $f'(t) = pe^t / (q + pe^t)$, and $f''(t) = pqe^t / (q + pe^t)^2$, hence $f'(0) = p$ and

$$f''(\theta t) = \frac{pqe^{\theta t}}{(q + pe^{\theta t})^2} \leqslant \frac{1}{4},$$

we obtain

$$\log(q + pe^t) = pt + \frac{t^2}{2} f''(\theta t) \leqslant pt + \frac{t^2}{8}, \qquad t \in \mathbb{R}.$$

The inequality $4pqe^{\theta t} \leqslant (q + pe^{\theta t})^2$ can be proved by noting that it is equivalent to $(q - pe^{\theta t})^2 \geqslant 0$.

f) By differentiating $t \mapsto zt - t^2/8$ with respect to $t > 0$ we find that the maximizing value $t(z)$ is given by $t(z) = 4z$, $z \in (0, 1)$.

g) We have

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - p) \geqslant z\right) &\leqslant e^{-n(t(p+z) - \log(q+pe^t))} \\ &\leqslant e^{-n(zt(z) - t(z)^2/8)} \\ &\leqslant e^{-2nz^2}, \qquad z \geqslant 0.\end{aligned}$$

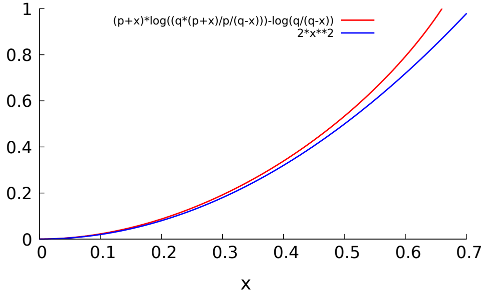


Fig. S.5: Comparison of rate functions.

Problem 3.11

a) We have

$$\mathbb{E}\left[\exp\left(\alpha\sum_{l=1}^n f(X_l)\right)\right] = \prod_{l=1}^n \mathbb{E}\left[e^{\alpha f(l)}\right]$$
$$= \left(\mathbb{E}\left[e^{\alpha f(l)}\right]\right)^n$$
$$= (\lambda_0(\alpha))^n, \quad n \geqslant 1.$$

b) For any $\alpha \in \mathbb{R}$ and $\gamma > 0$, we have

$$e^{\alpha\gamma n}\mathbb{P}\left(\sum_{l=1}^n f(X_l) \geqslant n\gamma\right) = e^{\alpha\gamma n}\mathbb{E}\left[\mathbf{1}_{\left\{\sum_{l=1}^n f(X_l)\geqslant n\gamma\right\}}\right]$$
$$\leqslant \mathbb{E}\left[\exp\left(\alpha\sum_{l=1}^n f(X_l)\right)\right]$$
$$= e^{-\alpha\gamma n}(\lambda_0(\alpha))^n$$
$$= e^{-n(\alpha\gamma - \log\lambda_0(\alpha))}, \qquad n \geqslant 1,$$

hence

$$\mathbb{P}\left(\sum_{l=1}^n f(X_l) \geqslant n\gamma\right) = e^{-n(\alpha\gamma - \log\lambda_0(\alpha))}, \qquad n \geqslant 1. \tag{S.29}$$

c) Since

$$\sum_{l=1}^d \pi_l f(l) = \mathbb{E}[f(X_1)] = 0,$$

we have

$$\lambda_0(\alpha) = \sum_{l=1}^d \pi_l e^{\alpha f(l)}$$
$$= \sum_{l=1}^d \pi_l + \alpha\sum_{l=1}^d \pi_l f(l) + \sum_{l=1}^d \pi_l(e^{\alpha f(l)} - \alpha f(l) - 1)$$
$$= 1 + \sum_{l=1}^d \pi_l(e^{\alpha f(l)} - \alpha f(l) - 1), \qquad \alpha \geqslant 1.$$

d) We have

$$\lambda_0(\alpha) = 1 + \sum_{l=1}^d \pi_l(e^{\alpha f(l)} - \alpha f(l) - 1),$$
$$= 1 + \sum_{k=2}^\infty \sum_{l=1}^d \pi_l \frac{(\alpha f(l))^n}{n!}$$

$$\begin{aligned}
&\leqslant 1 + \sum_{k=2}^{\infty} \sum_{l=1}^{d} \pi_l \alpha^n \\
&= 1 + \sum_{k=2}^{\infty} \alpha^n \\
&= 1 + \frac{\alpha^2}{1-\alpha}, \qquad \alpha \in [0,1).
\end{aligned}$$

e) By (S.29) and Question (d), for any $\alpha \in [0,1)$ and $\gamma > 0$ we have

$$\mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} f(X_l) \geqslant \gamma \right) \leqslant e^{-n(\alpha\gamma - \frac{\alpha^2}{1-\alpha})}, \qquad n \geqslant 1.$$

f) The value of $\alpha \in [0,1)$ which maximizes $\alpha\gamma - \alpha^2/(1-\alpha)$ satisfies

$$\gamma - 2\frac{\alpha}{1-\alpha} - \frac{\alpha^2}{(1-\alpha)^2} = 0$$

i.e.

$$\alpha = \frac{\gamma}{\gamma + 1 + \sqrt{\gamma+1}} < 1$$

and

$$1 - \alpha = \frac{1 + \sqrt{\gamma+1}}{\gamma + 1 + \sqrt{\gamma+1}}.$$

g) We have

$$\begin{aligned}
\alpha\gamma - \frac{\alpha^2}{1-\alpha} &= \frac{\gamma^2}{\gamma + 1 + \sqrt{\gamma+1}} - \frac{\gamma^2}{\gamma + 1 + \sqrt{\gamma+1}(1+\sqrt{\gamma+1})} \\
&= \frac{\gamma^2\sqrt{\gamma+1}}{(\gamma + 1 + \sqrt{\gamma+1})(1+\sqrt{\gamma+1})} \\
&\geqslant \frac{\gamma^2}{6},
\end{aligned}$$

hence for all $\gamma \in [0,1)$ and $n \geqslant 0$ we have

$$\mathbb{P}\left(\frac{1}{n} \sum_{l=1}^{n} f(X_l) \geqslant \gamma \right) \leqslant e^{-n\gamma^2/6}.$$

We note that this bound is better than the upper bound $e^{-(1-\lambda_1)n\gamma^2/12}$ where λ_1 is the second largest eigenvalue of P, since $0 \leqslant 1 - \lambda_1 \leqslant 2$.

Problem 3.12

a) For all $i = 1, \ldots, d$, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right|\right] &= \frac{1}{n}\sum_{j=1}^d \mathbb{E}\left[\left|\sum_{k=1}^n (\mathbf{1}_{\{X_k=j\}} - \pi_j)\right|\right] \\
&\leqslant \frac{1}{n}\sum_{j=1}^d \sqrt{\mathbb{E}\left[\left|\sum_{k=1}^n (\mathbf{1}_{\{X_k=j\}} - \pi_j)\right|^2\right]} \\
&= \frac{1}{n}\sum_{j=1}^d \sqrt{\mathbb{E}\left[\sum_{k=1}^n \left|\mathbf{1}_{\{X_k=j\}} - \pi_j\right|^2\right]} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^d \sqrt{\mathbb{E}[\left|\mathbf{1}_{\{X_k=j\}} - \pi_j\right|^2]} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^d \sqrt{\pi_j(1-\pi_j)} \\
&\leqslant \frac{1}{\sqrt{n}}\sum_{j=1}^d \sqrt{\pi_j} \\
&\leqslant \frac{\sqrt{d}}{\sqrt{n}}\sqrt{\sum_{j=1}^d \pi_j} \\
&= \sqrt{\frac{d}{n}}.
\end{aligned}$$

b) We have

$$\begin{aligned}
&\operatorname{Sup}_{y\in\mathbb{R}} |f(x_1,\ldots,x_n) - f(x_1,\ldots,x_{i-1},y,x_{i+1},\ldots,x_n)| \\
&\quad = \operatorname{Sup}_{y\in\mathbb{R}} \sum_{j=1}^d \left|\frac{1}{n}(\mathbf{1}_{\{x_i=j\}} - \mathbf{1}_{\{y=j\}})\right| \\
&\quad \leqslant \operatorname{Sup}_{y\in\mathbb{R}} \sum_{j=1}^d \frac{1}{n}\left|\mathbf{1}_{\{x_i=j\}} + \mathbf{1}_{\{y=j\}}\right| \\
&\quad \leqslant \frac{2}{n},
\end{aligned}$$

$x_1,\ldots,x_n \in \mathbb{R}$, $i=1,\ldots,n$.

c) For all $i=1,\ldots,d$ we have

$$\mathbb{P}\left(\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right| > \varepsilon\right)$$

$$= \mathbb{P}\left(\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right| - \mathbb{E}\left[\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right|\right]\right.$$
$$\left. > \varepsilon - \mathbb{E}\left[\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right|\right]\right)$$
$$\leqslant \mathbb{P}\left(\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right| - \mathbb{E}\left[\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right|\right] > \varepsilon - \sqrt{\frac{d}{n}}\right)$$
$$\leqslant \exp\left(-\frac{n}{2}\left(\varepsilon - \sqrt{\frac{d}{n}}\right)^2\right),$$

provided that $\varepsilon - \sqrt{d/n} > 0$, which implies

$$\mathbb{P}\left(\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} - \pi_j\right| > \varepsilon\right) \leqslant \exp\left(-\frac{n}{2}\operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}}\right)^2\right).$$

d) When $n \geqslant 4d/\varepsilon^2$, *i.e.* $\varepsilon \geqslant 2\sqrt{d/n}$, we have

$$\mathbb{P}\left(\sum_{j=1}^d \left|\widetilde{\pi}_j(n) - \pi_j\right| > \varepsilon\right) \leqslant \exp\left(-\frac{n}{2}\operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}}\right)^2\right)$$
$$= e^{-n\varepsilon^2/8}.$$

e) Setting $n > -8(\log\delta)/\varepsilon^2$, we have

$$\mathbb{P}\left(\sum_{j=1}^d \left|\widetilde{\pi}_j(n) - \pi_j\right| > \varepsilon\right) \leqslant e^{-n\varepsilon^2/8} < \delta,$$

which allows us to conclude by taking $c = 8$.

Problem 3.13

a) If none of the stated conditions, hold, *i.e.* if

$$\widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log n}{T_{n-1}^{(N,\alpha^*)}}} > p_N, \quad \widehat{m}_{n-1}^{(i,\alpha^*)} \leqslant p_i + \sqrt{\frac{2\log n}{T_{n-1}^{(i,\alpha^*)}}}, \quad T_{n-1}^{(i,\alpha^*)} \geqslant \frac{2\log n}{(p_N - p_i)^2},$$

then we have

$$\widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log n}{T_{n-1}^{(N,\alpha^*)}}} > p_N$$

$$= p_i + p_N - p_i$$
$$\geqslant p_i + \sqrt{\frac{2\log n}{T_{n-1}^{(i,\alpha^*)}}}$$
$$\geqslant \widehat{m}_{n-1}^{(i,\alpha^*)},$$

which implies $\alpha_n^* \neq i$.

b) We have

$$\mathbb{E}[T_n^{(i,\alpha^*)}] = \mathbb{E}\left[\sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}}\right]$$
$$\leqslant \mathbb{E}\left[\sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)} < \frac{2\log n}{(p_N-p_i)^2}\}}\right] + \mathbb{E}\left[\sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)} \geqslant \frac{2\log n}{(p_N-p_i)^2}\}}\right]$$
$$\leqslant \widehat{n}_i + \mathbb{E}\left[\sum_{\widehat{n}_i<k\leqslant n} \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)} \geqslant \frac{2\log n}{(p_N-p_i)^2}\}}\right]$$
$$\leqslant \widehat{n}_i + \sum_{\widehat{n}_i<k\leqslant n} \mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant p_N\right)$$
$$+ \sum_{\widehat{n}_i<k\leqslant n} \mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} > p_i + \sqrt{\frac{2\log k}{T_{k-1}^{(i,\alpha^*)}}}\right).$$

c) We have

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{2\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant p_N\right)$$
$$\leqslant \mathbb{P}\left(\exists l \in \{1,\ldots,k\} \ : \ \frac{1}{l}\sum_{j=1}^l (X_j^{(N,\alpha^*)} - p_N) + \sqrt{\frac{2\log k}{l}} \leqslant p_N\right)$$
$$\leqslant \sum_{l=1}^k \mathbb{P}\left(\frac{1}{l}\sum_{j=1}^l (X_j^{(N,\alpha^*)} - p_N) + \sqrt{\frac{2\log k}{l}} \leqslant p_N\right)$$
$$\leqslant \sum_{l=1}^k \mathbb{P}\left(\frac{1}{l}\sum_{j=1}^l (1 - X_j^{(N,\alpha^*)} - (1-p_N)) \geqslant \sqrt{\frac{2\log k}{l}}\right)$$
$$\leqslant \sum_{l=1}^k e^{-4\log k} = \sum_{l=1}^k \frac{1}{k^4} = \frac{1}{k^3}.$$

The argument is similar for

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > p_i + \sqrt{\frac{2\log k}{T_{k-1}^{(i,\alpha^*)}}}\right) \leqslant \frac{1}{k^3}, \quad i = 1, \ldots, N, \quad k > N.$$

d) We have

$$\begin{aligned}
\mathbb{E}[T_n^{(i)}] &\leqslant \widehat{n}_i + \sum_{k=1}^{n} \frac{2}{k^3} \\
&= \frac{8\log n}{(p_N - p_i)^2} + \sum_{k=1}^{n} \frac{2}{k^3} \\
&\leqslant \frac{8\log n}{(p_N - p_i)^2} + \int_1^n \frac{2}{t^3} dt \\
&\leqslant \frac{8\log n}{(p_N - p_i)^2} + \left(1 - \frac{1}{n^2}\right),
\end{aligned}$$

hence

$$\begin{aligned}
\overline{\mathcal{R}}n_n^{\alpha^*} &= np_N - \mathbb{E}\left[\sum_{k=1}^{n} p_{\alpha_k^*}\right] \\
&= \sum_{k=1}^{n} \mathbb{E}[p_N - p_{\alpha_k^*}] \\
&= np_N - \sum_{i=1}^{N} p_i \mathbb{E}[T_n^{i,\alpha^*}] \\
&= \sum_{i=1}^{N} (p_N - p_i) \mathbb{E}[T_n^{i,\alpha^*}] \\
&\leqslant 8 \sum_{i=1}^{N-1} \frac{\log n}{p_N - p_i} + \sum_{i=1}^{N-1} (p_N - p_i).
\end{aligned}$$

Problem 3.14

a) i) By first step analysis, the probability generating function

$$G_i(s) := \mathbb{E}\left[s^{T_{0,L}} \mid S_0 = i\right], \qquad s \in [-1, 1],$$

of $T_{0,L}$ satisfies the equation

$$G_i(s) = psG_{i+1}(s) + qsG_{i-1}(s), \qquad i = 1, \ldots, L-1,$$

with the boundary conditions $G_0(s) = G_L(s) = 1$. This equation can be solved as

$$G_i(s) = C_+(s)\left(\frac{1+\sqrt{1-4pqs^2}}{2ps}\right)^i + C_-(s)\left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^i,$$

$i = 0, \ldots, L$, where

$$\begin{cases} C_+(s) := \dfrac{(2ps)^L - (1-\sqrt{1-4pqs^2})^L}{(1+\sqrt{1-4pqs^2})^L - (1-\sqrt{1-4pqs^2})^L} \\ \\ C_-(s) := \dfrac{(1-\sqrt{1+4pqs^2})^L - (2ps)^L}{(1+\sqrt{1-4pqs^2})^L - (1-\sqrt{1-4pqs^2})^L}. \end{cases}$$

ii) The Laplace transform

$$L_i(\lambda) := \mathbb{E}\big[\mathrm{e}^{-\lambda T_{0,L}} \mid S_0 = i\big], \qquad i = 0,1,\ldots,L, \quad \lambda \geqslant 0.$$

of $T_{0,L}$ is then evaluated as

$$\begin{aligned} L_i(\lambda) &= G_i(\mathrm{e}^{-\lambda}) \\ &= C_+(\mathrm{e}^{-\lambda})\left(\frac{1+\sqrt{1-4pq\mathrm{e}^{-2\lambda}}}{2p\mathrm{e}^{-\lambda}}\right)^i + C_-(\mathrm{e}^{-\lambda})\left(\frac{1-\sqrt{1-4pq\mathrm{e}^{-2\lambda}}}{2p\mathrm{e}^{-\lambda}}\right)^i, \end{aligned}$$

$i = 0, \ldots, L$.

b) i) When $\mu = 0$, taking the limit as ε tends to zero yields the Laplace transform

$$L_x(\lambda) := \frac{\sinh(x\sqrt{2\lambda}) + \sinh((y-x)\sqrt{2\lambda})}{\sinh(y\sqrt{2\lambda})},$$

$x \in [0, y]$, $\lambda \geqslant 0$, of the first hitting time of the boundary $\{0, y\}$ by a standard Brownian motion $(B_t)_{t\in\mathbb{R}_+}$ started at $x \in [0, y]$, which recovers Equation (3) in Antal and Redner (2005), see also Equation (2.2.10) in Redner (2001), Theorem 1 in Williams (1992), and Relation (2.12) in Borodin (2017).

ii) When $\mu \neq 0$, we find the Laplace transform

$$\begin{aligned} L_x(\lambda) &= C_1(\lambda)\mathrm{e}^{\mu+\sqrt{2\lambda+\mu^2}} + C_2(\lambda)\mathrm{e}^{\mu-\sqrt{2\lambda+\mu^2}} \\ &= \frac{\mathrm{e}^{(x-y)\mu}\sinh(x\sqrt{2\lambda+\mu^2}) + \mathrm{e}^{\mu x}\sinh((y-x)\sqrt{2\lambda+\mu^2})}{\sinh(y\sqrt{2\lambda+\mu^2})}, \end{aligned}$$

$x \in [0, y]$, $\lambda \geqslant 0$, of the first hitting time of the boundary $\{0, y\}$ by a Brownian motion $(B_t + \mu t)_{t\in\mathbb{R}_+}$ with drift $\mu \in \mathbb{R}$ and started at

$x \in [0, y]$, which recovers Equation (3), where

$$\begin{cases} C_1(s) := \dfrac{1 - \mathrm{e}^{(\mu-\sqrt{2\lambda+\mu^2})y}}{\mathrm{e}^{(\mu+\sqrt{2\lambda+\mu^2})y} - \mathrm{e}^{(\mu-\sqrt{2\lambda+\mu^2})y}} \\ \\ C_2(s) := \dfrac{\mathrm{e}^{(\mu+\sqrt{2\lambda+\mu^2})y}}{\mathrm{e}^{(\mu+\sqrt{2\lambda+\mu^2})y} - \mathrm{e}^{(\mu-\sqrt{2\lambda+\mu^2})y}}, \end{cases}$$

see Theorem 1 in Williams (1992) in the case $x = 0$, by taking $\alpha = 0$ and $C = -1$ therein.

c) i) By first step analysis, the probability generating function

$$G_i(s) := \mathbb{E}\big[s^{T_{0,L}} \mid S_0 = i\big], \qquad s \in [-1, 1],$$

of $T_{0,L}$ satisfies the same equation

$$G_i(s) = psG_{i+1}(s) + qsG_{i-1}(s), \qquad i = 1, \ldots, L-1,$$

as above. However, the boundary conditions are modified into $G_0(s) = psG_1(s) + qsG_0(s)$, with $G_L(s) = 1$. The finite difference equation can now be solved as

$$G_i(s) = C_+(s)\left(\frac{1+\sqrt{1-4pqs^2}}{2ps}\right)^i + C_-(s)\left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^i,$$

$i = 0, \ldots, L$, where

$$\begin{cases} C_+(s) := \dfrac{ps\alpha_-(s) + qs - 1}{(1-qs)(\alpha_-^L(s) - \alpha_+^L(s)) - ps(\alpha_+(s)\alpha_-(s)^L - \alpha_+^L(s)\alpha_-(s))} \\ \\ C_-(s) := \dfrac{ps\alpha_+(s) + qs - 1}{(qs-1)(\alpha_-^L(s) - \alpha_+^L(s)) + ps(\alpha_+(s)\alpha_-(s)^L - \alpha_+^L(s)\alpha_-(s))} \end{cases}$$

and

$$\alpha_+(s) = \frac{1+\sqrt{1-4pqs^2}}{2ps}, \quad \alpha_-(s) = \frac{1-\sqrt{1-4pqs^2}}{2ps}.$$

ii) The Laplace transform is then evaluated as

$$\begin{aligned} L_i(\lambda) &= G_i(\mathrm{e}^{-\lambda}) \\ &= C_+(\mathrm{e}^{-\lambda})\left(\frac{1+\sqrt{1-4pq\mathrm{e}^{-2\lambda}}}{2p\mathrm{e}^{-\lambda}}\right)^i + C_-(\mathrm{e}^{-\lambda})\left(\frac{1-\sqrt{1-4pq\mathrm{e}^{-2\lambda}}}{2p\mathrm{e}^{-\lambda}}\right)^i, \end{aligned}$$

$i = 0, \ldots, L$.

i) When $\mu = 0$, taking the limit as ε tends to zero yields the Laplace transform

$$L_x(\lambda) := \frac{\cosh(x\sqrt{2\lambda})}{\cosh(y\sqrt{\lambda})}, \qquad x \in [0, y], \quad \lambda \geqslant 0,$$

of the first hitting time of the boundary $\{y\}$ by a standard Brownian motion reflected at 0, which recovers Equation (5) in Antal and Redner (2005), see also Equation (2.2.21) in Redner (2001).*

ii) When $\mu \neq 0$ we find the Laplace transform

$$L_x(\lambda) := \mathrm{e}^{(x-y)\mu} \frac{\mu \sinh(x\sqrt{2\lambda + \mu^2}) - \sqrt{2\lambda + \mu^2} \cosh(x\sqrt{2\lambda + \mu^2})}{\mu \sinh(y\sqrt{2\lambda + \mu^2}) - \sqrt{2\lambda + \mu^2} \cosh(y\sqrt{2\lambda + \mu^2})},$$

$x \in [0, y]$, $\lambda \geqslant 0$, of the first hitting time of the boundary $\{y\}$ by a Brownian motion $(B_t + \mu t)_{t \in \mathbb{R}_+}$ with drift $\mu \in \mathbb{R}$ reflected at 0 and started at $x \in [0, y]$.

Problem 3.15

a) By first step analysis, we have

$$H_i(s) = psH_{i+1}(s) + qsH_{i-1}(s), \qquad -1 \leqslant s \leqslant 1, \quad i \leqslant -2, \ i \geqslant 2,$$

and

$$H_1(s) = psH_2(s) + qs(1 + H_0(s)), \quad H_{-1}(s) = psH_{-2}(s) + qs(1 + H_0(s)),$$

and

$$H_0(s) = psH_1(s) + qsH_{-1}(s), \qquad -1 \leqslant s \leqslant 1.$$

b) Letting

$$H_i(s) := \begin{cases} \dfrac{1}{\sqrt{1-4pqs^2}} \left(\dfrac{1-\sqrt{1-4pqs^2}}{2ps} \right)^i, & i \geqslant 1, \\[2ex] \dfrac{1-\sqrt{1-4pqs^2}}{\sqrt{1-4pqs^2}}, & i = 0, \\[2ex] \dfrac{1}{\sqrt{1-4pqs^2}} \left(\dfrac{1-\sqrt{1-4pqs^2}}{2qs} \right)^{-i}, & i \leqslant -1, \end{cases}$$

* Equation (2.2.21) in Redner (2001) is stated for a reflecting boundary at $x = L$ ("Reflection mode" page 48), however in Antal and Redner (2005) the reflecting boundary is at $x = 0$, and therefore (5) therein has to be corrected accordingly.

we check that

$$\begin{aligned}
& psH_{i+1}(s) + qsH_{i-1}(s) \\
&= \frac{ps}{\sqrt{1-4pqs^2}} \left(\frac{1-\sqrt{1-4pqs^2}}{2ps} \right)^{i+1} + \frac{qs}{\sqrt{1-4pqs^2}} \left(\frac{1-\sqrt{1-4pqs^2}}{2ps} \right)^{i-1} \\
&= \frac{1}{\sqrt{1-4pqs^2}} \left(\frac{1-\sqrt{1-4pqs^2}}{2ps} \right)^{i} \left(\frac{1-\sqrt{1-4pqs^2}}{2} + \frac{2pqs^2}{1-\sqrt{1-4pqs^2}} \right) \\
&= \frac{1}{\sqrt{1-4pqs^2}} \left(\frac{1-\sqrt{1-4pqs^2}}{2ps} \right)^{i}, \qquad i \geqslant 1.
\end{aligned}$$

c) We have

$$H_i(s) = (1 + H_0(s))G_i(s), \qquad i \in \mathbb{Z}, \quad -1 \leqslant s \leqslant 1.$$

d) As a direct consequence of the answers to Questions (b) and (c), we have

$$G_i(s) := \begin{cases} \left(\dfrac{1-\sqrt{1-4pqs^2}}{2ps} \right)^{i}, & i \geqslant 1, \\ 1 - \sqrt{1-4pqs^2}, & i = 0, \\ \left(\dfrac{1-\sqrt{1-4pqs^2}}{2qs} \right)^{-i}, & i \leqslant -1. \end{cases}$$

e) We find

$$\mathbb{P}(T_0 < \infty \mid S_0 = i) = G_i(1) = \begin{cases} \min\left(1, \left(\dfrac{q}{p}\right)^i\right), & i \neq 0, \\ 1 - |p-q|, & i = 0, \end{cases}$$

see (3.4.15) and (3.4.16).

f) Using the relations $\mathbb{E}[T_0^r \mid S_0 = i] = G_i'(1)$ when $\mathbb{P}(T_0^r \mid S_0 = i) = 1$, see (1.7.6), and $\mathbb{E}[T_0^r \mid S_0 = i] = +\infty$ when $\mathbb{P}(T_0^r \mid S_0 = i) < 1$, We find

$$\mathbb{E}[T_0^r \mid S_0 = i] = \begin{cases} \dfrac{i}{q-p}, & i \geqslant 1, \quad q > p, \\ +\infty, & i \geqslant 1, \quad q \leqslant p, \\ +\infty, & i = 0, \\ \dfrac{i}{q-p}, & i \leqslant -1, \quad p > q \\ +\infty, & i \leqslant -1, \quad p \leqslant q, \end{cases}$$

see (3.4.18).

Problem 3.16

a) We have

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} p^{n+k} q^{n-k}, \quad -n \leqslant k \leqslant n.$$

b) We partition the event $\{S_{2n} = 0\}$ into

$$\{S_{2n} = 0\} = \bigcup_{k=1}^{2n} \{S_1 \neq 0, \ldots, S_{2k-1} \neq 0, \ S_{2k} = 0\}, \qquad n \geqslant 1,$$

according to all possible times $2k = 2, 4, \ldots, 2n$ of *first* return to state ⓪ *before* time $2n$, see Figure S.6.

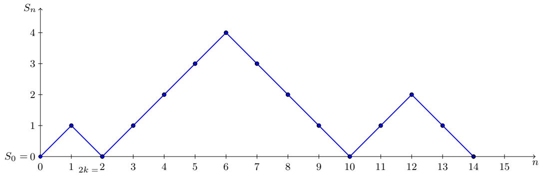


Fig. S.6: Last return to state 0 at time $k = 10$.

Then we have

$$\begin{aligned} \mathbb{P}(S_{2n} = 0) &= \sum_{r=1}^{n} \mathbb{P}\big(S_2 \neq 0, \ldots, S_{2r-1} \neq 0, \ S_{2r} = 0, \ S_{2n} = 0\big) \\ &= \sum_{r=1}^{n} \mathbb{P}\big(S_{2n} = 0 \mid S_{2r} = 0, \ S_{2r-1} \neq 0, \ldots, \ S_2 \neq 0\big) \end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{P}\big(S_2 \neq 0, \ldots, S_{2r-1} \neq 0,\ S_{2r} = 0\big) \\
&= \sum_{r=1}^{n} \mathbb{P}\big(S_{2n} = 0 \mid S_{2r} = 0\big)\mathbb{P}\big(T_0 = 2r\big) \\
&= \sum_{k=1}^{n} \mathbb{P}\big(S_{2n-2r} = 0\big)\mathbb{P}\big(T_0 = 2r\big), \qquad n \geqslant 1.
\end{aligned}
$$

c) The idea of the proof is to note that after starting from $S_0 = 0$, one may move up with probability $1/2$, in which case $T_0 = 2r$ time steps strictly above 0 will be counted from time 0 until time T_0, after which the remaining $2r - 2k$ time steps will be counted from time T_0 until time $2n$. On the other hand, if one moves down with probability $1/2$, zero time step strictly above 0 will be counted from time 0 until time $T_0 = 2r$, after which the remaining $2k$ time steps strictly above zero will be counted from time $T_0 = 2r$ until time $2n$. Hence we have

$$
\begin{aligned}
&\mathbb{P}(T_{2n}^{+} = 2k) = \sum_{r=1}^{n} \mathbb{P}\big(S_0 = 0, T_0 = 2r, T_{2n}^{+} = 2k\big) \\
&= \sum_{r=1}^{n} \mathbb{P}\big(S_0 = 0, S_1 = 1, T_0 = 2r, T_{2n}^{+} = 2k\big) \\
&\quad + \sum_{r=1}^{n} \mathbb{P}\big(S_0 = 0, S_1 = -1, T_0 = 2r, T_{2n}^{+} = 2k\big) \\
&= \sum_{r=1}^{k} \mathbb{P}\big(S_0 = 0, S_1 = 1, T_0 = 2r\big)\mathbb{P}\big(T_{2n}^{+} = 2k \mid S_1 = 1, T_0 = 2r\big) \\
&\quad + \sum_{r=1}^{n-k} \mathbb{P}\big(S_0 = 0, S_1 = -1, T_0 = 2r\big)\mathbb{P}\big(T_{2n}^{+} = 2k \mid S_1 = -1, T_0 = 2r\big) \\
&= \sum_{r=1}^{k} \mathbb{P}\big(S_0 = 0, S_1 = 1, T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^{+} = 2k - 2r\big) \\
&\quad + \sum_{r=1}^{n-k} \mathbb{P}\big(S_0 = 0, S_1 = -1, T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^{+} = 2k\big) \\
&= \frac{1}{2}\sum_{r=1}^{k} \mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^{+} = 2k - 2r\big) + \frac{1}{2}\sum_{r=1}^{n-k} \mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T_{2n-2r}^{+} = 2k\big),
\end{aligned}
$$

$n \geqslant 1$.

d) We check that, when

$$
\mathbb{P}\big(T_{2n-2r}^{+} = 2k - 2r\big) = 2^{-(2n-2r)}\binom{2k-2r}{k-r}\binom{2n-2k}{n-k}
$$

and

$$\mathbb{P}\big(T^+_{2n-2r} = 2k\big) = 2^{-(2n-2r)}\binom{2k}{k}\binom{2n-2r-2k}{n-r-k},$$

we have

$$\begin{aligned}
&\frac{1}{2}\sum_{r=1}^{k}\mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T^+_{2n-2r} = 2k-2r\big) + \frac{1}{2}\sum_{r=1}^{n-k}\mathbb{P}\big(T_0 = 2r\big)\mathbb{P}\big(T^+_{2n-2r} = 2k\big)\\
&\quad = \frac{1}{2}\sum_{r=1}^{k}\mathbb{P}\big(T_0 = 2r\big)2^{-2n+2r}\binom{2k-2r}{k-r}\binom{2n-2k}{n-k}\\
&\qquad + \frac{1}{2}\sum_{r=1}^{n-k}2^{-2n+2r}\mathbb{P}\big(T_0 = 2r\big)\binom{2k}{k}\binom{2n-2r-2k}{n-r-k}\\
&\quad = \frac{1}{2}2^{-2n}\binom{2n-2k}{n-k}2^{2k}\sum_{r=1}^{k}\mathbb{P}\big(T_0 = 2r\big)\frac{1}{2^{2(k-r)}}\binom{2k-2r}{k-r}\\
&\qquad + \frac{1}{2}2^{-2n}\binom{2k}{k}2^{2(n-k)}\sum_{r=1}^{n-k}\mathbb{P}\big(T_0 = 2r\big)\frac{1}{2^{2(n-k-r)}}\binom{2n-2r-2k}{n-r-k}\\
&\quad = \frac{1}{2}2^{-2(n-k)}\binom{2n-2k}{n-k}\sum_{r=1}^{k}\mathbb{P}\big(T_0 = 2r\big)\mathbb{P}(S_{2k-2r} = 0)\\
&\qquad + \frac{1}{2}\binom{2k}{k}2^{-2k}\sum_{r=1}^{n-k}\mathbb{P}\big(T_0 = 2r\big)\mathbb{P}(S_{2n-2k+2r} = 0)\\
&\quad = \frac{1}{2}2^{-2(n-k)}\binom{2n-2k}{n-k}\mathbb{P}(S_{2k} = 0) + \frac{1}{2}2^{-2k}\binom{2k}{k}\mathbb{P}(S_{2n-2k} = 0)\\
&\quad = \frac{1}{2}2^{-2n}\binom{2n-2k}{n-k}\binom{2k}{k} + \frac{1}{2}2^{-2n}\binom{2k}{k}\binom{2n-2k}{n-k}\\
&\quad = 2^{-2n}\binom{2k}{k}\binom{2n-2k}{n-k}\\
&\quad = \mathbb{P}\big(T^+_{2n} = 2k\big), \qquad n \geqslant 1.
\end{aligned}$$

e) We have

$$\begin{aligned}
\mathbb{P}\big(T^+_{2n} = 2k\big) &= 2^{-2n}\binom{2k}{k}\binom{2n-2k}{n-k}\\
&= 2^{-2n}\frac{(2k)!}{k!^2}\frac{(2n-2k)!}{(n-k)!^2}\\
&\simeq 2^{-2n}\frac{(2k/e)^{2k}\sqrt{4\pi k}}{(k/e)^{2k}2\pi k}\frac{((2n-2k)/e)^{(2n-2k)}\sqrt{2\pi(2n-2k)}}{((n-k)/e)^{(2n-2k)}2\pi(n-k)}
\end{aligned}$$

$$= \frac{1}{\pi\sqrt{k(n-k)}}, \qquad k, n-k \to \infty.$$

Next, we compute the limit

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\big(T_{2n}^{+}/2n \leqslant x\big) &= \lim_{n\to\infty} \sum_{k=0}^{nx} \mathbb{P}\big(T_{2n}^{+}/2n = k/n\big) \\
&= \lim_{n\to\infty} \sum_{0\leqslant k/n\leqslant x} 2^{-2n}\binom{2k}{k}\binom{2n-2k}{n-k} \\
&\simeq \frac{1}{\pi} \lim_{n\to\infty} \frac{1}{n} \sum_{0\leqslant k/n\leqslant x} \frac{1}{\sqrt{k(1-k/n)/n}} \\
&= \frac{1}{\pi}\int_0^x \frac{1}{\sqrt{t(1-t)}}dt \\
&= \frac{1}{2} + \frac{\arcsin(2x-1)}{\pi} \\
&= \frac{2}{\pi}\arcsin\sqrt{x}, \qquad x \in [0,1],
\end{aligned}$$

which yields the arcsine distribution.

Problem 3.17

a) Since the increment X_k takes its values in $\{-1, 1\}$, the set of distinct values in $\{S_0, S_1, \ldots, S_n\}$ is the integer interval

$$\left[\inf_{k=0,1,\ldots,n} S_k, \ \operatorname{Sup}_{k=0,1,\ldots,n} S_k\right],$$

which has

$$R_n = 1 + \left(\operatorname{Sup}_{k=0,1,\ldots,n} S_k\right) - \left(\inf_{k=0,1,\ldots,n} S_k\right)$$

elements. In addition we have $R_0 = 1$ and $R_1 = 2$.

b) At each time step $k \geqslant 1$ the range can only either increase by one unit or remain constant, hence $R_k - R_{k-1} \in \{0, 1\}$ is a Bernoulli random variable. In addition we have the identity

$$\{R_k - R_{k-1} = 1\} = \{S_k \neq S_0, \ S_k \neq S_1, \ldots, \ S_k \neq S_{k-1}\},$$

hence, applying the probability $\mathbb{P}$ to both sides, we get

$$\mathbb{P}(R_k - R_{k-1} = 1) = \mathbb{P}(S_k - S_0 \neq 0, \ S_k - S_1 \neq 0, \ldots, \ S_k - S_{k-1} \neq 0).$$

c) By the change of index

$$(X_1, X_2, \ldots, X_{k-1}, X_k) \longmapsto (X_k, X_{k-1}, \ldots, X_2, X_1)$$

under which $X_1 + X_2 + \cdots + X_l$ becomes $X_k + \cdots + X_{k-l+1}$, $l = 1, 2, \ldots, k$, we have

$$\begin{aligned}\mathbb{P}(R_k - R_{k-1} = 1) &= \mathbb{P}(S_k - S_0 \neq 0,\ S_k - S_1 \neq 0, \ldots,\ S_k - S_{k-1} \neq 0)\\ &= \mathbb{P}(X_1 + \cdots + X_k \neq 0,\ X_2 + \cdots + X_k \neq 0, \ldots,\ X_k \neq 0)\\ &= \mathbb{P}(X_1 \neq 0,\ X_1 + X_2 \neq 0, \ldots,\ X_1 + \cdots + X_k \neq 0),\end{aligned}$$

for all $k \geqslant 1$, since the sequence $(X_k)_{k\geqslant 1}$ is made of independent and identically distributed random variables.

d) We have the telescoping sum

$$\begin{aligned}R_n &= R_0 + (R_1 - R_0) + \cdots + (R_n - R_{n-1})\\ &= R_0 + \sum_{k=1}^{n} (R_k - R_{k-1}), \qquad n \geqslant 0.\end{aligned}$$

e) By (1.2.6) we have

$$\mathbb{P}(T_0 = \infty) = \mathbb{P}\left(\bigcap_{k\geqslant 1} \{T_0 > k\}\right) = \lim_{k\to\infty} \mathbb{P}(T_0 > k),$$

since

$$\{T_0 > k+1\} \Longrightarrow \{T_0 > k\}, \qquad k \geqslant 1,$$

i.e. $(\{T_0 > k\})_{k\geqslant 1}$ is a decreasing sequence of events.

f) Noting that $R_k - R_{k-1} \in \{0, 1\}$ is a Bernoulli random variable with

$$\mathbb{E}[R_k - R_{k-1}] = \mathbb{P}(R_k - R_{k-1} = 1),$$

we find that

$$\begin{aligned}\mathbb{E}[R_n] &= \mathbb{E}\left[R_0 + \sum_{k=1}^{n} (R_k - R_{k-1})\right] = R_0 + \sum_{k=1}^{n} \mathbb{E}[R_k - R_{k-1}]\\ &= R_0 + \sum_{k=1}^{n} \mathbb{P}(R_k - R_{k-1} = 1)\\ &= R_0 + \sum_{k=1}^{n} \mathbb{P}(X_1 \neq 0, X_1 + X_2 \neq 0, \ldots, X_1 + \cdots + X_k \neq 0)\\ &= R_0 + \sum_{k=1}^{n} \mathbb{P}(S_1 \neq 0, S_2 \neq 0, \ldots, S_k \neq 0) = R_0 + \sum_{k=1}^{n} \mathbb{P}(T_0 > k)\end{aligned}$$

$$= 1 + \sum_{k=1}^{n} \mathbb{P}(T_0 > k) = \mathbb{P}(T_0 > 0) + \sum_{k=1}^{n} \mathbb{P}(T_0 > k) = \sum_{k=0}^{n} \mathbb{P}(T_0 > k).$$

g) Let $\varepsilon > 0$. Since by Question (e) we have $\mathbb{P}(T_0 = \infty) = \lim_{k\to\infty} \mathbb{P}(T_0 > k)$, there exists $N \geqslant 1$ such that

$$|\mathbb{P}(T_0 = \infty) - \mathbb{P}(T_0 > k)| < \varepsilon, \qquad k \geqslant N.$$

Hence for $n \geqslant N$ we have

$$\left|\mathbb{P}(T_0 = \infty) - \frac{1}{n}\sum_{k=1}^{n} \mathbb{P}(T_0 > k)\right| = \left|\frac{1}{n}\sum_{k=1}^{n} (\mathbb{P}(T_0 = \infty) - \mathbb{P}(T_0 > k))\right|$$
$$\leqslant \frac{1}{n}\sum_{k=1}^{n} |\mathbb{P}(T_0 = \infty) - \mathbb{P}(T_0 > k)|$$
$$\leqslant \frac{1}{n}\sum_{k=1}^{N} |\mathbb{P}(T_0 = \infty) - \mathbb{P}(T_0 > k)| + \frac{1}{n}\sum_{k=N+1}^{n} |\mathbb{P}(T_0 = \infty) - \mathbb{P}(T_0 > k)|$$
$$\leqslant \frac{N}{n} + \frac{n-N}{n}\varepsilon$$
$$\leqslant \frac{N}{n} + \varepsilon.$$

Then, choosing $N_0 \geqslant 1$ such that $(N+1)/n \leqslant \varepsilon$ for $n \geqslant N_0$, we get

$$\left|\mathbb{P}(T_0 = \infty) - \frac{1}{n}\mathbb{E}[R_n]\right| \leqslant \frac{1}{n} + \left|\mathbb{P}(T_0 = \infty) - \frac{1}{n}\sum_{k=1}^{n} \mathbb{P}(T_0 > k)\right| \leqslant 2\varepsilon,$$

$n \geqslant N_0$, which concludes the proof.

Alternatively, the answer to that question can be derived by applying the Cesàro theorem, which states that in general we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} a_k = a$$

when the sequence $(a_k)_{k\in\mathbb{N}}$ has the limit a, by taking $a_k = \mathbb{P}(T_0 > k)$, $k \in \mathbb{N}$, since we have $\lim_{k\to\infty} \mathbb{P}(T_0 > k) = \mathbb{P}(T_0 = \infty)$.

h) From Relation (3.4.15) in Section 3.4 we have

$$\mathbb{P}(T_0 = +\infty) = |p - q|,$$

hence by the result of Question (g) we get

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[R_n] = |p-q|,$$

when $p \neq q$, and

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[R_n] = 0,$$

when $p = q = 1/2$.

Problem 3.18 The question of recurrence of the d-dimensional random walk has been first solved in Pólya (1921), cf. Champion et al. (2007) for the solution proposed in this problem.

a) We partition the event $\{S_n = \vec{0}\}$ into

$$\{S_n = \vec{0}\} = \bigcup_{k=2}^{n} \{S_{n-k} = \vec{0},\ S_{n-k+1} \neq \vec{0}, \ldots, S_{n-1} \neq \vec{0},\ S_n = \vec{0}\}, \quad n \geqslant 1,$$

according to the time of *last* return to state $\vec{0}$ *before* time n, with $\mathbb{P}(\{S_1 = \vec{0}\}) = 0$ since we are starting from $S_0 = \vec{0}$. Then we have

$$\begin{aligned}
&\mathbb{P}(S_n = \vec{0}) := \mathbb{P}(S_n = \vec{0} \mid S_0 = \vec{0}) \\
&= \sum_{k=2}^{n} \mathbb{P}(S_{n-k} = \vec{0},\ S_{n-k+1} \neq \vec{0}, \ldots, S_{n-1} \neq \vec{0},\ S_n = \vec{0} \mid S_0 = \vec{0}) \\
&= \sum_{k=2}^{n} \mathbb{P}(S_{n-k+1} \neq \vec{0}, \ldots, S_{n-1} \neq \vec{0},\ S_n = \vec{0} \mid S_{n-k} = \vec{0},\ S_0 = \vec{0}) \\
&\qquad\qquad\qquad\qquad \times \mathbb{P}(S_{n-k} = \vec{0} \mid S_0 = \vec{0}) \\
&= \sum_{k=2}^{n} \mathbb{P}(S_1 \neq \vec{0}, \ldots, S_{k-1} \neq \vec{0},\ S_k = \vec{0} \mid S_0 = \vec{0}) \mathbb{P}(S_{n-k} = \vec{0} \mid S_0 = \vec{0}) \\
&= \sum_{k=2}^{n} \mathbb{P}(T^r_{\vec{0}} = k \mid S_0 = \vec{0}) \mathbb{P}(S_{n-k} = \vec{0} \mid S_0 = \vec{0}) \\
&= \sum_{k=2}^{n} \mathbb{P}(S_{n-k} = \vec{0}) \mathbb{P}(T^r_{\vec{0}} = k), \qquad n \geqslant 1.
\end{aligned}$$

b) We have

$$\begin{aligned}
\sum_{n=1}^{m} \mathbb{P}(S_n = \vec{0}) &= \sum_{n=1}^{m}\sum_{k=2}^{n} \mathbb{P}(T^r_{\vec{0}} = k)\mathbb{P}(S_{n-k} = \vec{0}) \\
&= \sum_{k=2}^{m}\sum_{n=k}^{m} \mathbb{P}(T^r_{\vec{0}} = k)\mathbb{P}(S_{n-k} = \vec{0})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=2}^{m} \mathbb{P}\big(T_{\vec{0}}^{r} = k\big) \sum_{l=0}^{m-k} \mathbb{P}\big(S_l = \vec{0}\big) \\
&\leqslant \sum_{k=2}^{m} \mathbb{P}\big(T_{\vec{0}}^{r} = k\big) \sum_{l=0}^{m} \mathbb{P}\big(S_l = \vec{0}\big) \\
&= \left(\sum_{n=0}^{m} \mathbb{P}\big(S_n = \vec{0}\big)\right) \left(\sum_{n=2}^{m} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big)\right).
\end{aligned}
$$

c) By letting m tend to ∞ in (3.4.31) we obtain

$$
\mathbb{P}\big(T_{\vec{0}}^{r} < \infty\big) = \sum_{n \geqslant 2} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big) \geqslant 1,
$$

which allows us to conclude. Note that the sum of the series $\sum_{n\geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big)$ actually represents the average number of visits to state ⓪.

d) We have

$$
\begin{aligned}
\sum_{n=1}^{2m} \mathbb{P}\big(S_n = \vec{0}\big) &= \sum_{n=2}^{2m} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big) \sum_{l=0}^{2m-n} \mathbb{P}\big(S_l = \vec{0}\big) \\
&\geqslant \sum_{n=2}^{m} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big) \sum_{l=0}^{2m-n} \mathbb{P}\big(S_l = \vec{0}\big) \\
&\geqslant \sum_{n=2}^{m} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big) \sum_{l=0}^{m} \mathbb{P}\big(S_l = \vec{0}\big).
\end{aligned}
$$

e) Letting m tend to $+\infty$ in (3.4.32) we find

$$
\mathbb{P}\big(T_{\vec{0}}^{r} < \infty\big) = \sum_{n \geqslant 2} \mathbb{P}\big(T_{\vec{0}}^{r} = n\big) \leqslant \frac{\sum_{n\geqslant 2} \mathbb{P}\big(S_n = \vec{0}\big)}{\sum_{n\geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big)} = 1 - \frac{1}{\sum_{n\geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big)} < 1.
$$

f) When $\underline{d = 1}$ we have

$$
\mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n} = \frac{(2n)!}{2^{2n}(n!)^2} \simeq_{n\to\infty} \frac{1}{\sqrt{\pi n}},
$$

by Stirling's approximation, hence

$$
\sum_{n\geqslant 0} \mathbb{P}(S_n = 0) = \infty.
$$

and the result of Question (c) shows that $\mathbb{P}(T_0^r < \infty) = 1$.

g) We note that the random walk can return to state $\vec{0}$ in $2n$ time steps by

- k forward steps in the direction e_1,
- k backward steps in the direction $-e_1$,
- $n-k$ forward steps in the direction e_2,
- $n-k$ backward steps in the direction $-e_2$,

where k ranges from 0 to $2n$. For each $k = 0, 1, \ldots, n$ the number of ways to arrange those four types of moves among $2n$ time steps is the multinomial coefficient

$$\binom{2n}{k, k, n-k, n-k} = \frac{(2n!)}{k!k!(n-k)!(n-k)!},$$

hence, since every sequence of $2n$ moves occur with the same probability $(1/4)^{2n}$, by summation over $k = 0, 1, \ldots, n$ we find

$$\begin{aligned}\mathbb{P}\big(S_{2n} = \vec{0}\big) &= \left(\frac{1}{4}\right)^{2n} \sum_{k=0}^{n} \frac{(2n!)}{(k!)^2((n-k)!)^2} \\ &= \frac{(2n)!}{4^{2n}(n!)^2} \sum_{k=0}^{n} \binom{n}{k}^2 \\ &= \frac{(2n)!}{4^{2n}(n!)^2} \binom{2n}{n} \\ &= \frac{((2n)!)^2}{4^{2n}(n!)^4} \simeq_{n\to\infty} \frac{1}{2\pi n},\end{aligned}$$

where we used Stirling's approximation, and this yields

$$\sum_{n \geqslant 0} \mathbb{P}\big(S_n = \vec{0}\big) = \infty.$$

To conclude, the result of Question (c) shows again that $\mathbb{P}\big(T_{\vec{0}}^r < \infty\big) = 1$.

h) In order to come back to $\vec{0}$ we need to take i_1 forward steps in the direction e_1 and i_1 backward steps in the direction e_1, and similarly for $i_2, \ldots, i_d$. The number of ways to arrange such paths is given by the multinomial coefficient

$$\binom{2n}{i_1, i_1, i_2, i_2, \ldots, i_d, i_d} = \frac{(2n)!}{(i_1!)^2 \cdots (i_d!)^2},$$

and by summation over all possible indices $i_1, i_2, \ldots, i_d \geqslant 0$ satisfying $i_1 + \cdots + i_d = n$ and multiplying by the probability $(1/(2d))^{2n}$ of each path we find

$$\mathbb{P}(S_{2n}=\vec{0}) = \frac{1}{(2d)^{2n}} \sum_{\substack{i_1+\cdots+i_d=n\\ i_1,i_2,\ldots,i_d\geqslant 0}} \binom{2n}{i_1,i_1,i_2,i_2,\ldots,i_d,i_d}$$
$$= \frac{1}{(2d)^{2n}} \sum_{\substack{i_1+\cdots+i_d=n\\ i_1,i_2,\ldots,i_d\geqslant 0}} \frac{(2n)!}{(i_1!)^2\cdots(i_d!)^2}.$$

i) By the hint provided that we have

$$\sum_{n\geqslant 1}\mathbb{P}(S_{2n}=\vec{0}) = \sum_{n\geqslant 1}\frac{1}{(2d)^{2n}}\binom{2n}{n!} \sum_{\substack{i_1+\cdots+i_d=n\\ i_1,i_2,\ldots,i_d\geqslant 0}} \frac{(n!)^2}{(i_1!)^2\cdots(i_d!)^2}$$
$$\leqslant \sum_{n\geqslant 1}\frac{1}{(2d)^{2n}}\binom{2n}{n!}\frac{n!}{(a_n!)^d(a_n+1)^{b_n}} \sum_{\substack{i_1+\cdots+i_d=n\\ i_1,i_2,\ldots,i_d\geqslant 0}} \frac{n!}{i_1!\cdots i_d!}$$
$$\leqslant \sum_{n\geqslant 1}\frac{1}{(2d)^{2n}}\binom{2n}{n!}\frac{n!d^n}{(a_n!)^d a_n^{b_n}} = \sum_{n\geqslant 1}\frac{(2n)!}{2^{2n}d^n n!(a_n!)^d a_n^{b_n}}.$$

j) For sufficiently large n, by the Stirling approximation we have

$$\frac{(2n)!}{2^{2n}d^n n!(a_n!)^d a_n^{b_n}} \simeq \frac{(2n/e)^{2n}\sqrt{4\pi n}}{2^{2n}d^n(n/e)^n\sqrt{2\pi n}\left((a_n/e)^{a_n}\sqrt{2\pi a_n}\right)^d a_n^{b_n}}$$
$$= \frac{\sqrt{2}}{(2\pi)^{d/2}}\frac{n^n}{\mathrm{e}^{b_n}(a_n d)^n a_n^{d/2}}$$
$$= \frac{\sqrt{2}}{(2\pi)^{d/2}}\frac{(1-b_n/n)^{-n}}{\mathrm{e}^{b_n}a_n^{d/2}}$$
$$\leqslant \frac{\sqrt{2}d^{d/2}}{(2\pi)^{d/2}}\frac{(1-(d-1)/n)^{-n}}{(a_n d)^{d/2}}$$
$$\simeq \frac{\sqrt{2}d^{d/2}\mathrm{e}^{d-1}}{(2\pi)^{d/2}}\frac{1}{n^{d/2}},$$

since $a_n d \simeq n$ as n goes to infinity from the relation $a_n d/n = 1 - b_n/n$.

k) The result of Question (j) shows that

$$\sum_{n\geqslant 0}\mathbb{P}(S_n=\vec{0}) < \infty,$$

hence $\mathbb{P}(T_{\vec{0}}^r=\infty) > 0$ by the result of Question (e), and the random walk is *not* recurrent when $d \geqslant 3$.

Problem 3.19

a) The cookie random walk does not have the Markov property when $p \neq 1/2$ because in this case the transition probabilities at a given state may depend on the past behavior of the chain starting from time 1. On the other hand, the cookie random walk has the Markov property when $p = 1/2$ because in this case it coincides with the usual symmetric random walk with independent increments.

b) Knowing that the probability for a symmetric random walk to reach state $\boxed{x+1}$ before hitting state ⓪ starting from ⓚ is $k/(x+1)$, by first step analysis we find the probability

$$p + (1-p)\frac{x-1}{x+1} = 1 - \frac{2q}{x+1}, \qquad x \geqslant 1.$$

c) We have $\mathbb{P}(\tau_1 < \tau_0 \mid S_0 = 0) = 1/2$ and by the (strong) Markov property, by reasoning inductively on the transitions from state ⓪ to state ①, then from state ② to state ②, etc, up to state ⓧ, we find

$$\begin{aligned}\mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) &= \prod_{l=0}^{x-1} \mathbb{P}(\tau_{l+1} < \tau_0 \mid S_0 = l) \\ &= \mathbb{P}(\tau_1 < \tau_0 \mid S_0 = 0) \prod_{l=1}^{x-1} \mathbb{P}(\tau_{l+1} < \tau_0 \mid S_0 = l) \\ &= \frac{1}{2} \prod_{l=1}^{x-1} \left(1 - \frac{2q}{l+1}\right) \\ &= \frac{1}{2} \prod_{l=2}^{x} \left(1 - \frac{2q}{l}\right), \qquad x \geqslant 1.\end{aligned}$$

d) Using the inequality $\log(1+z) \leqslant z$ for $z > -1$, we have

$$\begin{aligned}\sum_{l=2}^{x} \log\left(1 - \frac{2q}{l}\right) &\leqslant \int_2^x \log\left(1 - \frac{2q}{y}\right) dy \\ &\leqslant -2q \int_2^x \frac{1}{y} dy \\ &= -2q \log \frac{x}{2},\end{aligned}$$

hence

$$\begin{aligned}\mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) &= \frac{1}{2} \exp\left(\sum_{l=2}^{x}\left(1 - \frac{2q}{l}\right)\right) \\ &\leqslant \exp\left(-2q \log \frac{x}{2}\right) \leqslant \left(\frac{x}{2}\right)^{-2q},\end{aligned}$$

$x \geqslant 2$.

e) We have

$$\begin{aligned}
\mathbb{P}(\tau_0 < \infty \mid S_0 = 0) &\geqslant \mathbb{P}\left(\bigcup_{x \geqslant 1} \tau_0 < \tau_x \;\middle|\; S_0 = 0 \right) \\
&= \lim_{x \to \infty} \mathbb{P}(\tau_0 < \tau_x \mid S_0 = 0) \\
&= \lim_{x \to \infty} (1 - \mathbb{P}(\tau_x \leqslant \tau_0 \mid S_0 = 0)) \\
&= \lim_{x \to \infty} (1 - \mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0)) \\
&\geqslant 1 - \lim_{x \to \infty} (x/2)^{-2q} \\
&= 1.
\end{aligned}$$

Remark. We note that $\mathbb{P}(\tau_x < \infty \mid S_0 = 0) = 1$, as can be seen by bounding from Question (a) of Problem 2.7 with rebound. Note also that $\mathbb{P}(B \cap A) = \mathbb{P}(B)$ when $\mathbb{P}(A) = 1$.

f) This question can be solved in three possible ways.

i) By first step analysis. We have

$$\mathbb{E}[\tau_1 \mid S_0 = 0] = \frac{1}{2} \times 1 + \frac{1}{2}(1 + \mathbb{E}[\tau_1 \mid S_0 = 0]),$$

hence

$$\mathbb{E}[\tau_1 \mid S_0 = 0] = 2.$$

ii) By pathwise analysis. We have

$$\mathbb{E}[\tau_1 \mid S_0 = 0] = \sum_{k \geqslant 1} \frac{k}{2^k} = \frac{1}{2} \sum_{k \geqslant 1} \frac{k}{2^{k-1}} = \frac{1/2}{(1-1/2)^2} = 2.$$

iii) By applying the result of Question (d) of Problem 2.7 page 81 with $S = 1$ and $k = 0$, which shows that

$$\mathbb{E}[\tau_1 \mid S_0 = 0] = (S + k + 1)(S - k) = 2.$$

g) Applying the result of Question (d) of Problem 2.7 together with first step analysis, we find the mean time

$$p + q(1 + (x+1+(x-1)+1)(x+1-(x-1))) = p + q(1 + 4x + 2) = 1 + q(4x + 2).$$

h) We have

$$\mathbb{E}[\tau_x \mid S_0 = 0] = 2 + \sum_{k=1}^{x-1} (1 + q(4k + 2))$$

$$
\begin{aligned}
&= 2 + (1+2q)(x-1) + 4q\sum_{k=1}^{x-1} k \\
&= 2 + (1+2q)(x-1) + 2qx(x-1) \\
&= 2 + (1+2q)x - (1+2q) + 2qx^2 - 2qx \\
&= 1 - 2q + x + 2qx^2, \qquad x \geqslant 1.
\end{aligned}
$$

i) Using the notation of Problem 2.7 we have

$$
\begin{aligned}
\mathbb{P}(S_1 = x+1 \mid S_0 = x \text{ and } \tau_{x+1} < \tau_0) &= p\frac{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x+1)}{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x)} \\
&= \frac{p}{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}(S_1 = x-1 \mid S_0 = x, \tau_{x+1} < \tau_0) &= q\frac{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x-1)}{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x)} \\
&= \frac{q(x-1)/(x+1)}{\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x)},
\end{aligned}
$$

as we have $\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x+1) = 1$ and

$$
\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x-1) = \frac{x-1}{x+1},
$$

because the random walk becomes symmetric when started from state $\boxed{x-1}$. Next, we note that, according to Question (b), $\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x)$ can be computed as

$$
\begin{aligned}
\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = x) &= p\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_1 = x+1) \\
&\quad + q\mathbb{P}(\tau_{x+1} < \tau_0 \mid S_1 = x-1) \\
&= p + q\frac{x-1}{x+1},
\end{aligned}
$$

and in this case we get

$$
\mathbb{P}(S_1 = x+1 \mid S_0 = x \text{ and } \tau_{x+1} < \tau_0) = \frac{p}{p + q(x-1)/(x+1)} = \frac{p}{1 - 2q/(x+1)}
$$

and

$$
\mathbb{P}(S_1 = x-1 \mid S_0 = x \text{ and } \tau_{x+1} < \tau_0) = \frac{q(x-1)/(x+1)}{p + q(x-1)/(x+1)} = \frac{q(x-1)/(x+1)}{1 - 2q/(x+1)}.
$$

j) The mean time is given by

$$\begin{aligned}\mathbb{E}[\tau_{x+1} \mid S_0 = x, \tau_{x+1} < \tau_0] &= 1 + \frac{((x+1)^2 - (x-1)^2)}{3} \frac{q(x-1)/(x+1)}{1-2q/(x+1)} \\ &= 1 + \frac{4qx(x-1)/(x+1)}{3(1-2q/(x+1))} \\ &= 1 + \frac{4qx(x-1)}{3(x+1-2q)},\end{aligned}$$

which yields $1 + (2x-2)/3$ when $p = q = 1/2$.

Remark. Using the above result we can also compute the average time to reach state ⓧ with $x \geqslant 2$ after starting from state ① and *given one does not hit* ⓪ by the summation

$$\begin{aligned}&\mathbb{E}[\tau_x \mid S_0 = 1,\ \tau_x < \tau_0] = \sum_{k=2}^{x} \mathbb{E}[\tau_k \mid S_0 = k-1, \tau_k < \tau_0] \\ &= x - 1 + \frac{4q}{3} \sum_{k=2}^{x} \frac{(k-1)(k-2)}{k-2q} \\ &= x - 1 + \frac{2qx(x-1)}{3} - \frac{8p\tilde{q}}{3} \sum_{k=1}^{x-1} \frac{k}{k+1-2q} \\ &= x - 1 + \frac{2qx(x-1)}{3} - \frac{8pq}{3}(x-1) + \frac{8pq}{3}(p-q) \sum_{k=1}^{x-1} \frac{1}{k+1-2q} \\ &= x - 1 + \frac{4q}{3} \left(\frac{x(x-1)}{2} - 2p(x-1) + 2p(p-q) \sum_{k=1}^{x-1} \frac{1}{k+1-2q} \right).\end{aligned}$$

When $p = q = 1/2$ we recover the known expression

$$\begin{aligned}\mathbb{E}[\tau_x \mid S_0 = 1,\ \tau_x < \tau_0] &= x - 1 + \frac{2}{3} \sum_{k=1}^{x-2} k \\ &= x - 1 + \frac{(x-1)(x-2)}{3} \\ &= \frac{x^2 - 1}{3}, \qquad x \geqslant 2,\end{aligned}$$

cf. Question (j) of Problem 2.7.

Problem 3.20

a) The probability $\mathbb{P}(X = 0)$ that the random walk eats no cookies before hitting the origin is the probability of going directly from ⓪ to ⓪ in one time step, which is $1/2$.

The probability $\mathbb{P}(X=1)$ that the random walk eats exactly *one* cookie before hitting the origin is the probability of first moving from ⓪ to ① in one time step and then back to ⓪ in one time step, that is $q \times (1/2) = q/2$.

In general, we have

$$\begin{aligned}\mathbb{P}(X=x) &= \mathbb{P}(\tau_x < \tau_0 \mid S_0 = 0) - \mathbb{P}(\tau_{x+1} < \tau_0 \mid S_0 = 0) \\ &= \frac{1}{2}\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right) - \frac{1}{2}\prod_{l=2}^{x+1}\left(1-\frac{2q}{l}\right), \\ &= \frac{1}{2}\left(1-\left(1-\frac{2q}{x+1}\right)\right)\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right) \\ &= \frac{q}{x+1}\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right).\end{aligned}$$

b) We have

$$\mathbb{E}[X] = \sum_{x\geqslant 0} x\mathbb{P}(X=x) = q\sum_{x\geqslant 0}\frac{x}{x+1}\prod_{l=2}^{x}\left(1-\frac{2q}{l}\right),$$

hence

$$qc_q\sum_{x\geqslant 0}\frac{x}{(x+1)x^{2q}} \leqslant \mathbb{E}[X] \leqslant qC_q\sum_{x\geqslant 0}\frac{x}{(x+1)x^{2q}},$$

and $\mathbb{E}[X]$ is finite if and only if $2q > 1$.

Remark. One could show in addition that the mean return time to ⓪ is always infinite, see Antal and Redner (2005).

Chapter 4 - Discrete-Time Markov Chains

Exercise 4.1

a) This process is a Markov chain because its increments $Z_{n+1} - Z_n = 2(S_{n+1} - S_n)$, $n \in \mathbb{N}$, are independent random variables.
b) We have

$$\begin{aligned}Z_{n+1} &= (S_{n+1})^2 \\ &= (S_{n+1} - S_n + S_n)^2\end{aligned}$$

$$
\begin{aligned}
&= (S_n)^2 + 2(S_{n+1} - S_n)S_n + (S_{n+1} - S_n)^2 \\
&= (S_n)^2 + 2(S_{n+1} - S_n)S_n + 1 \\
&= Z_n + 2X_{n+1}S_n + 1,
\end{aligned}
$$

where $(X_n)_{n\geqslant 1} = (S_n - S_{n-1})_{n\geqslant 1}$ denotes the independent increments of $(S_n)_{n\geqslant 0}$, and the increment of $(Z_n)_{n\in\mathbb{N}}$ is

$$
Z_{n+1} - Z_n = 2X_{n+1}S_n + 1, \qquad n \geqslant 0.
$$

In the symmetric case $\mathbb{P}(X_{n+1} = \pm 1) = 1/2$, the process $(Z_n)_{n\geqslant 0}$ is a Markov chain because $X_{n+1}S_n$ takes the values

$$
\{S_n, -S_n\} = \{\sqrt{Z_n}, -\sqrt{Z_n}\}
$$

with probabilities $1/2$, hence the probability distribution of $X_{n+1}S_n$ is always $(1/2, 1/2)$ over the two values $\{\sqrt{Z_n}, -\sqrt{Z_n}\}$, and the distribution of Z_{n+1} is fully determined given the value of $Z_n = (S_n)^2$.

In the non symmetric case, with

$$
\mathbb{P}(X_{n+1} = 1) = p, \quad \mathbb{P}(X_{n+1} = -1) = q, \quad \text{and} \quad p \neq q,
$$

the process $(Z_n)_{n\geqslant 0}$ will not be a Markov chain because when $S_n > 0$ we have

$$
(S_n, -S_n) = (\sqrt{Z_n}, -\sqrt{Z_n})
$$

and the probability distribution of $X_{n+1}S_n$ over $(\sqrt{Z_n}, -\sqrt{Z_n})$ is (p, q). On the other hand, when $S_n < 0$ we have

$$
(S_n, -S_n) = (-\sqrt{Z_n}, \sqrt{Z_n})
$$

and the probability distribution of $X_{n+1}S_n$ over $(\sqrt{Z_n}, -\sqrt{Z_n})$ becomes (q, p) which differs from (p, q) since $p \neq q$. Hence the sign of S_n is needed and the knowledge of $Z_n = |S_n|$ is no longer sufficient in order to determine the distribution of Z_{n+1} given Z_n, therefore $(Z_n)_{n\geqslant 0}$ is not a Markov chain.

Exercise 4.2

a) By the Markov property, we have

$$
\begin{aligned}
&\mathbb{P}(X_7 = 1 \text{ and } X_5 = 2 \mid X_4 = 1 \text{ and } X_3 = 2) \\
&\qquad = \mathbb{P}(X_7 = 1 \text{ and } X_5 = 2 \mid X_4 = 1) \\
&\qquad = \frac{\mathbb{P}(X_7 = 1,\ X_5 = 2,\ X_4 = 1)}{\mathbb{P}(X_4 = 1)}
\end{aligned}
$$

$$= \frac{\mathbb{P}(X_7 = 1,\ X_5 = 2,\ X_4 = 1)}{\mathbb{P}(X_5 = 2 \text{ and } X_4 = 1)} \frac{\mathbb{P}(X_5 = 2 \text{ and } X_4 = 1)}{\mathbb{P}(X_4 = 1)}$$
$$= \mathbb{P}(X_7 = 1 \mid X_5 = 2 \text{ and } X_4 = 1)\mathbb{P}(X_5 = 2 \mid\ X_4 = 1)$$
$$= 0.6 \times \mathbb{P}(X_7 = 1 \mid X_5 = 2).$$

Next, we note that

$$P^2 = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \times \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.64 & 0.36 \\ 0.48 & 0.52 \end{bmatrix},$$

hence $\mathbb{P}(X_7 = 1 \mid X_5 = 2) = 0.48$, and we conclude that

$$\mathbb{P}(X_7 = 1 \text{ and } X_5 = 2 \mid X_4 = 1 \text{ and } X_3 = 2) = 0.6 \times 0.48 = 0.288.$$

b) We have

$$\mathbb{E}[X_2 \mid X_1 = 1] = 1 \times \mathbb{P}(X_2 = 1 \mid X_1 = 1) + 2 \times \mathbb{P}(X_2 = 2 \mid X_1 = 1)$$
$$= 0.4 + 2 \times 0.6 = 1.6.$$

Exercise 4.3

a) We have

$$P^n = P = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_N \end{bmatrix}, \qquad \text{(S.30)}$$

for all $n \geqslant 1$.

b) The vector π is an invariant (or stationary) distribution for P because it satisfies

$$[\pi_0, \pi_1, \ldots, \pi_N] = [\pi_0, \ldots, \pi_N] \times \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots & \pi_N \end{bmatrix},$$

as in (4.2.5).

c) Since $P^n = P$, $n \geqslant 1$, we have

$$\mathbb{P}(Z_n = i \mid Z_k = j) = \mathbb{P}(Z_n = i) = \pi_i, \quad i, j = 0, 1, \ldots, N, \quad 0 \leqslant k < n,$$

hence by the Markov property we find

$$\begin{aligned} \mathbb{P}(Z_n = i \text{ and } Z_k = j) &= \mathbb{P}(Z_n = i \mid Z_k = j)\mathbb{P}(Z_k = j) \\ &= \mathbb{P}(Z_n = i)\mathbb{P}(Z_k = j), \end{aligned} \tag{S.31}$$

$i, j = 0, 1, \ldots, N$, which shows that Z_n is independent of Z_k for $0 \leqslant k < n$. Similarly, using induction and the Markov property, or (4.1.1), shows that

$$\begin{aligned} &\mathbb{P}(Z_n = i_n,\ Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) \\ &= \mathbb{P}(Z_n = i_n \mid Z_{n-1} = i_{n-1}) \cdots \mathbb{P}(Z_1 = i_1 \mid Z_0 = i_0)\mathbb{P}(Z_0 = i_0) \\ &= \mathbb{P}(Z_n = i_n)\mathbb{P}(Z_{n-1} = i_{n-1}) \cdots \mathbb{P}(Z_0 = i_0) \\ &= \pi_{i_n}\pi_{i_{n-1}} \cdots \pi_{i_0}, \end{aligned}$$

$i_0, i_1, \ldots, i_n \in \{0, 1, \ldots, N\}$, which shows that $(Z_n)_{n\in\mathbb{N}}$ is an *i.i.d* sequence of random variables with distribution π over $\{0, 1, \ldots, N\}$.

Exercise 4.4

a) We find

$$\begin{cases} \mathbb{P}(O_0 = a \mid X_0 = 0)\mathbb{P}(O_1 = b \mid X_1 = 0) = m_{0,a}m_{0,b}, \\ \\ \mathbb{P}(O_0 = a \mid X_0 = 0)\mathbb{P}(O_1 = b \mid X_1 = 1) = m_{0,a}m_{1,b}, \\ \\ \mathbb{P}(O_0 = a \mid X_0 = 1)\mathbb{P}(O_1 = b \mid X_1 = 0) = m_{1,a}m_{0,b}, \\ \\ \mathbb{P}(O_0 = a \mid X_0 = 1)\mathbb{P}(O_1 = b \mid X_1 = 1) = m_{1,a}m_{1,b}, \end{cases}$$

b) We find $\mathbb{P}(X_0 = 1, X_1 = 1) = \mathbb{P}\big((X_0, X_1) = (1, 1)\big) = \pi_1 P_{1,1}$, and

$$\begin{cases} \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (0, 0)\big) = \pi_0 P_{0,0} m_{1,a} m_{0,b}, \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (0, 1)\big) = \pi_0 P_{0,1} m_{0,a} m_{1,b}, \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (1, 0)\big) = \pi_1 P_{1,0} m_{1,a} m_{0,b}, \\ \\ \mathbb{P}\big((O_0, O_1) = (a, b) \text{ and } (X_0, X_1) = (1, 1)\big) = \pi_1 P_{1,1} m_{1,a} m_{1,b}, \end{cases}$$

c) We have

$$\begin{aligned}
&\mathbb{P}((O_0,O_1)=(a,b))\\
&=\sum_{x,y\in\{0,1\}}\mathbb{P}((O_0,O_1)=(a,b)\text{ and }(X_0,X_1)=(x,y))\\
&=\sum_{x,y\in\{0,1\}}\pi_xP_{x,y}m_{x,a}m_{y,b}\\
&=\pi_0P_{0,0}m_{0,a}m_{0,b}+\pi_0P_{0,1}m_{0,a}m_{1,b}+\pi_1P_{1,0}m_{1,a}m_{0,b}+\pi_1P_{1,1}m_{1,a}m_{1,b}.
\end{aligned}$$

d) We have

$$\begin{cases}
\mathbb{P}((X_0,X_1)=(0,1)\mid(O_0,O_1)=(a,b))=\pi_0P_{0,1}m_{0,a}m_{1,b}/\mathbb{P}((O_0,O_1)=(a,b)),\\
\\
\mathbb{P}((X_0,X_1)=(1,0)\mid(O_0,O_1)=(a,b))=\pi_1P_{1,0}m_{1,a}m_{0,b}/\mathbb{P}((O_0,O_1)=(a,b)).
\end{cases}$$

On the other hand, we have

$$\begin{aligned}
\{X_1=1\}&=\{(X_0,X_1)=(0,1)\}\bigcup\{(X_0,X_1)=(0,1)\}\\
&\qquad\bigcup\{(X_0,X_1)=(1,1)\}\bigcup\{(X_0,X_1)=(1,1)\}\\
&=\bigcup_{x\in\{0,1\}}\{(X_0,X_1)=(x,1)\}\\
&=\{(X_0,X_1)=(0,1)\}\bigcup\{(X_0,X_1)=(1,1)\},
\end{aligned}$$

where the above union is a partition, hence

$$\begin{aligned}
&\mathbb{P}(X_1=1\mid(O_0,O_1)=(a,b))\\
&=\sum_{x\in\{0,1\}}\mathbb{P}((X_0,X_1)=(x,1)\mid(O_0,O_1)=(a,b))\\
&=\frac{1}{\mathbb{P}((O_0,O_1)=(a,b))}\sum_{x\in\{0,1\}}\mathbb{P}((X_0,X_1)=(x,1)\text{ and }(O_0,O_1)=(a,b))\\
&=\frac{1}{\mathbb{P}((O_0,O_1)=(a,b))}\sum_{x\in\{0,1\}}\pi_xP_{x,1}m_{x,a}m_{1,b}\\
&=\frac{1}{\mathbb{P}((O_0,O_1)=(a,b))}\left(\pi_0P_{0,1}m_{0,a}m_{1,b}+\pi_1P_{1,1}m_{1,a}m_{1,b}\right).
\end{aligned}$$

Exercise 4.5

a) The process $(Z_n)_{n\in\mathbb{N}}$ cannot be a Markov chain because when *e.g.* $Z_n=25$ the distribution of the next value of the chain at time $n+1$ is not uniquely determined from the data of $Z_n=25$. Indeed, it may depend on how the state 25 has been reached, *e.g.* through $(5,0)$ or $(4,3)$, and this information will influence the next transition probabilities, which can then can depend on past chain values from time 0. Other examples include $(99, 101)$ and $(11, 141)$ whose sums of squares are both equal to $99^2 + 101^2 = 11^2 + 141^2 = 20002$.

b) The random walk $(S_n)_{n\in\mathbb{N}}$ cannot have the Markov property because the behavior of the chain at time $n+1$ may depend on the past of the chain from time 0 to time $n-1$. Indeed, the fact that have visited a given state between time 0 to time $n-1$ will affect the probability of visiting that state at time n.

Exercise 4.6 The Elephant Random Walk $(S_n)_{n\in\mathbb{N}}$ does *not* have the Markov property because the distribution of S_{n+1} given the value of S_n depends on extra information that relies on the past data of $(X_1, \ldots, X_{n-1})$.

Exercise 4.7 The state space of $(Z_n)_{n\geqslant 1}$ is $\{(0,0), (0,1), (1,0), (1,1)\}$. Based on $Z_n = (X_{n-1}, X_n)$, the distribution of $Z_{n+1} = (X_n, X_{n+1})$ at time $n+1$ is fully determined from the data of X_n and the transition matrix of $(X_n)_{n\in\mathbb{N}}$ hence $(Z_n)_{n\geqslant 1}$ is a $(0,1)$-valued Markov chain and its transition matrix is given by

$$\begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{bmatrix} 1-a & a & 0 & 0 \\ 0 & 0 & b & 1-b \\ 1-a & a & 0 & 0 \\ 0 & 0 & b & 1-b \end{bmatrix} \end{array},$$

Exercise 4.8

a) The probability distribution of T_2 is negative binomial with parameters $(p, 2)$, *i.e.*

$$\mathbb{P}(T_2 = k \mid X_0 = 0) = \binom{k-1}{k-2}(1-p)^2 p^{k-2}, \qquad k \geqslant 2.$$

b) We have

$$\begin{aligned} \mathbb{E}[T_2 \mid X_0 = 0] &= \sum_{k\geqslant 2} k\mathbb{P}(T_2 = k \mid X_0 = 0) \\ &= (1-p)^2 \sum_{k\geqslant 2} k\binom{k-1}{k-2} p^{k-2} \\ &= (1-p)^2 \sum_{k\geqslant 2} k(k-1)p^{k-2} \end{aligned}$$

$$
\begin{aligned}
&= (1-p)^2 \frac{\partial^2}{\partial p^2} \sum_{k \geqslant 0} p^k \\
&= (1-p)^2 \frac{\partial^2}{\partial p^2} \frac{1}{1-p} \\
&= \frac{2}{1-p} \\
&= \frac{2}{q}.
\end{aligned}
$$

Remark. We could also recover the values $\mathbb{E}[T_1 \mid X_0 = 0]$ and $\mathbb{E}[T_2 \mid X_0 = 1]$ from the first step analysis equations

$$
\mathbb{E}[T_2 \mid X_0 = 1] = p(1 + \mathbb{E}[T_2 \mid X_0 = 1]) + q \times 1.
$$

and

$$
\mathbb{E}[T_2 \mid X_0 = 0] = p(1 + \mathbb{E}[T_2 \mid X_0 = 0]) + q \times (1 + \mathbb{E}[T_2 \mid X_0 = 1]),
$$

which yield

$$
\mathbb{E}[T_2 \mid X_0 = 1] = \frac{1}{q} \quad \text{and} \quad \mathbb{E}[T_2 \mid X_0 = 0] = \frac{2}{q}.
$$

Exercise 4.9 When $N = 5$, the transition matrix of this chain is

$$
P = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
1/25 & 8/25 & 16/25 & 0 & 0 & 0 \\
0 & 4/25 & 12/25 & 9/25 & 0 & 0 \\
0 & 0 & 9/25 & 12/25 & 4/25 & 0 \\
0 & 0 & 0 & 16/25 & 8/25 & 1/25 \\
0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}.
$$

For general $N \geqslant 1$, the transition matrix reads

$$
P = \begin{bmatrix}
0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
1/N^2 & 2(N-1)/N^2 & (N-1)^2/N^2 & 0 & \cdots & \cdots & 0 & 0 \\
0 & 2^2/N^2 & 4(N-2)/N^2 & (N-2)^2/N^2 & \cdots & \cdots & 0 & 0 \\
0 & 0 & 3^2/N^2 & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \vdots & 0 & 3^2/N^2 & 0 & 0 \\
0 & 0 & \cdots & 0 & (N-2)^2/N^2 & 4(N-2)/N^2 & 2^2/N^2 & 0 \\
0 & 0 & \cdots & 0 & 0 & (N-1)^2/N^2 & 2(N-1)/N^2 & 1/N^2 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0
\end{bmatrix},
$$

and the matrix entries are given in general as

$$\begin{cases} P_{k,k-1} = \mathbb{P}(X_1 = k-1 \mid X_0 = k) = \dfrac{k^2}{N^2}, \\ \\ P_{k,k} = \mathbb{P}(X_1 = k \mid X_0 = k) = \dfrac{2k(N-k)^2}{N^2}, \\ \\ P_{k,k+1} = \mathbb{P}(X_1 = k+1 \mid X_0 = k) = \dfrac{(N-k)^2}{N^2}, \end{cases}$$

$k = 1, 2, \ldots, N-1$, with $P_{0,1} = P_{N,N-1} = 1$. We check that

$$P_{k,k-1} + P_{k,k} + P_{k,k+1} = \frac{k^2 + 2k(N-k)^2 + (N-k)^2}{N^2} = 1, \quad k = 1, 2, \ldots, N-1.$$

Exercise 4.10

a) Let S_n denote the wealth of the player at time $n \in \mathbb{N}$. The process $(S_n)_{n\in\mathbb{N}}$ is a Markov chain whose transition matrix is given by

$$P = \left[P_{i,j} \right]_{i,j\in\mathbb{N}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & 0 & \cdots \\ q & 0 & 0 & p & 0 & 0 & \cdots \\ q & 0 & 0 & 0 & p & 0 & \cdots \\ q & 0 & 0 & 0 & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

After n time steps we have

$$\mathbb{P}(S_n = n+1 \mid S_0 = 1) = p^n, \qquad n \geqslant 1,$$

and, by decomposing over n possible paths made of l steps up and one step down, $l = 0, 1, \ldots, n-1$, we find

$$\begin{aligned} \mathbb{P}(S_n = 0 \mid S_0 = 1) &= \sum_{l=0}^{n-1} qp^l \\ &= q\frac{1-p^n}{1-p} \\ &= 1 - p^n \\ &= 1 - \mathbb{P}(S_n = n+1 \mid S_0 = 1), \qquad n \geqslant 1. \end{aligned}$$

More generally, we have

$$\mathbb{P}(S_n = n+k \mid S_0 = k) = p^n, \qquad k, n \geqslant 1,$$

and

$$\begin{aligned}\mathbb{P}(S_n = 0 \mid S_0 = k) &= q \sum_{l=0}^{n-1} p^l \\ &= q \frac{1-p^n}{1-p} \\ &= 1 - p^n \\ &= 1 - \mathbb{P}(S_n = n+k \mid S_0 = k), \quad k, n \geqslant 1,\end{aligned}$$

by (A.2), hence P^n is given by

$$P^n = \big[[P^n]_{i,j} \big]_{i,j \in \mathbb{N}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & p^n & 0 & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & p^n & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & p^n & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & 0 & p^n & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & p^n & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{S.32}$$

in which the n columns n^o 2 to $n+1$ are identically 0. We can also check by matrix multiplication that this relation is consistent with $P^{n+1} = P \times P^n$, $n \geqslant 1$, *i.e.*

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p^{n+1} & 0 & \cdots & 0 & p^{n+1} & 0 & 0 & 0 & 0 & \cdots \\ 1-p^{n+1} & 0 & \cdots & 0 & 0 & p^{n+1} & 0 & 0 & 0 & \cdots \\ 1-p^{n+1} & 0 & \cdots & 0 & 0 & 0 & p^{n+1} & 0 & 0 & \cdots \\ 1-p^{n+1} & 0 & \cdots & 0 & 0 & 0 & 0 & p^{n+1} & 0 & \cdots \\ 1-p^{n+1} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & p^{n+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p & 0 & p & 0 & 0 & 0 & 0 & \cdots \\ 1-p & 0 & 0 & p & 0 & 0 & 0 & \cdots \\ 1-p & 0 & 0 & 0 & p & 0 & 0 & \cdots \\ 1-p & 0 & 0 & 0 & 0 & p & 0 & \cdots \\ 1-p & 0 & 0 & 0 & 0 & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \times \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & p^n & 0 & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & p^n & 0 & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & p^n & 0 & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & 0 & p^n & 0 & \cdots \\ 1-p^n & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & p^n & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

from the relation $1 - p + p(1-p^n) = 1 - p^{n+1}$, $n \in \mathbb{N}$.

b) In this case the transition matrix P becomes

$$P = \left[P_{i,j} \right]_{i,j \in \mathbb{N}} = \begin{bmatrix} q & p & 0 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & 0 & \cdots \\ q & 0 & 0 & p & 0 & 0 & \cdots \\ q & 0 & 0 & 0 & p & 0 & \cdots \\ q & 0 & 0 & 0 & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

with

$$P^2 = \begin{bmatrix} q & qp & p^2 & 0 & 0 & 0 & \cdots \\ q & qp & 0 & p^2 & 0 & 0 & \cdots \\ q & qp & 0 & 0 & p^2 & 0 & \cdots \\ q & qp & 0 & 0 & 0 & p^2 & \cdots \\ q & qp & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

and more generally, by induction on $n \geqslant 2$ we find

$$P^n = \begin{bmatrix} q & qp & qp^2 & \ldots & qp^{n-1} & p^n & 0 & 0 & 0 & \cdots \\ q & qp & qp^2 & \ldots & qp^{n-1} & 0 & p^n & 0 & 0 & \cdots \\ q & qp & qp^2 & \ldots & qp^{n-1} & 0 & 0 & p^n & 0 & \cdots \\ q & qp & qp^2 & \ldots & qp^{n-1} & 0 & 0 & 0 & p^n & \cdots \\ q & qp & qp^2 & \ldots & qp^{n-1} & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{S.33}$$

Exercise 4.11

a) We have

$$\begin{aligned}
&\mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 2) = \frac{\mathbb{P}(Z_{n+1} = 2,\ Z_n = 0,\ Z_{n-1} = 2)}{\mathbb{P}(Z_n = 0 \text{ and } Z_{n-1} = 2)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2,\ X_n \in \{0,1\},\ X_{n-1} = 2)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 2)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 0,\ X_{n-1} = 2)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 2)} + \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 1,\ X_{n-1} = 2)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 2)} \\
&= 0,
\end{aligned}$$

$n \geqslant 1$. Next, we have

$$\begin{aligned}
&\mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 3) = \frac{\mathbb{P}(Z_{n+1} = 2,\ Z_n = 0,\ Z_{n-1} = 3)}{\mathbb{P}(Z_n = 0 \text{ and } Z_{n-1} = 3)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2,\ X_n \in \{0,1\},\ X_{n-1} = 3)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 3)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 0,\ X_{n-1} = 3)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 3)} + \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 1,\ X_{n-1} = 3)}{\mathbb{P}(X_n \in \{0,1\} \text{ and } X_{n-1} = 3)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 0,\ X_{n-1} = 3)}{\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3) + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)} \\
&\quad + \frac{\mathbb{P}(X_{n+1} = 2,\ X_n = 1,\ X_{n-1} = 3)}{\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3) + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)} \\
&= \frac{\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3)\mathbb{P}(X_{n+1} = 2 \mid\ X_n = 0 \text{ and } X_{n-1} = 3)}{\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3) + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)} \\
&\quad + \frac{\mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)\mathbb{P}(X_{n+1} = 2 \mid X_n = 1 \text{ and } X_{n-1} = 3)}{\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3) + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2 \mid\ X_n = 0)}{1 + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)/\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3)} \\
&\quad + \frac{\mathbb{P}(X_{n+1} = 2 \mid X_n = 1)}{1 + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)/\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3)} \\
&= \frac{\mathbb{P}(X_{n+1} = 2 \mid\ X_n = 0)}{1 + \mathbb{P}(X_n = 1 \text{ and } X_{n-1} = 3)/\mathbb{P}(X_n = 0 \text{ and } X_{n-1} = 3)} \\
&= \frac{1/2}{1 + (1/3)/(1/3)} \\
&= \frac{1}{4},
\end{aligned}
$$

$n \geqslant 1$.

b) The process $(Z_n)_{n\in\mathbb{N}}$ is *not* a Markov chain because by Question (a) we have

$$
\begin{aligned}
0 &= \mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 2) \\
&\neq \mathbb{P}(Z_{n+1} = 2 \mid Z_n = 0 \text{ and } Z_{n-1} = 3) \\
&= \frac{1}{4}.
\end{aligned}
$$

Exercise 4.12

a) We find

$$
360\eta_L + 505\eta_M + 640\eta_H = 360 \times 50\% + 505 \times 40\% + 640 \times 10\% \simeq 446.
$$

b) We find

$$
\begin{aligned}
\eta P &= [50\%, 40\%, 10\%] \times \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} \\
&= \left[\frac{2 \times 0.5}{3} + \frac{0.4}{3} + \frac{0.1}{6}, \frac{0.5}{6} + \frac{0.4}{2} + \frac{2 \times 0.1}{3}, \frac{0.5}{6} + \frac{0.4}{6} + \frac{0.1}{6}\right] \\
&= \left[\frac{29}{60}, \frac{21}{60}, \frac{10}{60}\right]
\end{aligned}
$$

$$\simeq [48.3\%, 35\%, 16.7\%] .$$

c) We find

$$360\eta_L + 505\eta_M + 640\eta_H = 360 \times \frac{29}{60} + 505 \times \frac{21}{60} + 640 \times \frac{10}{60} \simeq 457.$$

d) The equation $\pi = \pi P$ reads

$$[\pi_L, \pi_M, \pi_H] = [\pi_L, \pi_M, \pi_H] \times P = [\pi_L, \pi_M, \pi_H] \times \begin{bmatrix} 8/12 & 2/12 & 2/12 \\ 4/12 & 6/12 & 2/12 \\ 2/12 & 8/12 & 2/12 \end{bmatrix},$$

and it can be rewritten as

$$\pi(P - I) = \pi P - \pi = 0,$$

i.e.,

$$\begin{aligned} \pi(P - I) &= [\pi_L, \pi_M, \pi_H](P - I) \\ &= [\pi_L, \pi_M, \pi_H] \times \begin{bmatrix} -4/12 & 2/12 & 2/12 \\ 4/12 & -6/12 & 2/12 \\ 2/12 & 8/12 & -10/12 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \end{aligned}$$

which can be rewritten as

$$\begin{cases} \pi_L - 3\pi_M + 4\pi_H = 0 \\ \\ \pi_L + \pi_M - 5\pi_H = 0. \end{cases}$$

Combining the above with the additional condition

$$\pi_L + \pi_M + \pi_H = 1,$$

we arrive at

$$[\pi_L, \pi_M, \pi_H] = \left[\frac{11}{24}, \frac{9}{24}, \frac{4}{24}\right] \simeq [45.8\%, 37.5\%, 16.7\%].$$

Exercise 4.13

a) By summing over $o_1, \ldots, o_t$ we have

$$\begin{aligned} &\mathbb{P}(X_t = i_t, \ldots, X_0 = i_0) \\ &= \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}) \cdots \mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_0 = i_0) \end{aligned}$$

$$= \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1})\mathbb{P}(X_{t-1} = i_{t-1}, \ldots, X_0 = i_0),$$

which recovers (4.1.1) as

$$\mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}), \quad t \geqslant 1.$$

b) We have

$$\begin{aligned}
&\mathbb{P}(X_t = i_t, \ldots, X_0 = i_0,\ O_t = o_t, \ldots, O_1 = o_1) \\
&= \mathbb{P}(O_t = o_t \mid X_t = i_t)\mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}) \\
&\quad \mathbb{P}(X_{t-1} = i_{t-1}, \ldots, X_0 = i_0, O_{t-1} = o_{t-1}, \ldots, O_1 = o_1),
\end{aligned}$$

hence by summing over $i_0, i_1, \ldots, i_{t-2}$ and o_{t-1}, we have

$$\begin{aligned}
&\mathbb{P}(X_t = i_t, X_{t-1} = i_{t-1},\ O_{t-1} = o_{t-1}, \ldots, O_1 = o_1) \\
&= \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1})\mathbb{P}(X_{t-1} = i_{t-1}, O_{t-1} = o_{t-1}, \ldots, O_1 = o_1),
\end{aligned}$$

which implies

$$\begin{aligned}
&\mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1},\ O_{t-1} = o_{t-1}, \ldots, O_1 = o_1) \\
&\qquad = \mathbb{P}(X_t = i_t \mid X_{t-1} = i_{t-1}), \qquad t \geqslant 1. \qquad \text{(S.34)}
\end{aligned}$$

Problem 4.14

a) We have

$$P = \begin{bmatrix} 1 & 0 \\ q & p \end{bmatrix},$$

with $q := 1 - p$.

b) We have

$$P = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
q & p & 0 & \cdots & 0 & 0 \\
0 & q & p & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & q & p & 0 \\
0 & 0 & \cdots & 0 & q & p
\end{bmatrix},$$

with $q := 1 - p$.

c) Clearly, the first row of P has to be $[1, 0, \ldots, 0]$ because state ⓪ is absorbing, and the remaining of the matrix can take the form $[\alpha, Q]$. In addition, every row of the $d \times (d+1)$ matrix $[\alpha, Q]$ has to add up to one, *i.e.*

$$\alpha_k + \sum_{l=1}^{d} Q_{k,l} = 1, \qquad k = 1, 2, \ldots, d,$$

which can be rewritten as

$$(I-Q)e = \begin{bmatrix} 1-Q_{1,1} & -Q_{1,2} & \cdots & -Q_{1,d} \\ -Q_{1,1} & 1-Q_{1,2} & \cdots & -Q_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ -Q_{d,1} & \cdots & Q_{d,d-1} & 1-Q_{d,d} \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1-Q_{1,1}-\cdots-Q_{1,d} \\ 1-Q_{2,1}-\cdots-Q_{2,d} \\ \vdots \\ 1-Q_{d,1}-\cdots-Q_{d,d} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{bmatrix},$$

i.e.

$$\alpha = (I-Q)e.$$

d) Clearly, the conclusion holds for $n=0$, and also at the rank $n=1$ since $\alpha = (I-Q)e$. Next, we assume that the relation (4.5.9) holds at the rank $n \geqslant 0$. In this case, we have

$$\begin{aligned} P^{n+1} &= P \times P^n \\ &= \begin{bmatrix} 1 & 0 \\ \alpha & Q \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ (I-Q^n)e & Q^n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \alpha + Q(I-Q^n)e & Q^{n+1} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ (I-Q^{n+1})e & Q^{n+1} \end{bmatrix}, \end{aligned}$$

since

$$\alpha + Q(I-Q^n)e = (I-Q)e + (Q-Q^{n+1})e = (I-Q^{n+1})e.$$

e) Since $[P^{n-1}]_{i,k} = [Q^{n-1}]_{i,k}$ and $\alpha_k = P_{k,0}$, $k=1,2,\ldots,d$, we find that

$$\begin{aligned} \mathbb{P}(T_0 = n \mid X_0 = i) &= \mathbb{P}\left(\bigcup_{k=1}^{d} \{X_{n-1} = k,\ T_0 = n\} \;\middle|\; X_0 = i \right) \\ &= \sum_{k=1}^{d} \mathbb{P}(X_{n-1} = k \mid X_0 = i)\mathbb{P}(X_n = 0 \mid X_{n-1} = k) \\ &= \sum_{k=1}^{d} [P^{n-1}]_{i,k} P_{k,0} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{d} \alpha_k [Q^{n-1}]_{i,k} \\
&= [Q^{n-1}\alpha]_i, \qquad n \geqslant 1.
\end{aligned}
$$

f) We have

$$
\begin{aligned}
\mathbb{P}(T_0 = n) &= \sum_{i=1}^{d} \mathbb{P}(X_0 = i)\mathbb{P}(T_0 = n \mid X_0 = i) \\
&= \sum_{i=1}^{d}\sum_{k=1}^{d} \beta_i \alpha_k Q_{i,k}^{n-1} \\
&= \beta^\top Q^{n-1}\alpha, \qquad n \geqslant 1.
\end{aligned}
$$

g) We have

$$
\begin{aligned}
\mathbb{P}(T_0 \leqslant n) &= \sum_{k=1}^{n} \mathbb{P}(T_0 = k) \\
&= \sum_{k=1}^{n} \beta^\top Q^{k-1}\alpha \\
&= \beta^\top (I - Q^n)(I - Q)^{-1}\alpha \\
&= \beta^\top (I - Q^n)e \\
&= 1 - \beta^\top Q^n e, \qquad n \geqslant 1.
\end{aligned}
$$

Alternatively, using the relation $\alpha = (I - Q)e$ and a telescopic sum, we recover this result as

$$
\begin{aligned}
\mathbb{P}(T_0 \leqslant n) &= \sum_{k=1}^{n} \mathbb{P}(T_0 = k) \\
&= \sum_{k=1}^{n} \beta^\top Q^{k-1}\alpha \\
&= \sum_{k=1}^{n} \beta^\top Q^{k-1}(I - Q)e \\
&= \sum_{k=0}^{n-1} \beta^\top Q^k e - \sum_{k=1}^{n} \beta^\top Q^k e \\
&= \beta^\top e - \beta^\top Q^n e \\
&= 1 - \beta^\top Q^n e, \qquad n \geqslant 1.
\end{aligned}
$$

h) We have

$$\begin{aligned}
\mathbb{P}(T_0 \leqslant n) &= 1 - \sum_{k=1}^{d} \mathbb{P}(X_n = k) \\
&= 1 - \sum_{k=1}^{d} \sum_{i=1}^{d} \beta_i \mathbb{P}(X_n = k \mid X_0 = i) \\
&= 1 - \sum_{k=1}^{d} \sum_{i=1}^{d} \beta_i Q^n_{i,k} \\
&= 1 - \beta^\top Q^n e, \qquad n \geqslant 1.
\end{aligned}$$

Alternatively we could also write

$$\begin{aligned}
\mathbb{P}(T_0 \leqslant n) &= \mathbb{P}(X_n = 0) \\
&= \sum_{i=1}^{d} \beta_i \mathbb{P}(X_n = 0 \mid X_0 = i) \\
&= \sum_{i=1}^{d} \beta_i [P^n]_{i,0} \\
&= \sum_{i=1}^{d} \beta_i [(I - Q^n)e]_i \\
&= \sum_{i=1}^{d} \beta_i - \sum_{i=1}^{d} \beta_i [Q^n e]_i \\
&= 1 - \beta^\top Q^n e, \qquad n \geqslant 1.
\end{aligned}$$

i) We note that T_0 is finite with probability one, since

$$\begin{aligned}
\mathbb{P}(T_0 < \infty) &= \sum_{n=0}^{\infty} \mathbb{P}(T_0 = n) \\
&= \sum_{n=1}^{\infty} \beta^\top Q^{n-1} \alpha \\
&= \beta^\top \sum_{n=0}^{\infty} Q^n \alpha \\
&= \beta^\top (I - Q)^{-1} \alpha \\
&= \beta^\top (I - Q)^{-1} (I - Q) e \\
&= \sum_{i=1}^{d} \beta_i
\end{aligned}$$

$$= \sum_{i=1}^{d} \mathbb{P}(X_0 = i)$$
$$= 1.$$

Since the states $\{1, 2, \ldots, d\}$ are transient, Corollary 6.9 shows that the matrix inverse $(I - sQ)^{-1}$ exists and is given by the series

$$(I - sQ)^{-1} = \sum_{k \geqslant 0} s^k Q^k, \qquad s \in (-1, 1],$$

hence we have

$$\begin{aligned}
G_{T_0}(s) &= \mathbb{E}\left[s^{T_0}\right] \\
&= \sum_{k \geqslant 0} s^k \mathbb{P}(T_0 = k) \\
&= \mathbb{P}(X_0 = 0) + \sum_{k \geqslant 1} s^k \beta^\top Q^{k-1} \alpha \\
&= s \sum_{k \geqslant 0} s^k \beta^\top Q^k \alpha \\
&= s\beta^\top (I - sQ)^{-1} \alpha \\
&= s\beta^\top (I - sQ)^{-1} (I - Q) e.
\end{aligned}$$

We note that

$$\mathbb{P}(T_0 < \infty) = G_{T_0}(1) = \beta^\top (I - Q)^{-1} (I - Q) e = \beta^\top e = 1,$$

hence state ⓪ is reached in finite time with probability one.

j) We have

$$G'_{T_0}(s) = \beta^\top (I - sQ)^{-1} \alpha + s\beta^\top Q (I - sQ)^{-2} \alpha,$$

hence

$$\begin{aligned}
\mathbb{E}[T_0] &= G'_{T_0}(1) \\
&= \beta^\top (I - Q)^{-1} \alpha + \beta^\top Q (I - Q)^{-2} \alpha \\
&= \beta^\top (I - Q)(I - Q)^{-2} \alpha + \beta^\top Q (I - Q)^{-2} \alpha \\
&= \beta^\top (I - Q)^{-2} \alpha \\
&= \beta^\top (I - Q)^{-1} e.
\end{aligned}$$

We also have

$$G''_{T_0}(s) = \beta^\top Q (I - sQ)^{-2} \alpha + \beta^\top Q (I - sQ)^{-2} \alpha + 2s\beta^\top Q^2 (I - sQ)^{-3} \alpha,$$

hence

$$\begin{aligned}\mathbb{E}[T_0(T_0-1)] &= G''_{T_0}(1)\\ &= 2\beta^\top Q(I-Q)^{-2}\alpha + 2\beta^\top Q^2(I-Q)^{-3}\alpha\\ &= 2\beta^\top Q(I-Q)^{-3}\alpha,\\ &= 2\beta^\top Q(I-Q)^{-2}e,\end{aligned}$$

hence

$$\begin{aligned}\mathbb{E}[T_0^2] &= \mathbb{E}[T_0(T_0-1)] + \mathbb{E}[T_0]\\ &= 2\beta^\top Q(I-Q)^{-2}e + \beta^\top(I-Q)^{-1}e\\ &= 2\beta^\top Q(I-Q)^{-2}e + \beta^\top(I-Q)(I-Q)^{-2}e\\ &= \beta^\top(I+Q)(I-Q)^{-2}e.\end{aligned}$$

More generally, by (1.7.7) we could also compute the *factorial moment*

$$\mathbb{E}[T_0(T_0-1)\cdots(T_0-k+1)] = G^{(k)}_{T_0}(1) = k!\beta^\top Q^{k-1}(I-Q)^{-k}e,$$

for all $k \geqslant 1$.

Problem 4.15

a) We have

$$\begin{aligned}&\mathbb{P}((O_0,O_1,O_2)=(c,a,b) \text{ and } (X_0,X_1,X_2)=(1,1,0))\\ &= \mathbb{P}((O_0,O_1,O_2)=(c,a,b) \mid (X_0,X_1,X_2)=(1,1,0))\mathbb{P}((X_0,X_1,X_2)=(1,1,0))\\ &= \mathbb{P}(O_0=c \mid X_0=1)\mathbb{P}(O_1=a \mid X_1=1)\mathbb{P}(O_2=b \mid X_2=0)\mathbb{P}((X_0,X_1,X_2)=(1,1,\\ &= \pi_1 P_{1,1}P_{1,0}m_{1,c}m_{1,a}m_{0,b},\end{aligned}$$

since

$$\mathbb{P}((X_0,X_1,X_2)=(1,1,0)) = \pi_1 P_{1,1}P_{1,0}.$$

b) We have

$$\begin{aligned}&\mathbb{P}((O_0,O_1,O_2)=(c,a,b))\\ &\qquad = \sum_{x,y,z\in\{0,1\}} \mathbb{P}((O_0,O_1,O_2)=(c,a,b) \text{ and } (X_0,X_1,X_2)=(x,y,z))\\ &\qquad = \sum_{x,y,z\in\{0,1\}} \pi_x P_{x,y}P_{y,z}m_{x,c}m_{y,a}m_{z,b}.\end{aligned}$$

c) We have

$$\{X_1=1\} = \{(X_0,X_1,X_2)=(0,1,0)\}\bigcup\{(X_0,X_1,X_2)=(0,1,1)\}$$

$$\bigcup \{(X_0, X_1, X_2) = (1,1,0)\} \bigcup \{(X_0, X_1, X_2) = (1,1,1)\}$$
$$= \bigcup_{x,z\in\{0,1\}} \{(X_0, X_1, X_2) = (x,1,z)\},$$

where the above union is a partition, hence

$$\begin{aligned}
&\mathbb{P}(X_1 = 1 \mid (O_0, O_1, O_2) = (c,a,b)) \\
&= \sum_{x,z\in\{0,1\}} \mathbb{P}((X_0, X_1, X_2) = (x,1,z) \mid (O_0, O_1, O_2) = (c,a,b)) \\
&= \frac{1}{\mathbb{P}((O_0, O_1, O_2) = (c,a,b))} \\
&\times \sum_{x,z\in\{0,1\}} \mathbb{P}((X_0, X_1, X_2) = (x,1,z) \text{ and } (O_0, O_1, O_2) = (c,a,b)) \\
&= \frac{1}{\mathbb{P}((O_0, O_1, O_2) = (c,a,b))} \sum_{x,z\in\{0,1\}} \pi_x P_{x,1} P_{1,z} m_{x,c} m_{1,b} m_{z,c}.
\end{aligned}$$

d) By the result of Question (a) we find

$$\begin{cases}
\mathbb{P}((X_0, X_1, X_2) = (0,0,0) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.00588, \\
\mathbb{P}((X_0, X_1, X_2) = (0,0,1) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.00126, \\
\mathbb{P}((X_0, X_1, X_2) = (0,1,0) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.0101, \\
\mathbb{P}((X_0, X_1, X_2) = (0,1,1) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.00756, \\
\mathbb{P}((X_0, X_1, X_2) = (1,0,0) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.000448, \\
\mathbb{P}((X_0, X_1, X_2) = (1,0,1) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.0000960, \\
\mathbb{P}((X_0, X_1, X_2) = (1,1,0) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.00269, \\
\mathbb{P}((X_0, X_1, X_2) = (1,1,1) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.00202.
\end{cases}$$

From the above computation we deduce that the most likely sample sequence of values for (X_0, X_1, X_2) given the observation $(O_0, O_1, O_2) = (c,a,b)$ is $(X_0, X_1, X_2) = (0,1,0)$, with the probability

$$\mathbb{P}((X_0, X_1, X_2) = (0,1,0) \text{ and } (O_0, O_1, O_2) = (c,a,b)) = 0.0101.$$

e) By the result of Question (b) we find

$$\mathbb{P}((O_0, O_1, O_2) = (c,a,b)) = 0.030028 \simeq 3\%.$$

f) By proceeding as in Question (c) we find

$$\begin{cases} \mathbb{P}(X_0 = 0 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.825, \\ \mathbb{P}(X_0 = 1 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.175, \\ \mathbb{P}(X_1 = 0 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.256, \\ \mathbb{P}(X_1 = 1 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.744, \\ \mathbb{P}(X_2 = 0 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.636, \\ \mathbb{P}(X_2 = 1 \mid (O_0, O_1, O_2) = (c, a, b)) = 0.364. \end{cases}$$

According to the above estimates, the most likely sequence for (X_0, X_1, X_2) given the observation $(O_0, O_1, O_2) = (c, a, b)$ is $(0, 1, 0)$.

g) We find $\hat{\pi} = [\hat{\pi}_0, \hat{\pi}_1] = [0.825, 0.175]$.

h) We find

$$\widehat{P} = \begin{bmatrix} 0.415 & 0.585 \\ \\ 0.482 & 0.518 \end{bmatrix}.$$

i) We find

$$\widehat{M} = \begin{bmatrix} 0.149 & 0.370 & 0.481 \\ 0.580 & 0.284 & 0.136 \end{bmatrix}.$$

j) The estimates of the matrix M obtained from the R code can be plotted as follows:

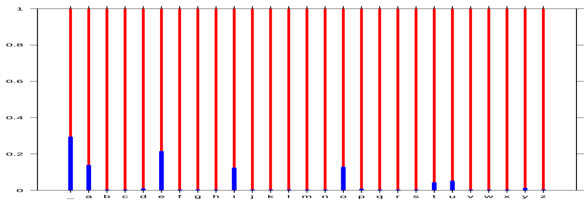


Fig. S.7: Plot of estimatehmmemissionProbs[1,].

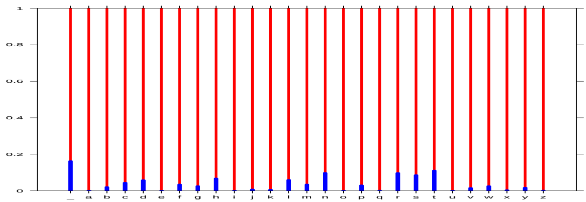


Fig. S.8: Plot of estimatehmmemissionProbs[2,].

From Figures S.7 and S.8 we can infer that the vowels $\{a, e, i, u, o\}$ are more frequently associated to the state ⓪ of the hidden chain $(X_n)_{n\in\mathbb{N}}$.

The vowels “a,e,i,o,u”, together with the spacing character “_” total 93% of emission probabilities from state ⓪, and the combined probabilities of vowels from state ① is only 6.2×10^{-9} %.

Human intervention can be nevertheless required in order to set a probability threshold that can distinguish vowels from consonants, *e.g.* to separate “u” from “t”. The classification effect is enhanced in the following Figure S.9 that plots $\eta \mapsto (M_{0,\eta}/M_{0,"\_"})((M_{1,"\_"} - M_{1,\eta})/M_{1,"\_"})^2$ by combining the information available in the two rows of the emission matrix M, showing that “y” recovers its “semi-vowel” status.

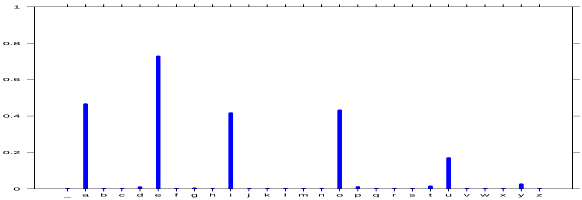


Fig. S.9: Plot of $\eta \mapsto (M_{0,\eta}/M_{0,"\_"})((M_{1,"\_"} - M_{1,\eta})/M_{1,"\_"})^2$.

Note that the graphs of Figures S.7 and S.8 do *not* represent a frequency analysis. A frequency analysis of letters can be represented as the histogram of Figure S.10 using the command

```
text = readChar("brown.txt",nchars=10000)
data <- unlist (strsplit (gsub ("[^a-z]", "_", tolower (text)), ""))
barplot(col = rainbow(30), table(data), cex.names=0.7)
```

with the following output:

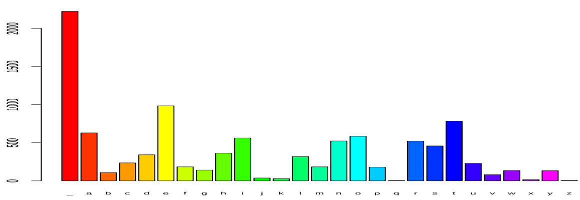


Fig. S.10: Frequency analysis of alphabet letters.

The command

```
estimate$hmm$transProbs[1,]
```

yields the transition probability estimate

$$\widehat{P} = \begin{bmatrix} 0 & 1 \\ 0.1906356 & 0.8093644 \end{bmatrix}$$

for the hidden chain $(X_n)_{n\in\mathbb{N}}$.

Note that $\widehat{P}$ is not the transition matrix of vowels *vs.* consonants. For example of the word “universities” contains eleven letter transitions, including:

five vowel-consonant transitions,
one vowel-vowel transition,
four consonant-vowel transitions,
one consonant-consonant transitions,

which would yield the transition probability estimate

$$\begin{bmatrix} 5/6 & 1/6 \\ 4/5 & 1/5 \end{bmatrix},$$

assuming the alphabet has *already* been partitioned. Such a matrix can be estimated on the whole text, from the following code:

```
x <- unlist (strsplit (gsub ("[^a-z]", "", tolower (text)), ""))
y <- unlist (strsplit (gsub ("[^a,e,i,o,u]", "2", tolower (x)), ""))
z <- as.numeric(noquote(unlist (strsplit (gsub ("[a,e,i,o,u]", "1",y), ""))))
p <- matrix(nrow = 2, ncol = 2, 0)
for (t in 1:(length(z) - 1)) p[z[t], z[t + 1]] <- p[z[t], z[t + 1]] + 1
for (i in 1:2) p[i, ] <- p[i, ] / sum(p[i, ])
```

as

$$\begin{bmatrix} 0.1424749 & 0.8575251 \\ 0.5360502 & 0.4639498 \end{bmatrix},$$

which means that inside the text, a vowel is followed by a consonant for 85.7% of the time, while a consonant is followed by a vowel for 53% of the time.

The Baum-Welch algorithm does more than a simple frequency/transition analysis, as it can estimate the emission probability matrix M, which can be used to partition the alphabet. However, the algorithm is not making a one-to-one association between states $(0, 1)$ of $(X_n)_{n\in\mathbb{N}}$ and letters; the association is only probabilistic and expressed through the estimate $\widehat{M}$ of the emission matrix.

Using a three-state model shows a more definite identification of vowels from state 3 in Figure S.11, and a special weight given to the letters h and t from state ① in Figure S.12.

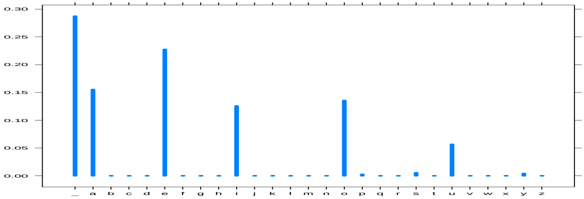


Fig. S.11: Plot of estimatehmmemissionProbs[3,].

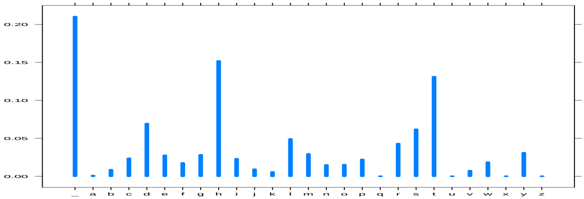


Fig. S.12: Plot of estimatehmmemissionProbs[1,].

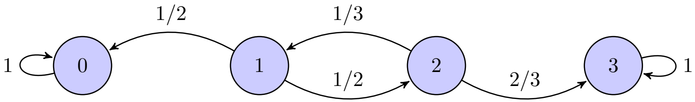
Chapter 5 - First Step Analysis

Exercise 5.1

a) The chain has the transition matrix

$$[\ P_{i,j}\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and the graph



b) Compute $\alpha := \mathbb{P}(T_3 < \infty \mid X_0 = 1)$ and $\beta := \mathbb{P}(T_3 < \infty \mid X_0 = 2)$, where

$$T_3 := \inf\{n \geqslant 0 \ : \ X_n = 3\}.$$

By first step analysis, we have

$$\begin{cases} \alpha = \mathbb{P}(T_3 < \infty \mid X_0 = 1) = \dfrac{1}{2} \times 0 + \dfrac{1}{2} \times \beta \\ \\ \beta = \mathbb{P}(T_3 < \infty \mid X_0 = 2) = \dfrac{1}{3} \times \alpha + \dfrac{2}{3} \times 1, \end{cases}$$

hence

$$\alpha = \mathbb{P}(T_3 < \infty \mid X_0 = 1) = \frac{2}{5} \quad \text{and} \quad \beta = \mathbb{P}(T_3 < \infty \mid X_0 = 2) = \frac{4}{5}.$$

c) By first step analysis, we have

$$\begin{cases} \mathbb{E}[T_{0,3} \mid X_0 = 1] = \dfrac{1}{2}\left(1 + \mathbb{E}[T_{0,3} \mid X_0 = 0]\right) + \dfrac{1}{2}\left(1 + \mathbb{E}[T_{0,3} \mid X_0 = 2]\right) \\ \qquad\qquad\qquad\quad = 1 + \dfrac{1}{2}\mathbb{E}[T_{0,3} \mid X_0 = 2], \\ \\ \mathbb{E}[T_{0,3} \mid X_0 = 2] = 1 + \dfrac{1}{3}\mathbb{E}[T_{0,3} \mid X_0 = 1] + \dfrac{2}{3}\left(1 + \mathbb{E}[T_{0,3} \mid X_0 = 3]\right) \\ \qquad\qquad\qquad\quad = 1 + \dfrac{1}{3}\mathbb{E}[T_{0,3} \mid X_0 = 1], \end{cases}$$

i.e.

$$\begin{cases} \mathbb{E}[T_{0,3} \mid X_0 = 1] = \dfrac{3}{2} + \dfrac{1}{6}\mathbb{E}[T_{0,3} \mid X_0 = 1] \\ \\ \mathbb{E}[T_{0,3} \mid X_0 = 2] = \dfrac{4}{3} + \dfrac{1}{6}\mathbb{E}[T_{0,3} \mid X_0 = 2], \end{cases}$$

hence

$$\begin{cases} \mathbb{E}[T_{0,3} \mid X_0 = 1] = \dfrac{9}{5}, \\ \\ \mathbb{E}[T_{0,3} \mid X_0 = 2] = \dfrac{8}{5}. \end{cases}$$

Exercise 5.2 By interpreting the mean duration between two visits to state (1) as the mean return time $\mu_1(1)$ from state (1) to state (1), we find

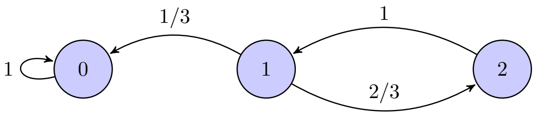
$$\mu_1(1) = 1 + \frac{b}{a} = 1 + \frac{0.8}{0.6} = \frac{7}{3}.$$

from *e.g.* (5.3.6). This mean time can also be recovered by pathwise analysis using the mean $1/a$ of the geometric distribution on $\{1, 2, 3, \ldots\}$ with parameter $a \in (0, 1]$, as

$$(1 - b) \times 1 + b \times \left(1 + \frac{1}{a}\right) = 1 + \frac{b}{a} = \frac{7}{3}.$$

Exercise 5.3

a) The chain has the following graph:



Noting that state ⓪ is absorbing, by first step analysis we have

$$\begin{cases} g_0(0) = 1 \\ \\ g_0(1) = \frac{1}{3} g_0(0) + \frac{2}{3} g_0(2) \\ \\ g_0(2) = g_0(1), \end{cases}$$

which has for solution

$$g_0(0) = g_0(1) = g_0(2) = 1.$$

```
install.packages("igraph");install.packages("markovchain")
library("igraph");library(markovchain)
P<-matrix(c(1,0,0,1/3,0,2/3,0,1,0),nrow=3,byrow=TRUE);
       MC<-new("markovchain",transitionMatrix=P)
graph <- as(MC, "igraph")
plot(graph,vertex.size=50,edge.label.cex=2,edge.label=E(graph)$prob,edge.color='black',
       vertex.color='dodgerblue',vertex.label.cex=3)
hittingProbabilities(object = MC)
  1 2 3
1 1 0.0000000 0.0000000
2 1 0.6666667 0.6666667
3 1 1.0000000 0.6666667
```

b) By first step analysis, we have

$$\begin{cases} h_0(0) = 0 \\ \\ h_0(1) = 1 + \frac{1}{3} h_0(0) + \frac{2}{3} h_0(2) \\ \\ h_0(2) = 1 + h_0(1), \end{cases}$$

which has for solution

$$h_0(0) = 0, \quad h_0(1) = 5, \quad h_0(2) = 6.$$

```
meanAbsorptionTime(object = MC)
5 6
```

Exercise 5.4

a) The boundary conditions are given by

$$f(x,0) = -x \qquad \text{and} \qquad f(0,y) = y, \qquad x, y \geqslant 0.$$

b) The finite difference equation satisfied by $f(x,y)$ is given by

$$f(x,y) = \frac{x}{x+y}(f(x-1,y)-1) + \frac{y}{x+y}(f(x,y-1)+1), \quad x, y \geqslant 1.$$

c) We have

$$\begin{cases} f(1,1) = \frac{1}{2}(f(0,1)-1) + \frac{1}{2}(f(1,0)+1) = 0, \\ f(1,2) = \frac{1}{3}(f(0,2)-1) + \frac{2}{3}(f(1,1)+1) = 1, \\ f(2,2) = \frac{1}{2}(f(1,2)-1) + \frac{1}{2}(f(2,1)+1) = 0, \\ f(1,3) = \frac{1}{4}(f(0,3)-1) + \frac{3}{4}(f(1,2)+1) = 2, \\ f(2,3) = \frac{2}{5}(f(1,3)-1) + \frac{3}{5}(f(2,2)+1) = 1, \\ f(3,3) = \frac{1}{2}(f(2,3)-1) + \frac{1}{2}(f(3,2)+1) = 0. \end{cases}$$

d) We check that $f(x,y) := y - x$ solves the finite difference equation

$$\begin{aligned} & \frac{x}{x+y}(f(x-1,y)-1) + \frac{y}{x+y}(f(x,y-1)+1) \\ & = \frac{x}{x+y}(y-(x-1)-1) + \frac{y}{x+y}(y-1-x+1) \\ & = \frac{x}{x+y}(y-x) + \frac{y}{x+y}(y-x) \\ & = y - x \\ & = f(x,y), \end{aligned}$$

with the correct boundary conditions.

Exercise 5.5

a) By first step analysis, we have

$$\begin{cases} p_2(0) = 0.2 \times p_2(1) + 0.8 \\ p_2(1) = 0.1 \times p_2(1) + 0.3 \times p_2(0) + 0.6 \times p_2(3) \\ p_2(2) = 0.5 \times p_2(0) + 0.5 \times p_2(3) \\ p_2(3) = 0, \end{cases}$$

hence

$$\begin{cases} p_2(0) = 0.2 \times p_2(1) + 0.8 \\ \\ p_2(1) = \dfrac{p_2(0)}{3} \\ \\ p_2(2) = \dfrac{p_2(0)}{2} \\ \\ p_2(3) = 0, \end{cases}$$

and

$$\begin{cases} p_2(0) = \dfrac{6}{7} \\ \\ p_2(1) = \dfrac{2}{7} \\ \\ p_2(2) = \dfrac{3}{7} \\ \\ p_2(3) = 0. \end{cases}$$

We note that

$$\mathbb{P}(T_2^r = \infty \mid X_0 = 2) = 1 - p_2(2) = \frac{4}{7} \geqslant \mathbb{P}(X_1 = 3 \mid X_0 = 2) = 0.5,$$

as seen in (6.3.6).

b) By first step analysis, we have

$$\begin{cases} p_1(0) = 0.2 + 0.8 \times p_1(2) \\ p_1(1) = 0.1 + 0.3 \times p_1(0) + 0.6 \times p_1(3) \\ p_1(2) = 0.5 \times p_1(0) + 0.5 \times p_1(3) \\ p_1(3) = 0, \end{cases}$$

hence

$$\begin{cases} p_1(0) = 0.2 + 0.8 \times p_1(2) = 0.2 + 0.4 \times p_1(0) \\ p_1(1) = 0.1 + 0.3 \times p_1(0) \\ p_1(2) = 0.5 \times p_1(0) \\ p_1(3) = 0, \end{cases}$$

and

$$\begin{cases} p_1(0) = \frac{1}{3} \\ \\ p_1(1) = 0.2 \\ \\ p_1(2) = \frac{1}{6} \\ \\ p_1(3) = 0. \end{cases}$$

We note that

$$\mathbb{P}(T_1^r = \infty \mid X_0 = 1) = 1 - p_1(1) = 0.8 \geqslant \mathbb{P}(X_1 = 3 \mid X_0 = 1) = 0.6,$$

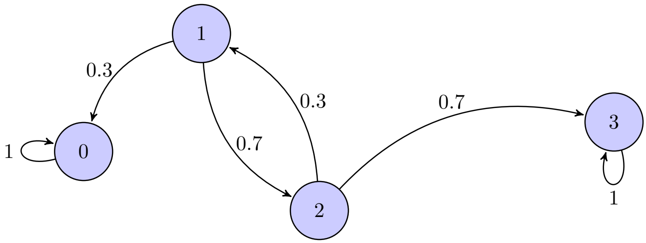
as seen in (6.3.7).

c) We have

$$\begin{cases} \mu_2(0) = 0.2 \times (1 + \mu_2(1)) + 0.8 \\ \mu_2(1) = 0.3 \times (1 + \mu_2(0)) + 0.1 \times (1 + \mu_2(1)) + 0.6 \times (1 + \mu_2(3)) \\ \mu_2(2) = 0.5 \times (1 + \mu_2(0)) + 0.5 \times (1 + \mu_2(3)) \\ \mu_2(3) = +\infty, \end{cases}$$

hence $\mu_0(1) = \mu_2(1) = \mu_2(2) = \mu_3(2) = +\infty$, as can be checked from $\mathbb{P}(T_2^r = \infty \mid X_0 = k) = 1 - p_2(k) > 0$, $k = 0, 1, 2, 3$.

Exercise 5.6 This exercise is a particular case of the Example of Section 5.1, by taking $a := 0.3$, $b := 0$, $c := 0.7$, $d := 0$, $\alpha := 0$, $\beta := 0.3$, $\gamma := 0$, $\eta := 0.7$, and the chain has the graph



a) States ⓪ and ③ are absorbing.

b) By first step analysis we find the equations

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = 0.3 + 0.7 g_0(2) \\ g_0(2) = 0.3 g_0(1) \\ g_0(3) = 0, \end{cases}$$

hence

$$g_0(1) = \frac{0.3}{1 - 0.7 \times 0.3} \quad \text{and} \quad g_0(2) = \frac{0.3 \times 0.3}{1 - 0.7 \times 0.3},$$

which is consistent with the answer

$$\begin{cases} g_0(1) = \dfrac{a}{1 - \beta c} \\ g_0(2) = \dfrac{a\beta}{1 - \beta c} \end{cases}$$

obtained from (5.1.11). Note that this chain is actually a particular case of the gambling process of Chapter 2 with $S = 3$, hence by (2.2.12) we can also write

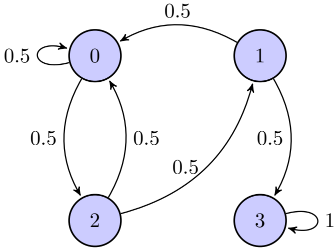
$$g_0(1) = \mathbb{P}(R_A \mid X_0 = 1) = \frac{(3/7) - (3/7)^3}{1 - (3/7)^3} = \frac{(7/3)^{3-1} - 1}{(7/3)^3 - 1} \simeq 0.3797.$$

c) By first step analysis we find the equations

$$\begin{cases} h_1(0) = 1 + h_1(0) \\ h_1(1) = 0 \\ h_1(2) = 0.3(1 + h_1(1)) + 0.7(1 + h_1(3)) = 1 + 0.7 h_1(3) \\ h_1(3) = 1 + h_1(3), \end{cases}$$

which admit the only solution $h_1(1) = 0$ and $h_1(0) = h_1(2) = h_1(2) = +\infty$.

Exercise 5.7 We observe that state ③ is absorbing:



Let

$$h_3(k) := \mathbb{E}[T_3 \mid X_0 = k]$$

denote the mean (hitting) time needed to reach state ③ starting from state $k = 0, 1, 2, 3$. We have

$$\begin{cases} h_3(0) = 1 + \dfrac{1}{2} h_3(0) + \dfrac{1}{2} h_3(2) \\ h_3(1) = 1 + \dfrac{1}{2} h_3(0) \\ h_3(2) = 1 + \dfrac{1}{2} h_3(0) + \dfrac{1}{2} h_3(1) \\ h_3(3) = 0, \end{cases}$$

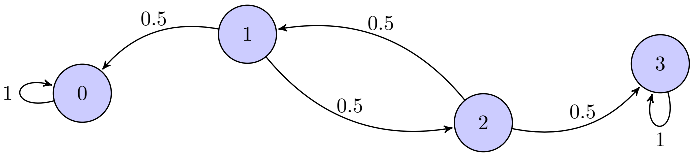
which yields

$$h_3(3) = 0, \qquad h_3(1) = 8, \qquad h_3(2) = 12, \qquad h_3(0) = 14.$$

We check that $h_3(3) < h_3(1) < h_3(2) < h_3(0)$, as can be expected from the graph.

Exercise 5.8 This exercise has some similarities with the gambling problem of Chapter 2, with the maze problems of Section 5.3, and with Exercise 5.12 below.

a) The chain has the following graph:



Note that this process is in fact a fair gambling process on the state space $\{0,1,2,3\}$.

b) Since the states ⓪ and ③ are absorbing, by first step analysis we have

$$\begin{cases} g_0(0)=1 \\ g_0(1)=\frac{1}{2}+\frac{1}{2}g_0(2) \\ g_0(2)=\frac{1}{2}g_0(1) \\ g_0(3)=0, \end{cases}$$

which has for solution

$$g_0(0)=1, \quad g_0(1)=\frac{2}{3}, \quad g_0(2)=\frac{1}{3}, \quad g_0(3)=0.$$

c) By first step analysis, we have

$$\begin{cases} h_{0,3}(0)=0 \\ h_{0,3}(1)=1+\frac{1}{2}h_{0,3}(2) \\ h_{0,3}(2)=1+\frac{1}{2}h_{0,3}(1) \\ h_{0,3}(3)=0, \end{cases}$$

which has for solution

$$h_{0,3}(0)=0, \quad h_{0,3}(1)=2, \quad h_{0,3}(2)=2, \quad h_{0,3}(3)=0.$$

Exercise 5.9 Using the law of total probability with respect to the partition $\Omega=\bigcup_{m\geqslant 0}\{T_A=m\}$, we have, provided that $i_0\in A$,

$$\begin{aligned}
&\mathbb{P}(X_{T_A+n}=j \mid X_{T_A}=i_0,\ldots,X_0=i_{T_A} \text{ and } T_A<+\infty)\\
&=\sum_{m\geqslant 0}\mathbb{P}(X_{T_A+n}=j \text{ and } \ T_A=m \mid X_{T_A}=i_0,\ldots,X_0=i_{T_A} \text{ and } T_A<+\infty)\\
&=\sum_{m\geqslant 0}\frac{\mathbb{P}(X_{T_A+n}=j,\ T_A=m,\ X_{T_A}=i_0,\ldots,X_0=i_{T_A} \text{ and } T_A<+\infty)}{\mathbb{P}(X_{T_A}=i_0,\ldots,X_0=i_{T_A} \text{ and } T_A<+\infty)}\\
&=\sum_{m\geqslant 0}\frac{\mathbb{P}(X_{n+m}=j,\ T_A=m,\ X_m=i_0,\ldots,X_0=i_m)}{\mathbb{P}(X_{T_A}=i_0,\ldots,X_0=i_{T_A} \text{ and } T_A<+\infty)}
\end{aligned}$$

$$= \sum_{m\geqslant 0} \frac{\mathbb{P}(X_{n+m} = j,\ X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A,\ X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \sum_{m\geqslant 0} \frac{\mathbb{P}(X_{n+m} = j,\ X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}$$

$$\times \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \sum_{m\geqslant 0} \mathbb{P}(X_{n+m} = j \mid X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)$$

$$\times \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \sum_{m\geqslant 0} \mathbb{P}(X_{n+m} = j \mid X_m = i_0 \text{ and } X_m \in A)$$

$$\times \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m,\ X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \sum_{m\geqslant 0} \mathbb{P}(X_{n+m} = j \mid X_m = i_0)$$

$$\times \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m \text{ and } X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin A)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \mathbb{P}(X_n = j \mid X_0 = i_0) \sum_{m\geqslant 0} \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m \text{ and } X_m \in A, X_{m-1} \notin A, \ldots, X_0 \notin}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \mathbb{P}(X_n = j \mid X_0 = i_0) \sum_{m\geqslant 0} \frac{\mathbb{P}(X_m = i_0, \ldots, X_0 = i_m \text{ and } T_A = m)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \mathbb{P}(X_n = j \mid X_0 = i_0) \sum_{m\geqslant 0} \frac{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A = m)}{\mathbb{P}(X_{T_A} = i_0, \ldots, X_0 = i_{T_A} \text{ and } T_A < +\infty)}$$

$$= \mathbb{P}(X_n = j \mid X_0 = i_0),$$

for all $n, m \geqslant 0$, $j \in \mathbb{S}$, and $(i_k)_{k\in\mathbb{N}} \subset \mathbb{S}$.

Exercise 5.10

a) Letting $f(k) := \mathbb{P}(T_0 < \infty \mid S_0 = k)$ we have the boundary condition $f(0) = 1$ and by first step analysis we find that $f(k)$ satisfies

$$f(k) = pf(k+1) + qf(k-1), \qquad k \geqslant 1,$$

which is (2.2.6), and has the general solution

$$f(k) = C_1 + C_2 r^k, \qquad k \in \mathbb{N}, \tag{S.35}$$

where $r = q/p$, by (2.2.17), see also the command

RSolve[f[k]=pf[k+1]+(1-p)f[k-1],f[k],k].

i) In case $q \geqslant p$, $f(k)$ would tend to (positive or negative) infinity if $C_2 \neq 0$, hence we should have $C_2 = 0$, and $C_1 = f(0) = 1$, showing that $f(k) = 1$ for all $k \in \mathbb{N}$.
ii) In case $q < p$, the probability of hitting $\textcircled{0}$ in finite time after starting from state $\textcircled{k}$ becomes 0 in the limit as k tends to infinity, *i.e.* we have

$$\lim_{k\to\infty} f(k) = \lim_{k\to\infty} \mathbb{P}(T_0 < \infty \mid S_0 = k) = 0, \qquad k \in \mathbb{N},$$

which shows that $C_1 = 0$.

On the other hand, the condition $f(0) = 1$ yields $C_2 = 1$, hence we find $f(k) = (q/p)^k$ for all $k \geqslant 0$.

Alternatively, denoting by T_k the hitting time of state $\textcircled{k}$ we could show the decomposition

$$\begin{aligned} f(k+1) &= \mathbb{P}(T_0 < \infty \mid S_0 = k+1) \\ &= \mathbb{P}(T_k < \infty \mid S_0 = k+1) \times \mathbb{P}(T_0 < \infty \mid S_0 = k) \\ &= \mathbb{P}(T_0 < \infty \mid S_0 = 1) \times \mathbb{P}(T_0 < \infty \mid S_0 = k) \\ &= f(1) \times f(k), \qquad k \in \mathbb{N}, \end{aligned}$$

based on the Markov property, *i.e.* the probability of going down from $k+1$ to 0 in a finite time should be the product of the probability to go down one step in finite time, multiplied by the probability of going down k steps in a finite time. This shows that

$$f(k) = f(0)(f(1))^k = (f(1))^k, \qquad k \geqslant 0,$$

hence we have $C_1 = 0$, $C_2 = 1$, and $f(1) = r = q/p$, which yields

$$f(k) = \left(\frac{q}{p}\right)^k, \qquad k \in \mathbb{N},$$

and recovers (2.2.14) and (3.4.16), cf. also (S.25) and (S.26).

b) Letting $h(k) := \mathbb{E}[T_0 \mid S_0 = k]$ we have the boundary condition $h(0) = 0$ and by first step analysis we find that $h(k)$ satisfies

$$\begin{aligned} h(k) &= p(1 + h(k+1)) + q(1 + h(k-1)) \\ &= 1 + ph(k+1) + qh(k-1), \qquad k \geqslant 1, \end{aligned}$$

which is (2.3.7a) and has a general solution of the form

$$h(k) = C_1 + C_2 r^k + \frac{1}{q-p} k, \qquad k \in \mathbb{N}, \tag{S.36}$$

by (2.3.10). Next, we note that by the Markov property we should have the decomposition

$$\begin{aligned} h(k+1) &= \mathbb{E}[T_0 \mid S_0 = k+1] \\ &= \mathbb{E}[T_0 \mid S_0 = 1] + \mathbb{E}[T_0 \mid S_0 = k] \\ &= h(1) + h(k), \qquad k \in \mathbb{N}, \end{aligned}$$

i.e. the mean time to go down from $k+1$ to 0 should be the sum of the mean time needed to go down one step plus the mean time needed to go down k steps. This shows that

$$h(k) = h(0) + kh(1) = kh(1),$$

hence by (S.36) we have $C_1 = C_2 = 0$, $h(1) = 1/(q-p)$, and

$$h(k) = \frac{k}{q-p}, \qquad k \in \mathbb{N}.$$

In case $q \leqslant p$ the above argument would yield a negative value for $h(k)$, which is impossible, and $h(k)$ has to be infinite for all $k \geqslant 1$.

Exercise 5.11 First, we take a look at the complexity of the problem. Starting from ⓪ there are multiple ways to reach state ⑬ without reaching ⑪ or ⑫. For example:

$13 = 3+4+1+5$, <u>or</u> $13 = 1+6+3+3$, <u>or</u> $13 = 1+1+2+1+3+1+4$,

etc. Clearly, it would be difficult to enumerate all such possibilities, for this reason we use the framework of Markov chains.

We denote by X_n the cumulative sum of dice outcomes after n rounds, and choose to model it as a Markov chain with n as a time index. We can represent X_n as

$$X_n = \sum_{k=1}^{n} \xi_k, \qquad n \geqslant 0,$$

where $(\xi_k)_{k \geqslant 1}$ is a family of independent random variables uniformly distributed over $\{1, 2, 3, 4, 5, 6\}$. The process $(X_n)_{n \geqslant 0}$ is a Markov chain since given the history of $(X_k)_{k=0,1,\ldots,n}$ up to time n, the value

$$X_{n+1} = X_n + \xi_{n+1}$$

depends only on X_n and on ξ_{n+1} which is independent of $X_0, X_1, \ldots, X_n$. The process $(X_n)_{n\geqslant 0}$ is actually a random walk with independent increments $\xi_1, \xi_2, \ldots$.

The chain $(X_n)_{n\geqslant 0}$ has the transition matrix

$$[\,P_{i,j}\,]_{i,j\in\mathbb{N}} = \begin{bmatrix} 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Letting $A := \{11, 12, 13, 14, 15, 16\}$, we are looking at the probability

$$g_0 := \mathbb{P}(X_{T_A} = 13 \mid X_0 = 0)$$

of hitting the set A through state 13 after starting from state ⓪. More generally, letting

$$g_k := \mathbb{P}(X_{T_A} = 13 \mid X_0 = k)$$

denote the probability of hitting the set A *through state* ⑬ after starting from state $k \geqslant 0$, we have $g_k = 0$ for all $k \geqslant 14$, and for $k = 8, 9, 10, 11, 12, 13$ the probabilities g_k are easily computed as

$$\begin{aligned}
g_{13} &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 13) = 1, \\
g_{12} &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 12) = 0, \\
g_{11} &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 11) = 0, \\
g_{10} &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 10) = \frac{1}{6}, \\
g_9 &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 9) = \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} = \frac{7}{36}, \\
g_8 &= \mathbb{P}(X_{T_A} = 13 \mid X_0 = 8) = \frac{1}{6} + 2 \times \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{49}{216},
\end{aligned}$$

where for the computation of g_8 we note that there are only 4 ways to reach state ⑬ from state ⑧ without hitting states ⑪ or ⑫, by the four combinations

$$13 = 8 + 5, \qquad 13 = 8 + 1 + 4, \qquad 13 = 8 + 2 + 3 \qquad 13 = 8 + 1 + 1 + 3.$$

Clearly, things become easily complicated for $k \leqslant 7$, and we will rely on first step analysis to continue the calculation. We have

$$g_k = \frac{1}{6} \sum_{i=1}^{6} g_{k+i}, \qquad k \in \mathbb{N},$$

i.e.

$$\begin{aligned}
g_{10} &= \frac{1}{6} g_{13} \\
g_9 &= \frac{1}{6}(g_{10} + g_{13}) \\
g_8 &= \frac{1}{6}(g_9 + g_{10} + g_{13}) \\
g_7 &= \frac{1}{6}(g_8 + g_9 + g_{10} + g_{13}) \\
g_6 &= \frac{1}{6}(g_7 + g_8 + g_9 + g_{10}) \\
g_5 &= \frac{1}{6}(g_6 + g_7 + g_8 + g_9 + g_{10}) \\
g_4 &= \frac{1}{6}(g_5 + g_6 + g_7 + g_8 + g_9 + g_{10}) \\
g_3 &= \frac{1}{6}(g_4 + g_5 + g_6 + g_7 + g_8 + g_9) \\
g_2 &= \frac{1}{6}(g_3 + g_4 + g_5 + g_6 + g_7 + g_8) \\
g_1 &= \frac{1}{6}(g_2 + g_3 + g_4 + g_5 + g_6 + g_7) \\
g_0 &= \frac{1}{6}(g_1 + g_2 + g_3 + g_4 + g_5 + g_6).
\end{aligned}$$

In order to solve this system of equations we rewrite it as

$$\begin{cases} g_0 - g_1 = \frac{1}{6}(g_1 - g_7) \\ g_1 - g_2 = \frac{1}{6}(g_2 - g_8) \\ g_2 - g_3 = \frac{1}{6}(g_3 - g_9) \\ g_3 - g_4 = \frac{1}{6}(g_4 - g_{10}) \\ g_4 - g_5 = \frac{1}{6}g_5 \\ g_5 - g_6 = \frac{1}{6}g_6 \\ g_6 - g_7 = \frac{1}{6}(g_7 - 1) \\ g_7 - g_8 = \frac{1}{6}g_8 \\ g_8 - g_9 = \frac{1}{6}g_9 \\ g_9 - g_{10} = \frac{1}{6}g_{10} \\ g_{10} = \frac{1}{6}g_{13}, \end{cases} \quad \text{or} \quad \begin{cases} g_0 = \frac{7^{10} - 7^6 \times 6^4 - 4 \times 7^3 \times 6^6}{6^{11}} \\ g_1 = \frac{7^9 - 7^5 \times 6^4 - 3 \times 7^2 \times 6^6}{6^{10}} \\ g_2 = \frac{7^8 - 7^4 \times 6^4 - 2 \times 7 \times 6^6}{6^9} \\ g_3 = \frac{7^7 - 7^3 \times 6^4 - 6^6}{6^8} \\ g_4 = \frac{7^6 - 7^2 \times 6^4}{6^7} \\ g_5 = \frac{7^5 - 7 \times 6^4}{6^6} \\ g_6 = \frac{7^4}{6^5} - \frac{1}{6} = \frac{7^4 - 6^4}{6^5} \\ g_7 = \frac{7^3}{6^4} \\ g_8 = \frac{7^2}{6^3} \\ g_9 = \frac{7}{6^2} \\ g_{10} = \frac{1}{6}, \end{cases}$$

i.e.

$$g_0 = \frac{7^{10} - 7^6 \times 6^4 - 4 \times 7^3 \times 6^6}{6^{11}} \simeq 0.181892636.$$

The total number of paths leading to 13 in the above fashion is 928. More generally, the numbers of such paths starting from $k = 0, 1, \ldots, 10$ can be computed by induction according to the following table:

k	0	1	2	3	4	5	6	7	8	9	10
Number of paths	928	492	248	125	63	32	16	8	4	2	1

Table 14.1: Number of paths starting from $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

In Figure S.13 we enumerate the corresponding possible paths.

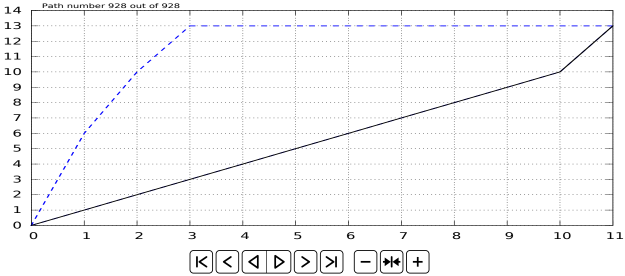


Fig. S.13: Graph of 928 paths leading from 0 to state 13.*

In addition, we could compute the number of paths leading to 13 in k rounds with $k = 3,4,5,6,7,8,9,10,11$ as in the following table:

k	3	4	5	6	7	8	9	10	11
Number of paths	18	88	191	246	209	120	45	10	1

Table 14.2: Number of paths starting taking k rounds, $k = 3,4,5,6,7,8,9,10,11$.

This would also allow us to recover the value of g_0 as

$$\begin{aligned} g_0 &= \frac{18}{6^3} + \frac{88}{6^4} + \frac{191}{6^5} + \frac{246}{6^6} + \frac{209}{6^7} + \frac{120}{6^8} + \frac{45}{6^9} + \frac{10}{6^{10}} + \frac{1}{6^{11}} \\ &= \frac{65990113}{362797056} \simeq 0.181892636, \end{aligned}$$

see *e.g. here*.

Exercise 5.12

a) The transition matrix is given by

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ q & 0 & p & 0 & 0 & 0 \\ 0 & q & 0 & p & 0 & 0 \\ 0 & 0 & q & 0 & p & 0 \\ 0 & 0 & 0 & q & 0 & p \\ \times & \times & \times & \times & \times & \times \end{bmatrix}.$$

* Animated figure (works in Acrobat Reader).

The information contained in the first and last lines of the matrix is not needed here because they have no influence on the result. We have $g(0) = 0$, $g(5) = 1$, and

$$g(k) = q \times g(k-1) + p \times g(k+1), \qquad 1 \leqslant k \leqslant 4. \tag{S.37}$$

b) When $p = q = 1/2$ the probability that starting from state $\textcircled{k}$ the fish finds the food before getting shocked is obtained by solving Equation (S.37) rewritten as

$$g(k) = \frac{1}{2} \times g(k-1) + \frac{1}{2} \times g(k+1), \qquad 1 \leqslant k \leqslant 4.$$

Trying a solution of the form $g(k) = C_1 + kC_2$ under the boundary conditions $g(0) = 0$ and $g(5) = 1$, shows that $C_1 = 0$ and $C_2 = 1/5$, which yields

$$g(k) = \frac{k}{5}, \qquad k = 0, 1, \ldots, 5.$$

Exercise 5.13

a) The transition matrix is given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 \cdots \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \cdots \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \cdots \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \cdots \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \ddots \end{bmatrix}.$$

Note that an arbitrary choice has been made for the first line (*i.e.* state $\textcircled{0}$ is absorbing), however other choices would not change the answer to the question.

b) We have

$$h_0(m) = \sum_{k=0}^{m-1} \frac{1}{m}(1 + h_0(k)) = 1 + \frac{1}{m} \sum_{k=0}^{m-1} h_0(k), \qquad m \geqslant 1,$$

and $h_0(0) = 0$, $h_0(1) = 1$.

c) We have

$$h_0(m) = 1 + \frac{1}{m} \sum_{k=0}^{m-1} h_0(k)$$

$$
\begin{aligned}
&= 1 + \frac{1}{m} h_0(m-1) + \frac{m-1}{m} \sum_{k=0}^{m-2} \frac{h_0(k)}{m-1} \\
&= 1 + \frac{1}{m} h_0(m-1) - \frac{m-1}{m} + \frac{m-1}{m} \left(1 + \sum_{k=0}^{m-2} \frac{h_0(k)}{m-1} \right) \\
&= 1 + \frac{1}{m} h_0(m-1) - \frac{m-1}{m} + \frac{m-1}{m} h_0(m-1) \\
&= 1 - \frac{m-1}{m} + \frac{1}{m} h_0(m-1) + \frac{m-1}{m} h_0(m-1) \\
&= h_0(m-1) + \frac{1}{m}, \qquad m \geqslant 1,
\end{aligned}
$$

hence

$$
\begin{aligned}
h_0(m) &= h_0(m-1) + \frac{1}{m} \\
&= h_0(m-2) + \frac{1}{m-1} + \frac{1}{m} = \sum_{k=1}^{m} \frac{1}{k}, \qquad m \geqslant 1.
\end{aligned}
$$

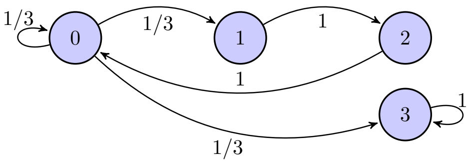
We can also note that from the Markov property, we have

$$
h_0(m) = h_{m-1}(m) + h_0(m-1),
$$

where the mean time $h_{m-1}(m)$ from state (m) to state $\boxed{m-1}$ is equal to $1/m$ since state $\boxed{m-1}$ is always reached in one step from state (m), with probability $1/m$.

Exercise 5.14

a) Assuming that it takes one day per state transition, the graph of the chain can be drawn as follows:



In this model, state (0) represents the tower, states (1) and (2) correspond to the path returning to the tower in three steps through exit A, the returning loop to state (0) accounts for the time to return to the tower through exit B, and state (3) represents the outside.

b) We have

$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

c) By first step analysis, the average time $h_3(i)$ from state ⓘ to state ③ solves

$$\begin{cases} h_3(0) = \frac{1}{3}(1 + h_3(0)) + \frac{1}{3}(1 + h_3(1)) + \frac{1}{3} \\ h_3(1) = 1 + h_3(2) \\ h_3(2) = 1 + h_3(0) \\ h_3(3) = 0, \end{cases}$$

which yields

$$h_3(0) = \frac{1}{3}(1 + h_3(0)) + \frac{1}{3}(3 + h_3(0)) + \frac{1}{3},$$

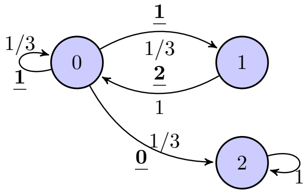
i.e. $h_3(0) = 5$, *i.e.* 4 times steps on average to reach the exit, plus one time step from the exit to the outside. The result $h_3(0) = 4$ can also be obtained directly by writing the equations

$$\begin{cases} h_3(0) = \frac{1}{3}(1 + h_3(0)) + \frac{1}{3}(1 + h_3(1)) + \frac{1}{3} \times 0 \\ h_3(1) = 1 + h_3(2) \\ h_3(2) = 1 + h_3(0). \end{cases}$$

The difficulty in this exercise is that the Markov chain itself is not specified and one had to first design a suitable Markov model, sometimes involving a Markov chain with weighted links. On the other hand, the equation $h = \mathbf{1} + Ph$ *cannot* be directly used when the links have different weights starting from state ①.

Alternative solutions:

(i) The problem can be solved in a simpler way with only 3 states, by adding a weight corresponding to the travel time (underlined) on each link:



Here, the average time $h_2(i)$ from state ⓘ to state ② solves

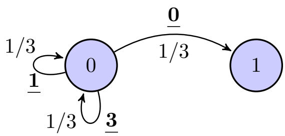
$$\begin{cases} h_2(0) = \frac{1}{3}(\underline{1} + h_2(0)) + \frac{1}{3}(\underline{1} + h_2(1)) + \frac{1}{3} \times \underline{0} \\ h_2(1) = \underline{2} + h_2(0) \\ h_2(2) = 0, \end{cases}$$

which yields

$$h_2(0) = \frac{1}{3}(\underline{1} + h_2(0)) + \frac{1}{3}(\underline{3} + h_2(0)) + \frac{1}{3} \times \underline{0},$$

i.e. $h_2(0) = 4$, which is consistent with the answer to Question (c) above since the time from the exit to the outside is here taken equal to 0.

(ii) The answer can even be further simplified using the following graph, which no longer uses the notion of Markov chain, and in which each link is weighted by a travel time:

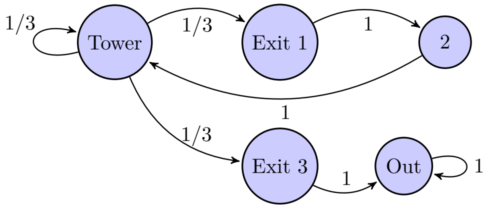


By first step analysis, the average time $t_1(0)$ to travel from state ⓪ to state ① is directly given by

$$t_1(0) = \frac{1}{3}(\underline{1} + t_1(0)) + \frac{1}{3}(\underline{3} + t_1(0)) + \frac{1}{3} \times \underline{0},$$

i.e. $t_1(0) = 4$.

(iii) The problem could also be solved using 4 transient states and one absorbing state, based on the following graph:



Exercise 5.15 The average time t spent inside the maze can be quickly computed by the following first step analysis using weighted links:

$$t = \frac{1}{2} \times (t+3) + \frac{1}{6} \times 2 + \frac{2}{6} \times (t+5),$$

which yields $t = 21$. We refer to Exercise 5.14 and its solution for a more detailed analysis of a similar problem.

Exercise 5.16

a) We have

$$(\pi_0, \pi_1) = \left(\frac{b}{a+b}, \frac{a}{a+b} \right).$$

b) We have

$$\mu_0(0) = 1 + \frac{a}{b}, \quad \mu_1(1) = 1 + \frac{b}{a}, \quad h_0(1) = \frac{1}{b}, \quad h_1(0) = \frac{1}{a}.$$

c) We have

$$\begin{aligned}
\mathbb{E}[\tau - 1 \mid X_0 = 0] &= a\mu_1(1) + (1-a)\mu_0(0) \\
&= a\left(1 + \frac{b}{a}\right) + (1-a)\left(1 + \frac{a}{b}\right) \\
&= (1 + b - a)\frac{a+b}{b} \\
&= \frac{1+b-a}{\pi_0},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\tau - 1 \mid X_0 = 1] &= (1-b)\mu_1(1) + b\mu_0(0) \\
&= (1-b)\left(1 + \frac{b}{a}\right) + b\left(1 + \frac{a}{b}\right) \\
&= (1 + a - b)\frac{a+b}{a}
\end{aligned}$$

$$= \frac{1+a-b}{\pi_1}.$$

d) We have

$$\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=1\}} \,\middle|\, X_0 = 1\right] = b(\mu_0(0)-1) + (1-b) = 1+a-b,$$

$$\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=1\}} \,\middle|\, X_0 = 0\right] = a + (1-a)(\mu_0(0)-1) = a + (1-a)\frac{a}{b} = (1+b-a)\frac{\pi_1}{\pi_0},$$

$$\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=0\}} \,\middle|\, X_0 = 1\right] = b + (1-b)(\mu_1(1)-1) = b + (1-b)\frac{b}{a} = (1+a-b)\frac{\pi_0}{\pi_1},$$

$$\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=0\}} \,\middle|\, X_0 = 0\right] = a(\mu_1(1)-1) + (1-a) = 1+b-a.$$

e) We note that

$$\begin{cases} \mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=1\}} \,\middle|\, X_0 = 1\right] = \mathbb{E}[\tau - 1 \mid X_0 = 1]\pi_1, \\ \mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=1\}} \,\middle|\, X_0 = 0\right] = \mathbb{E}[\tau - 1 \mid X_0 = 0]\pi_1, \\ \mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=0\}} \,\middle|\, X_0 = 1\right] = \mathbb{E}[\tau - 1 \mid X_0 = 1]\pi_0, \\ \mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=0\}} \,\middle|\, X_0 = 0\right] = \mathbb{E}[\tau - 1 \mid X_0 = 0]\pi_0, \end{cases}$$

hence for any initial distribution $(\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1))$ we have

$$\begin{aligned} &\frac{\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=i\}}\right]}{\mathbb{E}[\tau-1]} \\ &= \frac{\mathbb{P}(X_0=0)\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=i\}} \,\middle|\, X_0=0\right] + \mathbb{P}(X_0=1)\mathbb{E}\left[\sum_{l=1}^{\tau-1} \mathbb{1}_{\{X_l=i\}} \,\middle|\, X_0=1\right]}{\mathbb{E}[\tau-1]} \\ &= \frac{\mathbb{P}(X_0=0)\mathbb{E}[\tau-1 \mid X_0=0]\pi_i + \mathbb{P}(X_0=1)\mathbb{E}[\tau-1 \mid X_0=0]\pi_i}{\mathbb{E}[\tau-1]} \\ &= \pi_i \mathbb{P}(X_0=0) + \pi_i \mathbb{P}(X_0=1) \\ &= \pi_i, \qquad i = 0, 1. \end{aligned}$$

Exercise 5.17 By first step analysis, we have

$$\begin{cases} V_a(a) = 0 \\ V_a(b) = -1 + \dfrac{2}{3} V_a(a) + \dfrac{1}{3} V_a(c) \\ V_a(c) = 2 + V_a(b), \end{cases}$$

which has for solution $V_a(a) = 0$, $V_a(b) = -1/2$, $V_a(c) = 3/2$.

```
install.packages("igraph");install.packages("markovchain")
library("igraph");library(markovchain); statenames <- c("a", "b", "c")
P<-matrix(c(1,0,0,2/3,0,1/3,0,1,0),nrow=3,byrow=TRUE, dimnames =
      list(statenames,statenames));
MC <-new("markovchain",transitionMatrix=P); graph <- as(MC, "igraph")
plot(graph,vertex.size=50,edge.label.cex=2,edge.label=E(graph)$prob,edge.color='black',
      vertex.color='dodgerblue',vertex.label.cex=3)
expectedRewards(MC,100,c(0,-1,2))
0.0 -0.5 1.5
meanAbsorptionTime(object = MC)
  b c
a 2 3
```

Exercise 5.18 By first step analysis, we have

$$\begin{cases} V(1) = -2 + (1-p)\gamma V(1) + p\gamma V(2) \\ V(2) = 3 + (1-q)\gamma V(1) + q\gamma V(3) \\ V(3) = 1 + \gamma V(3) \end{cases}$$

hence

$$\begin{cases} V(1) = -2 + (1-p)\gamma V(1) + p\gamma V(2) \\ V(2) = 3 + (1-q)\gamma V(1) + \dfrac{q\gamma}{1-\gamma} \\ V(3) = \dfrac{1}{1-\gamma} = \displaystyle\sum_{n\geqslant 0} \gamma^n, \end{cases}$$

and

$$\begin{cases} V(1) = \dfrac{(3p\gamma - 2)(1-\gamma) + pq\gamma^2}{(1-(1-p)\gamma - (1-q)p\gamma^2)(1-\gamma)} \\ V(2) = 3 + \dfrac{q\gamma}{1-\gamma} + \dfrac{(1-q)((3p\gamma-2)(\gamma-\gamma^2) + pq\gamma^3)}{(1-(1-p)\gamma - (1-q)p\gamma^2)(1-\gamma)} \\ V(3) = \dfrac{1}{1-\gamma}. \end{cases}$$

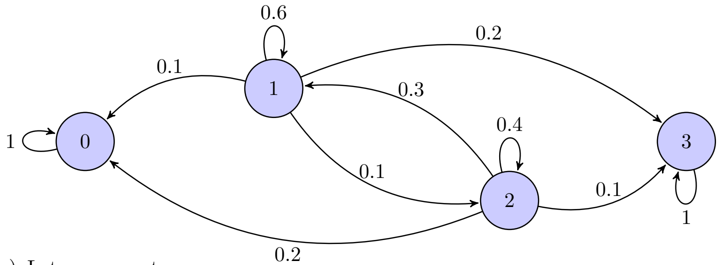
In particular, when $p = q = 1$ we check that

$$\begin{cases} V(1) = -2 + 3\gamma + \dfrac{\gamma^2}{1-\gamma} \\ V(2) = 3 + \dfrac{\gamma}{1-\gamma}, \\ \\ V(3) = \dfrac{1}{1-\gamma} = \displaystyle\sum_{n \geqslant 0} \gamma^n. \end{cases}$$

Exercise 5.19

a) It clearly takes S steps for Buffalo A to travel up from ⓪ to Ⓢ, and for Buffalo B to travel down from Ⓢ to ⓪ ?
b) After the buffalos collide they can be assumed to both continue their way without any impact on their travel times to the boundary {⓪, Ⓢ}, therefore the answer is S steps in this case as well.

Exercise 5.20 The chain has the following graph:



a) Let us compute

$$g_0(k) = \mathbb{P}(T_0 < \infty \mid X_0 = k) = \mathbb{P}(X_{T_{\{0,3\}}} = 0 \mid X_0 = k), \qquad k = 0, 1, 2, 3.$$

Since states ⓪ and ③ are absorbing, by first step analysis we have

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = 0.1 \times g_0(0) + 0.6 \times g_0(1) + 0.1 \times g_0(2) + 0.2 \times g_0(3) \\ g_0(2) = 0.2 \times g_0(0) + 0.3 \times g_0(1) + 0.4 \times g_0(2) + 0.1 \times g_0(3) \\ g_0(3) = 0, \end{cases}$$

i.e.

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = 0.1 + 0.6 \times g_0(1) + 0.1 \times g_0(2) \\ g_0(2) = 0.2 + 0.3 \times g_0(1) + 0.4 \times g_0(2) \\ g_0(3) = 0, \end{cases}$$

which has for solution

$$g_0(0) = 1, \quad g_0(1) = \frac{8}{21}, \quad g_0(2) = \frac{11}{21}, \quad g_0(3) = 0,$$

cf. also (5.1.11).

b) Let

$$h_{0,3}(k) = \mathbb{E}[T_{\{0,3\}} \mid X_0 = k]$$

denote the mean time to reach the set $A = \{0,3\}$ starting from $k = 0,1,2,3$. By first step analysis, we have

$$\begin{cases} h_{0,3}(0) = 0 \\ h_{0,3}(1) = 0.1 \times 1 + 0.6 \times (1 + h_{0,3}(1)) + 0.1 \times (1 + h_{0,3}(2)) + 0.2 \times (1 + h_{0,3}(3)) \\ h_{0,3}(2) = 0.2 \times 1 + 0.4 \times (1 + h_{0,3}(1)) + 0.3 \times (1 + h_{0,3}(2)) + 0.1 \times (1 + h_{0,3}(3)) \\ h_{0,3}(3) = 0, \end{cases}$$

i.e.

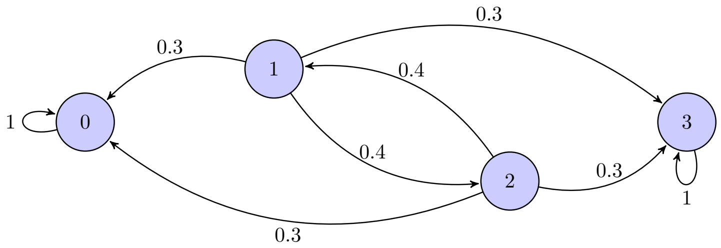
$$\begin{cases} h_{0,3}(0) = 0 \\ h_{0,3}(1) = 1 + 0.6 \times h_{0,3}(1) + 0.1 \times h_{0,3}(2) \\ h_{0,3}(2) = 1 + 0.4 \times h_{0,3}(1) + 0.3 \times h_{0,3}(2) \\ h_{0,3}(3) = 0, \end{cases}$$

which has for solution

$$h_{0,3}(0) = 0, \quad h_{0,3}(1) = \frac{10}{3}, \quad h_{0,3}(2) = \frac{10}{3}, \quad h_{0,3}(3) = 0.$$

Note that the relation $h_{0,3}(1) = h_{0,3}(2)$ can be guessed from the symmetry of the problem.

Exercise 5.21 The chain has the following graph:



a) The absorbing states are ⓪ and ③.
b) By (5.1.12) we have $g_0(1) = g_3(1) = 1/2$. On the other hand, we clearly have $g_1(0) = g_1(3) = 0$ and $g_1(1) = 1$, hence

$$g_1(2) = 0.3 \times g_1(0) + 0.4 \times g_1(1) + 0.3 \times g_1(3) = 0.4.$$

c) We clearly have $p_1(0) = p_1(3) = 0$, and

$$\begin{cases} p_1(1) = 0.3 \times p_1(0) + 0.4 \times p_1(2) + 0.3 \times p_1(3) = 0.4 \times p_1(2) \\ p_1(2) = 0.3 \times p_1(0) + 0.4 + 0.3 \times p_1(3) = 0.4, \end{cases}$$

hence $p_1(1) = 0.16$.
d) We have $h_1(1) = 0$ by construction and $h_1(0) = h_1(3) = +\infty$ because states ⓪ and ③ are absorbing, and $h_1(2) = +\infty$ because $g_0(2) \geqslant 0.3 > 0$. Regarding mean return times, we have $\mu_1(0) = \mu_1(1) = \mu_1(2) = \mu_1(3) = +\infty$ because states ① and ② communicate while states ⓪ and ③ are absorbing.

Exercise 5.22

a) Let $g_2(k) = \mathbf{P}(T_2 < \infty \mid X_0 = k)$, $k = 0, 1, 2$. We have $g_2(2) = 1$ and by first step analysis,

$$\begin{cases} g_2(0) = \dfrac{1}{3} g_2(0) + \dfrac{1}{3} g_2(1) + \dfrac{1}{3} \\ \\ g_2(1) = \dfrac{1}{4} g_2(0) + \dfrac{3}{4} g_2(1), \end{cases}$$

with $g_2(2) = 1$ since state ② is absorbing. This shows that

$$\begin{cases} \dfrac{2}{3} g_2(0) = \dfrac{1}{3} g_2(1) + \dfrac{1}{3} \\ \\ g_2(1) = g_2(0), \end{cases}$$

hence $\mathbf{P}(T_2 < \infty \mid X_0 = 1) = g_2(0) = g_2(1) = g_2(2) = 1$.

Next, let $g_1(k) = \mathbf{P}(T_1^r < \infty \mid X_0 = k)$, $k = 0, 1, 2$. We have $g_1(2) = 0$ and by first step analysis we have

$$\begin{cases} g_1(0) = \frac{1}{3} g_1(0) + \frac{1}{3} \\ \\ g_1(1) = \frac{1}{4} g_1(0) + \frac{3}{4} \\ \\ g_1(2) = 0, \end{cases}$$

hence

$$\begin{cases} g_1(0) = \frac{1}{2} \\ \\ g_1(1) = \frac{7}{8} \\ \\ g_1(2) = 0, \end{cases}$$

and $\mathbf{P}(T_1^r < \infty \mid X_0 = 1) = g_1(1) = 7/8$.

b) We have $\mathbb{E}[T_1^r \mid X_0 = 1] = +\infty \times \mathbb{P}(T_1^r = \infty \mid X_0 = 1) = +\infty$ because $\mathbb{P}(T_1^r = \infty \mid X_0 = 1) > 0$ as state ① is transient, cf. Proposition 7.4.

On the other hand, let $h_2(k) = \mathbb{E}[T_2 < \infty \mid X_0 = k]$, $k = 0, 1, 2$. We have $h_2(2) = 0$ and by first step analysis we have

$$\begin{cases} h_2(0) = 1 + \frac{1}{3} h_2(0) + \frac{1}{3} h_2(1) \\ \\ h_2(1) = 1 + \frac{1}{4} h_2(0) + \frac{3}{4} h_2(1) \\ \\ h_2(2) = 0, \end{cases}$$

hence

$$\begin{cases} 2h_2(0) = 3 + h_2(1) \\ \\ h_2(1) = 4 + h_2(0) \\ \\ h_2(2) = 0, \end{cases}$$

hence

$$\begin{cases} h_2(0) = 7 \\ \\ h_2(1) = 11 \\ \\ h_2(2) = 0, \end{cases}$$

and $\mathbb{E}[T_2^r < \infty \mid X_0 = 1] = h_2(1) = 11$.

Exercise 5.23

a) By a recurrence using Pascal's identity

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},$$

we find

$$[P^n]_{i,j} = \begin{cases} p^{j-i} q^{n-(j-i)} \binom{n}{j-i}, & 0 \leqslant j-i \leqslant n, \\ 0, & n < j-i, \\ 0, & i > j. \end{cases}$$

b) We have

$$\begin{aligned} 0 &\leqslant \lim_{n\to\infty} [P^n]_{i,j} \\ &= \frac{p^{j-i} q^{-(j-i)}}{(j-i)!} \lim_{n\to\infty} q^n \frac{n!}{(n-(j-i))!} \\ &= \lim_{n\to\infty} q^n n(n-1)\cdots(n-(j-i)+1) \\ &\leqslant \lim_{n\to\infty} q^n n^{j-i} \\ &= \lim_{n\to\infty} e^{\log(q^n n^{j-i})} \\ &= \lim_{n\to\infty} e^{n\log q + (j-i)\log n} \\ &= 0, \qquad 0 \leqslant j-i. \end{aligned}$$

c) We have

$$\begin{aligned} \sum_{n\geqslant 0} [P^n]_{i,j} &= \begin{cases} \displaystyle\sum_{n\geqslant j-i} p^{j-i} q^{n-(j-i)} \binom{n}{j-i}, & i \leqslant j, \\ 0, & i > j, \end{cases} \\ &= \begin{cases} \displaystyle\frac{p^{j-i}}{(j-i)!} \sum_{n\geqslant 0} q^n \frac{(n+j-i)!}{n!}, & i \leqslant j, \\ 0, & i > j, \end{cases} \\ &= \begin{cases} \displaystyle\frac{p^{j-i}}{(j-i)!} \sum_{n\geqslant 0} q^n \frac{(n+j-i)!}{n!}, & i \leqslant j, \\ 0, & i > j, \end{cases} \end{aligned}$$

$$= \begin{cases} \dfrac{p^{j-i}}{(j-i)!} \dfrac{\partial^{j-i}}{\partial q^{j-i}} \dfrac{1}{1-q}, & i \leqslant j, \\ 0, & i > j, \end{cases}$$

$$= \begin{cases} \dfrac{p^{j-i}}{(1-q)^{j-i+1}}, & i \leqslant j, \\ 0, & i > j, \end{cases}$$

$$= \begin{cases} \dfrac{1}{p}, & i \leqslant j, \\ 0, & i > j. \end{cases}$$

d) We have

$$p_{i,j} = \mathbb{P}(T_j < \infty \mid X_0 = i) = \begin{cases} 1, & i < j, \\ q < 1, & i = j, \\ 0, & i > j. \end{cases}$$

e) Since $p_{i,i} = q < 1$ for all $i \geqslant 0$, the chain $(X_n)_{n \geqslant 0}$ is transient as all of its states are transient.

f) As in Proposition 5.7, the mean number of returns from state (i) to state (j) is given by

$$\sum_{n \geqslant 1} [P^n]_{i,j} = \mathbb{E}[R_j \mid X_0 = i] = \begin{cases} p \displaystyle\sum_{n \geqslant 1} n q^{n-1} = \frac{1}{p} = \frac{p_{i,j}}{1 - p_{j,j}}, & i < j, \\ qp \displaystyle\sum_{n \geqslant 1} n q^{n-1} = \frac{q}{p} = \frac{p_{i,i}}{1 - p_{i,i}}, & i = j, \\ 0 = \dfrac{p_{i,j}}{1 - p_{j,j}}, & i > j. \end{cases}$$

g) The matrix

$$I - P = \begin{bmatrix} 1-q & -p & 0 & 0 & \cdots \\ 0 & 1-q & -p & 0 & \cdots \\ 0 & 0 & 1-q & -p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} p & -p & 0 & 0 & \cdots \\ 0 & p & -p & 0 & \cdots \\ 0 & 0 & p & -p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

is invertible, and as in (6.3.5), its inverse can be expressed as

$$(I - P)^{-1} = \left[\sum_{n \geqslant 0} [P^n]_{i,j} \right]_{i,j \in \mathbb{N}}$$

$$= \left[\mathbf{1}_{\{i=j\}} + \mathbb{E}[R_j \mid X_0 = i] \right]_{i,j\in\mathbb{N}}$$

$$= \begin{bmatrix} \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \cdots \\ 0 & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \cdots \\ 0 & 0 & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \cdots \\ 0 & 0 & 0 & \frac{1}{p} & \frac{1}{p} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Note that although the vector $e = (1,1,1,\ldots)$ satisfies $(I-P)e = 0$ it does not belong to $\ell^1(\mathbb{N})$, and $I-P$ is invertible as an operator from $\ell^1(\mathbb{N})$ into

$$\left\{(u_k)_{k\geqslant 0} : \sum_{n\geqslant 0}\Big|\sum_{k\geqslant n} u_k\Big| < \infty\right\}.$$

Exercise 5.24

a) We have $\mu_A(x,y) = 0$ for all $(x,y) \in A$.
b) For all $0 \leqslant x,y \leqslant 3$ we have

$$\mu_A(x,y) = 1 + \frac{1}{2}\mu_A(x+1,y) + \frac{1}{2}\mu_A(x,y+1). \qquad \text{(S.38)}$$

c) We have

$$\begin{cases}\mu_A(2,2) = 1 + \frac{1}{2}\mu_A(3,2) + \frac{1}{2}\mu_A(2,3) = 1, \\ \mu_A(1,2) = 1 + \frac{1}{2}\mu_A(2,2) + \frac{1}{2}\mu_A(1,3) = \frac{3}{2}, \\ \mu_A(2,1) = 1 + \frac{1}{2}\mu_A(2,2) + \frac{1}{2}\mu_A(3,1) = \frac{3}{2}, \\ \mu_A(0,2) = 1 + \frac{1}{2}\mu_A(1,2) + \frac{1}{2}\mu_A(0,3) = \frac{7}{4}, \\ \mu_A(2,0) = 1 + \frac{1}{2}\mu_A(2,1) + \frac{1}{2}\mu_A(3,0) = \frac{7}{4}, \\ \mu_A(1,1) = 1 + \frac{1}{2}\mu_A(2,1) + \frac{1}{2}\mu_A(1,2) = \frac{5}{2}, \\ \mu_A(0,1) = 1 + \frac{1}{2}\mu_A(1,1) + \frac{1}{2}\mu_A(0,2) = \frac{25}{8}, \\ \mu_A(1,0) = 1 + \frac{1}{2}\mu_A(1,1) + \frac{1}{2}\mu_A(2,0) = \frac{25}{8}, \\ \mu_A(0,0) = 1 + \frac{1}{2}\mu_A(1,0) + \frac{1}{2}\mu_A(0,1) = \frac{33}{8}. \end{cases}$$

4	0	0	0	0	0
3	0	0	0	0	0
2	7/4	3/2	1	0	0
1	25/8	5/2	3/2	0	0
0	33/8	25/8	7/4	0	0
	0	1	2	3	4

Table 14.3: Values of $\mu_A(x, y)$ with $N = 3$ and the set A in blue.

d) The mean number of rounds is $\mu_A(0,0) = 33/8 = 4.125$.

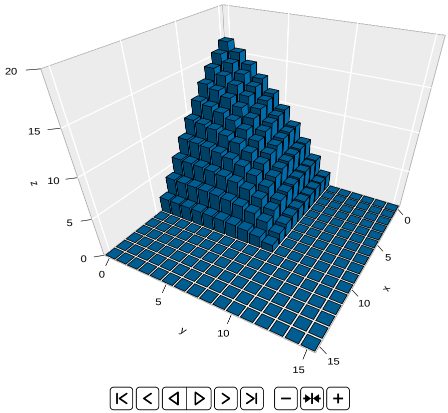


Fig. S.14: Backward solution of Equation (S.38) for $\mu_A(x,y)$ with $N=10$.*

```
install.packages("plot3D"); require(plot3D);N=10;M=15
X=array(1:2,c(M+1,M+1));
for (i in seq(1,M+1)) {for (j in seq(1,M+1)) X[i,j]=0;}
par(mar=c(1,2,0,0)+0.01)
for (k in seq(N,-N)) {for (i in seq(k,N)) {
if (i>=1 && N+k-i>=1) {X[i,N+k-i]=1+(X[i+1,N+k-i]+X[i,N+k-i+1])/2.0;dev.hold();
hist3D(x=0:M, y=0:M, z=X, scale=T, bty="g", phi=35, theta=120, border="black",
       zlim=c(0,20), shade=0.3, space=0.15, col="#0072B2", colkey=F,
       ticktype="detailed"); dev.flush();}}}
```

Exercise 5.25

a) When $X_0 = x \geqslant 2$ and $Y_0 = y \geqslant 2$ we have $T_A = 0$, hence

$$\mu_A(x,y) := \mathbb{E}[T_A < \infty \mid X_0 = x,\ Y_0 = y] = 0, \qquad x \geqslant 2, \quad y \geqslant 2.$$

b) This equation is obtained by first step analysis, noting that we can only move up to to the right with probability $1/2$ in both cases.

c) We note that $\mu_A(x,y) = \mu_A(x,y+1)$ for $y \geqslant 2$, and

$$\mu_A(1,y) = 1 + \frac{1}{2}\mu_A(2,y) + \frac{1}{2}\mu_A(1,y+1) = 1 + \frac{1}{2}\mu_A(1,y), \quad y \geqslant 2,$$

hence $\mu_A(1,y) = 2$ for all $y \geqslant 2$. We also have

* Animated figure (works in Acrobat Reader).

$$\mu_A(0,y) = 1 + \frac{1}{2}\mu_A(1,y) + \frac{1}{2}\mu_A(0,y+1) = 2 + \frac{1}{2}\mu_A(0,y), \quad y \geqslant 2,$$

hence $\mu_A(0,y) = 4$, $y \geqslant 2$. By symmetry we also have $\mu_A(x,1) = 2$ and $\mu_A(x,0) = 4$ for all $x \geqslant 2$.

These results can also be recovered using pathwise analysis as

$$\mu_A(1,y) = \sum_{k\geqslant 1} \frac{k}{2^k} = \frac{1}{2}\sum_{k\geqslant 0} \frac{k}{2^{k-1}} = \frac{1}{2(1-1/2)^2} = 2, \quad y \geqslant 2,$$

which yields similarly $\mu_A(x,1) = 2$ for all $x \geqslant 2$. Repeating this argument once also leads to $\mu_A(x,0) = \mu_A(0,y) = 4$ for all $x, y \geqslant 2$.

d) We have

$$\begin{cases} \mu_A(1,1) = 1 + \frac{1}{2}\mu_A(2,1) + \frac{1}{2}\mu_A(1,2) = 3, \\ \\ \mu_A(0,1) = 1 + \frac{1}{2}\mu_A(1,1) + \frac{1}{2}\mu_A(0,2) = \frac{9}{2}, \\ \\ \mu_A(1,0) = 1 + \frac{1}{2}\mu_A(2,0) + \frac{1}{2}\mu_A(1,1) = \frac{9}{2}, \\ \\ \mu_A(0,0) = 1 + \frac{1}{2}\mu_A(1,0) + \frac{1}{2}\mu_A(0,1) = \frac{11}{2}, \end{cases}$$

hence the mean time it takes until both cans contain at least \$2 is $\mu_A(0,0) = 11/2$.

4	4	2	0	0	0
3	4	2	0	0	0
2	4	2	0	0	0
1	9/2	3	2	2	2
0	11/2	9/2	4	4	4
	0	1	2	3	4

Table 14.4: Values of $\mu_A(x,y)$ with $N = 2$ and the set A in blue.

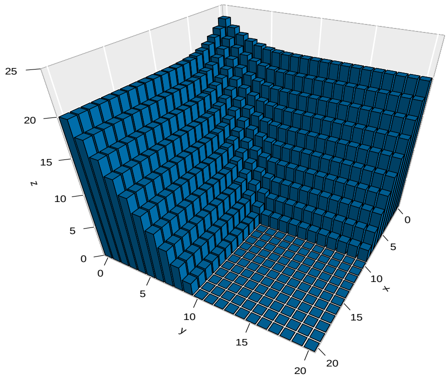


Fig. S.15: Backward solution of (5.4.11) for $\mu_A(x, y)$ with $N = 10$.*

```
require(plot3D);N=10;M=20;X=array(1:2,c(M+1,M+1));
for (i in seq(N+2,M+1)) {for (j in seq(N+2,M+1)) X[i,j]=0;}
for (i in seq(N+1,M+1)) {for (j in seq(1,N+1)) X[i,j]=2*(N+1-j);}
for (i in seq(1,N+1)) {for (j in seq(N+1,M+1)) X[i,j]=2*(N+1-i);}
for (k in seq(N,-N)) {for (i in seq(k,N)) {if (i>=1 && N+k-i>=1)
      X[i,N+k-i]=1+(X[i+1,N+k-i]+X[i,N+k-i+1])/2.0;}}
hist3D(x = 1:21, y = 1:21, z = X, scale = T, bty="g", phi=35, theta=120, border="black",
      zlim=c(0,25), shade = 0.3, space=0.15, col = "#0072B2", colkey = F, ticktype =
      "detailed")
```

Exercise 5.26

a) We have

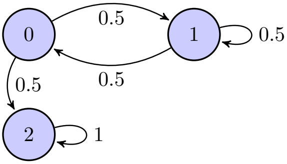
$$
\begin{aligned}
h(k) &= \mathbb{E}\left[\sum_{i \geqslant 0} \beta^i c(X_i) \;\middle|\; X_0 = k\right] \\
&= \mathbb{E}[c(X_0) \mid X_0 = k] + \mathbb{E}\left[\sum_{i \geqslant 1} \beta^i c(X_i) \;\middle|\; X_0 = k\right] \\
&= c(k) + \sum_{j \in S} P_{k,j} \mathbb{E}\left[\sum_{i \geqslant 1} \beta^i c(X_i) \;\middle|\; X_1 = j\right]
\end{aligned}
$$

* Animated figure (works in Acrobat Reader).

$$= c(k) + \beta \sum_{j \in S} P_{k,j} \mathbb{E}\left[\sum_{i \geqslant 0} \beta^i c(X_i) \ \Big| \ X_0 = j \right]$$
$$= c(k) + \beta \sum_{j \in S} P_{k,j} h(j), \qquad k \in S.$$

This type of equation may be difficult to solve in full generality.

b) The chain has the following graph:



The average utility $h(k)$ solves the first step analysis equations

$$\begin{cases} h(0) = c(0) + \frac{1}{2} h(1) = 5 + \frac{1}{2} h(1) \\ \\ h(1) = c(1) + \frac{1}{2} h(0) = -2 + \frac{1}{2} h(0) \\ \\ h(2) = 0, \end{cases}$$

which yields

$$h(0) = \frac{16}{3}, \quad h(1) = \frac{2}{3}, \quad h(2) = 0.$$

We also refer to Problem 5.34 for another example with explicit solution.

Exercise 5.27

a) We have

$$\mathbb{E}[T^{(m)}] = mp^m + \sum_{k=0}^{m-1} p^k q(k + 1 + \mathbb{E}[T^{(m)}]).$$

b) We find

$$\mathbb{E}[T^{(m)}] = \frac{mp^m + q \sum_{k=0}^{m-1} p^k (k+1)}{1 - q \sum_{k=0}^{m-1} p^k}$$

$$
\begin{aligned}
&= \frac{mp^m + \dfrac{1-(m+1)p^m + mp^{m+1}}{1-p}.}{p^m} \\
&= \frac{m(1-p)p^m + 1 - (m+1)p^m + mp^{m+1}}{(1-p)p^m} \\
&= \frac{1/p^m - 1}{1-p} \\
&= \sum_{k=1}^{m} \frac{1}{p^k}. \qquad \text{(S.39)}
\end{aligned}
$$

Alternative solution: We note the recurrence relation

$$
\mathbb{E}[T^{(m)}] = \mathbb{E}[T^{(m-1)}] + p \times 1 + (1-p)\mathbb{E}[T^{(m)}], \qquad m \geqslant 2,
$$

or

$$
\mathbb{E}[T^{(m)}] = \frac{\mathbb{E}[T^{(m-1)}] + p}{p}, \qquad m \geqslant 2,
$$

with $\mathbb{E}[T^{(1)}] = 1/p$, which also recovers (S.39).

Exercise 5.28 We have

$$
\begin{aligned}
V^*(k) &= \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0} \gamma^n R(X_n) \;\middle|\; X_0 = k \right] \\
&= \operatorname*{Max}_{\pi} \mathbb{E}\left[R(X_0) + \sum_{n\geqslant 1} \gamma^n R(X_n) \;\middle|\; X_0 = k \right] \\
&= \operatorname*{Max}_{\pi} \mathbb{E}\left[R(k) + \gamma \sum_{n\geqslant 1} \gamma^{n-1} R(X_n) \;\middle|\; X_0 = k \right] \\
&= \operatorname*{Max}_{\pi} \left(R(k) + \mathbb{E}\left[\gamma \sum_{n\geqslant 1} \gamma^{n-1} R(X_n) \;\middle|\; X_0 = k \right] \right) \\
&= R(k) + \gamma \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 1} \gamma^{n-1} R(X_n) \;\middle|\; X_0 = k \right] \\
&= R(k) + \gamma \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P^a_{k,l} \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 1} \gamma^{n-1} R(X_n) \;\middle|\; X_1 = l \right] \\
&= R(k) + \gamma \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P^a_{k,l} \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0} \gamma^{n} R(X_{n+1}) \;\middle|\; X_1 = l \right]
\end{aligned}
$$

$$= R(k) + \gamma \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P^a_{k,l}\mathbb{E}\left[\sum_{n\geqslant 0}\gamma^n R(X_n) \;\middle|\; X_0 = l\right]$$
$$= R(k) + \gamma \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P^a_{k,l} V^*(l), \qquad k \in \mathbb{S}.$$

The optimal policy $\pi^*_k \in \mathbb{A}$ starting from state $k \in \mathbb{S}$ is given by

$$\pi^*_k = \mathrm{argmax}_{a\in\mathbb{A}} \sum_{l\in\mathbb{S}} P^a_{k,l} V^*(l).$$

Exercise 5.29

a) We have $h_3(3) = 0$, and

$$\begin{cases} h_3(1) = 1 + (1-p)h_3(1) + ph_3(2) \\ h_3(2) = 1 + (1-q)h_3(1), \end{cases}$$

hence

$$\begin{cases} h_3(1) = 1 + (1-p)h_3(1) + ph_3(2) = 1 + p + ((1-p) + (1-q)p)h_3(1) \\ h_3(2) = 1 + (1-q)h_3(1), \end{cases}$$

hence

$$\begin{cases} h_3(1) = \dfrac{1+p}{1-(1-p)-(1-q)p} = \dfrac{1+p}{pq} \\ h_3(2) = 1 + \dfrac{(1+p)(1-q)}{1-((1-p)+(1-q)p)} = \dfrac{1+p-q}{pq}. \end{cases}$$

b) We have

$$\begin{cases} G_1(s) = (1-p)s\mathbb{E}\big[s^{T_3} \,\big|\, X_0 = 1\big] + ps\mathbb{E}\big[s^{T_3} \,\big|\, X_0 = 2\big] \\ G_2(s) = (1-q)s\mathbb{E}\big[s^{T_3} \,\big|\, X_0 = 1\big] + qs, \end{cases}$$

hence

$$\begin{cases} G_1(s) = (1-p)sG_1(s) + psG_2(s) \\ G_2(s) = (1-q)sG_1(s) + qs \end{cases}$$

or

$$\begin{cases} G_1(s) = (1+ps)G_2(s) - qs \\ G_2(s) = (1-q)sG_1(s) + qs \end{cases}$$

i.e.

$$\begin{cases} G_1(s) = (1+ps)(1-q)sG_1(s) + qs(1+ps) - qs \\ \\ G_2(s) = (1-q)s(1+ps)G_2(s) - q(1-p)s^2 + qs, \end{cases}$$

hence

$$\begin{cases} G_1(s) = \dfrac{pqs^2}{1-(1-p)s-p(1-q)s^2} \\ \\ G_2(s) = \dfrac{-q(1-p)s^2+qs}{1-(1-q)s(1+ps)} \end{cases}$$

c) Using the identity

$$\frac{\sqrt{(1-p)^2+4(1-q)p}}{1-(1-p)s-p(1-q)s^2} = \sum_{n=0}^{\infty} \frac{s^n}{z_+^{n+1}} - \sum_{n=0}^{\infty} \frac{s^n}{z_-^{n+1}},$$

we find

$$\begin{aligned} G_1(s) &= \frac{pqs^2}{1-(1-p)s-p(1-q)s^2} \\ &= \frac{pqs^2}{\sqrt{(1-p)^2+4(1-q)p}} \sum_{n=0}^{\infty} \left(\frac{s^n}{z_+^{n+1}} - \frac{s^n}{z_-^{n+1}} \right) \\ &= \frac{pq}{\sqrt{(1-p)^2+4(1-q)p}} \sum_{n=2}^{\infty} \left(\frac{s^n}{z_+^{n-1}} - \frac{s^n}{z_-^{n-1}} \right) \\ &= \sum_{n=0}^{\infty} s^n \mathbb{P}(T_3 = n \mid X_0 = 1), \qquad -1 \leqslant s \leqslant 1, \end{aligned}$$

hence by identification we find $\mathbb{P}(T_3 = n \mid X_0 = 1) = 0$, $n = 0, 1$, and

$$\mathbb{P}(T_3 = n \mid X_0 = 1) = \frac{pq}{\sqrt{(1-p)^2+4(1-q)p}} \left(\frac{1}{z_+^{n-1}} - \frac{1}{z_-^{n-1}} \right), \quad n \geqslant 2.$$

In particular, this recovers

$$\begin{aligned} \mathbb{P}(T_3 = 2 \mid X_0 = 1) &= \frac{pq}{\sqrt{(1-p)^2+4(1-q)p}} \left(\frac{1}{z_+} - \frac{1}{z_-} \right) \\ &= \frac{pq}{\sqrt{(1-p)^2+4(1-q)p}} \frac{z_- - z_+}{z_- z_+} \\ &= pq. \end{aligned}$$

d) We note that the hitting time is *a.s.** finite, *i.e.* $\mathbb{P}(T_3 < \infty \mid X_0 = 1) = 1$, hence the mean hitting time $\mathbb{E}[T_3 \mid X_0 = 1]$ is given from (1.7.6) as

$$\begin{aligned}\mathbb{E}[T_3 \mid X_0 = 1] &= G_1'(1) \\ &= \frac{2pqs}{1-(1-p)s-p(1-q)s^2}_{|s=1} \\ &\quad + \frac{pqs^2(1-p)+2p(1-q)s}{(1-(1-p)s-p(1-q)s^2)^2}_{|s=1} \\ &= \frac{1+p}{pq}.\end{aligned}$$

Exercise 5.30 (See also *here*). By first step analysis, we have

$$\begin{cases} h(3) = 1 + h(2), \\ h(2) = 1 + \frac{2}{3}h(1) + \frac{1}{3}h(3), \\ h(1) = 1 + \frac{1}{3} \times 0 + \frac{2}{3}h(2), \\ h(0) = 0, \end{cases}$$

which yields

$$\begin{aligned} h(2) &= 1 + \frac{2}{3}\left(1 + \frac{2}{3}h(2)\right) + \frac{1}{3}(1 + h(2)) \\ &= 1 + \frac{2}{3} + \frac{4}{9}h(2) + \frac{1}{3} + \frac{1}{3}h(2) \\ &= 2 + \frac{7}{9}h(2), \end{aligned}$$

hence

$$\begin{cases} h(3) = 10, \\ h(2) = 9, \\ h(1) = 7 \\ h(0) = 0. \end{cases}$$

Problem 5.31 (See also *here*).

a) We have $h(d) = 0$.
b) We have $h(0) = 1 + h(1)$.
c) We have

* Almost surely, *i.e.* with probability one.

$$h(r) = 1 + \frac{r}{d}h(r-1) + \frac{d-r}{d}h(r+1), \qquad r = 1, 2, \ldots, d-1.$$

d) We have

$$h(r) = 1 + \frac{r}{d}h(r-1) + \frac{d-r}{d}h(r+1), \qquad r = 1, 2, \ldots, d-1,$$

hence

$$\frac{r}{d}h(r) + \frac{d-r}{d}h(r) = 1 + \frac{r}{d}h(r-1) + \frac{d-r}{d}h(r+1),$$

hence

$$\frac{r}{d}f(r-1) = 1 + \frac{d-r}{d}f(r), \qquad r = 1, 2, \ldots, d-1.$$

e) We have $f(0) = h(1) - h(0) = -1$, and

$$f(r) = -\frac{d}{d-r} + \frac{r}{d-r}f(r-1), \qquad r = 1, 2, \ldots, d,$$

hence

$$f(r) = -\frac{1}{\binom{d-1}{r}}\sum_{l=0}^{r}\binom{d}{l}, \qquad r = 0, 1, \ldots, r.$$

f) We have

$$\begin{aligned} h(r) &= h(d) + \sum_{k=r}^{d-1}(h(k) - h(k+1)) \\ &= h(d) - \sum_{k=r}^{d-1} f(k) \\ &= \sum_{k=r}^{d-1}\frac{1}{\binom{d-1}{k}}\sum_{l=0}^{k}\binom{d}{l}, \qquad r = 0, 1, \ldots, d. \end{aligned}$$

g) We have

$$h(0) = \sum_{k=0}^{d-1}\frac{1}{\binom{d-1}{k}}\sum_{l=0}^{k}\binom{d}{l}$$

and

$$h(1) = \sum_{k=1}^{d-1}\frac{1}{\binom{d-1}{k}}\sum_{l=0}^{k}\binom{d}{l}, \quad \text{and} \quad h(2) = \sum_{k=2}^{d-1}\frac{1}{\binom{d-1}{k}}\sum_{l=0}^{k}\binom{d}{l}.$$

h) i) When $d = 1$ we find $h(0) = 1$, $h(1) = 0$.
ii) When $d = 2$ we find $h(0) = 4$, $h(1) = 3$, h(2) = 0.

iii) When $d = 3$ we have $h(0) = 10$, $h(1) = 9$, $h(2) = 7$, $h(3) = 0$.

Remark. This random walk is the same as the one in Exercises 6.7 and 7.4 below on the Ehrenfest chain.

Problem 5.32

a) The boundary conditions $g(0)$ and $g(N)$ are given by $g(0) = 1$ and $g(N) = 0$.
b) We have

$$\begin{aligned} g(k) &= \mathbb{P}(T_0 < T_N \mid X_0 = k) \\ &= \sum_{l=0}^{N} \mathbb{P}(T_0 < T_N \mid X_1 = l)\mathbb{P}(X_1 = l \mid X_0 = k) \\ &= \sum_{l=0}^{N} \mathbb{P}(T_0 < T_N \mid X_1 = l) P_{k,l} \\ &= \sum_{l=0}^{N} \mathbb{P}(T_0 < T_N \mid X_0 = l) P_{k,l} \\ &= \sum_{l=0}^{N} g(l) P_{k,l}, \qquad k = 0, 1, \ldots, N. \end{aligned}$$

c) We find

$$[\ P_{i,j}\ ]_{0 \leqslant i,j \leqslant 3} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ (2/3)^3 & (2/3)^2 & 2/3^2 & 1/3^3 \\ 1/3^3 & 2/3^2 & (2/3)^2 & (2/3)^3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

d) Letting $g(k) = 1 - k/N$, we check that $g(k)$ satisfies the boundary conditions $g(0) = 1$ and $g(N) = 0$, and in addition we have

$$\begin{aligned} \sum_{l=0}^{N} g(l) P_{k,l} &= \sum_{l=0}^{N} \frac{N!}{(N-l)!l!} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l} \frac{N-l}{N} \\ &= \sum_{l=0}^{N-1} \frac{N!}{(N-l)!l!} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l} \frac{N-l}{N} \\ &= \sum_{l=0}^{N-1} \frac{(N-1)!}{(N-l-1)!l!} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l} \end{aligned}$$

$$= \left(1 - \frac{k}{N}\right) \sum_{l=0}^{N-1} \binom{N-1}{l} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-1-l}$$
$$= \left(1 - \frac{k}{N}\right) \left(\frac{k}{N} + 1 - \frac{k}{N}\right)^{N-1}$$
$$= \frac{N-k}{N}$$
$$= g(k), \qquad k = 0, 1, \ldots, N,$$

which allows us to conclude by uniqueness of the solution given two boundary conditions, cf. Exercise 5.10 for cases of non-uniqueness under a single boundary condition.

e) The boundary conditions $h(0)$ and $h(N)$ are given by $h(0) = 0$ and $h(N) = 0$ since the states $\textcircled{0}$ and $\textcircled{N}$ are absorbing.

f) We have

$$\begin{aligned}
h(k) &= \mathbb{E}[T_{0,N} \mid X_0 = k] \\
&= \sum_{l=0}^{N} (1 + \mathbb{E}[T_{0,N} \mid X_1 = l]) \mathbb{P}(X_1 = l \mid X_0 = k) \\
&= \sum_{l=0}^{N} (1 + \mathbb{E}[T_{0,N} \mid X_1 = l]) P_{k,l} = \sum_{l=0}^{N} P_{k,l} + \sum_{l=0}^{N} \mathbb{E}[T_{0,N} \mid X_1 = l] P_{k,l} \\
&= 1 + \sum_{l=1}^{N-1} \mathbb{E}[T_{0,N} \mid X_1 = l] P_{k,l} = 1 + \sum_{l=1}^{N-1} h(l) P_{k,l}, \qquad \text{(S.40)}
\end{aligned}$$

$k = 1, 2, \ldots, N-1$.

g) In this case, the Equation (S.40) reads

$$\begin{cases}
h(0) = 0, \\
h(1) = 1 + \dfrac{4}{9} h(1) + \dfrac{2}{9} h(2) \\
h(2) = 1 + \dfrac{2}{9} h(1) + \dfrac{4}{9} h(2) \\
h(3) = 0,
\end{cases}$$

which yields

$$h(0) = 0, \quad h(1) = 3, \quad h(2) = 3, \quad h(3) = 0.$$

Problem 5.33 (Exercise 4.7 continued).

a) As in the solution of Exercise 4.7, the transition matrix is given by

$$\begin{array}{c} \\ aa \\ ab \\ ba \\ bb \end{array} \begin{array}{c} \begin{array}{cccc} aa & ab & ba & bb \end{array} \\ \left[\begin{array}{cccc} p & q & 0 & 0 \\ 0 & 0 & p & q \\ p & q & 0 & 0 \\ 0 & 0 & p & q \end{array} \right] \end{array}.$$

b) We have $\tau_{ab} = 1$ with probability one, hence

$$G_{ab}(s) = \mathbb{E}[s \mid Z_1 = (a, b)] = s.$$

c) We find

$$\begin{cases} G_{aa}(s) = psG_{aa}(s) + qsG_{ab}(s), \\ \\ G_{ba}(s) = psG_{aa}(s) + qsG_{ab}(s). \end{cases}$$

d) We have

$$\begin{cases} G_{aa}(s) = psG_{aa}(s) + qs^2, \\ \\ G_{ba}(s) = psG_{aa}(s) + qs^2, \end{cases}$$

hence

$$G_{aa}(s) = G_{ba}(s) = \frac{pqs^3}{1 - ps} + qs^2 = \frac{qs^2}{1 - ps}, \qquad s \in (-1, 1).$$

We note that

$$\begin{aligned} \mathbb{P}(\tau_{ab} < \infty \mid Z_1 = (a, a)) &= \mathbb{P}(\tau_{ab} < \infty \mid Z_1 = (b, a)) \\ &= G_{ba}(1^-) \\ &= \lim_{s \nearrow 1} G_{ba}(s) \\ &= \lim_{s \nearrow 1} \frac{qs^2}{1 - ps} \\ &= \frac{q}{1 - p} \\ &= 1. \end{aligned}$$

e) We have

$$\begin{aligned} \mathbb{E}[\tau_{ab} \mid Z_1 = (a, a)] &= \mathbb{E}[\tau_{ab} \mid Z_1 = (b, a)] \\ &= G'_{ba}(1) = G'_{aa}(1) \\ &= \frac{2q}{1 - p} + \frac{pq}{(1 - p)^2} = 2 + \frac{p}{q}. \end{aligned}$$

f) This average time is

$$p\mathbb{E}[\tau_{ab} \mid Z_1 = (a,a)] + q\mathbb{E}[\tau_{ab} \mid Z_1 = (a,b)] = p\left(2 + \frac{p}{q}\right) + q = 1 + \frac{p}{q}.$$

Problem 5.34

a) Since we consider the time until we hit either 0 or N, we have $h(0) = 0$ as well as $h(N) = 0$.

b) We have

$$\begin{aligned}
h(k) &= \mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k\right] \\
&= \mathbb{E}\left[X_0 \,\middle|\, X_0 = k\right] + \mathbb{E}\left[\sum_{i=1}^{\tau-1} X_i \,\middle|\, X_0 = k\right] \\
&= k + p\mathbb{E}\left[\sum_{i=1}^{\tau-1} X_i \,\middle|\, X_1 = k+1\right] + q\mathbb{E}\left[\sum_{i=1}^{\tau-1} X_i \,\middle|\, X_1 = k-1\right] \\
&= k + p\mathbb{E}\left[\sum_{i=0}^{\tau-2} X_{i+1} \,\middle|\, X_1 = k+1\right] + q\mathbb{E}\left[\sum_{i=0}^{\tau-2} X_{i+1} \,\middle|\, X_1 = k-1\right] \quad \text{(S.41)} \\
&= k + p\mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k+1\right] + q\mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k-1\right] \quad \text{(S.42)} \\
&= k + ph(k+1) + qh(k-1), \qquad 1 \leqslant k \leqslant N-1,
\end{aligned}$$

where we used the fact that $\tau - 1$ in (S.41) becomes τ in (S.42).

When passing from

$$p\mathbb{E}\left[\sum_{i=1}^{\tau-1} X_i \,\middle|\, X_1 = k+1\right] + q\mathbb{E}\left[\sum_{i=1}^{\tau-1} X_i \,\middle|\, X_1 = k-1\right]$$

to

$$p\mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k+1\right] + q\mathbb{E}\left[\sum_{i=0}^{\tau-1} X_i \,\middle|\, X_0 = k-1\right],$$

we are shifting the time summation index from i to $i-1$, and after that the summation starts from 0 instead of 1. Nevertheless the summation remains up to $\tau - 1$ at the second step because in both cases, τ denotes the first hitting time of 0 or N. Writing $\tau - 2$ in the second step would be wrong because the first hitting time of 0 or N within the expectation is τ (*not* $\tau - 1$).

From now on we take $p = q = 1/2$.

c) We check successively that $h(k) = C$, $h(k) = Ck$, $h(k) = Ck^2$ cannot be solution and that $h(k) = Ck^3$ is solution provided that $C = -1/3$.

Note that when trying successively $h(k) = C$, $h(k) = Ck$, $h(k) = Ck^2$ and $h(k) = Ck^3$, the quantity C has to be a **constant**.

Here, the index k is a **variable**, not a constant. Finding for example

$$C = 3k$$

can only be wrong because $3k$ is **not a constant**. On the other hand, finding $C = 2$ (say) would make sense because 2 **is a constant**. Note that N also has the status of a constant (parameter) here.

d) The general solution has the form

$$h(k) = -\frac{k^3}{3} + C_1 + C_2 k,$$

and the boundary conditions show that

$$\begin{cases} 0 = h(0) = C_1, \\ \\ 0 = h(N) = -\dfrac{N^3}{3} + C_1 + C_2 N, \end{cases}$$

hence $C_1 = 0$, $C_2 = N^2/3$, and

$$h(k) = -\frac{k^3}{3} + N^2\frac{k}{3} = \frac{k}{3}(N^2 - k^2) = k(N-k)\frac{N+k}{3}, \qquad k = 0, 1, \ldots, N. \tag{S.43}$$

e) When $N = 2$ we find $h(1) = 1$ since starting from $k = 1$ we can only move to state ⓪ or state $N = 2$ which ends the game with a cumulative sum equal to 1 in both cases.

f) i) We find an average of

$$\mathbb{E}[T_{0,N} \mid X_0 = 4] = 4(70-4) = 264 \text{ months } = 22 \text{ years}.$$

ii) By (S.43) we find

$$h(4) = \frac{4}{3}(70^2 - 4^2) = \$6512\text{K} = \$6.512\text{M}.$$

iii) In that case we find

$$\$4\text{K} \times 264 = \$1056\text{K} = \$1.056\text{M}.$$

iv) It appears that starting a (potentially risky) business is more profitable on the average than keeping the same fixed initial income over an equivalent (average) period of time.

However, this problem has been solved in the fair game case and without time discount, which may not be realistic. A more thorough analysis should consider

$$h(k) = \mathbb{E}\left[\sum_{i=0}^{\tau-1} \beta^i X_i \;\middle|\; X_0 = k\right]$$

with discount factor $\beta \in (0,1)$ in the case where $p < q$, with additional computational difficulties.

Note that the probability of hitting 0 first (*i.e.* ending with a bankruptcy) is given by

$$\mathbb{P}(X_{T_{0,N}} = 0 \mid X_0 = k) = 1 - \frac{k}{N} = \%94$$

when $k = 4$ and $N = 70$.

Problem 5.35

a) We have $f_{i,j}^{(1)} = P_{i,j}$, $i,j \in \mathbb{S}$.

b) We have

$$\begin{aligned} f_{i,j}^{(n+1)} &= \mathbb{P}(X_{n+1} = j, X_n \neq j, \ldots, X_1 \neq j \mid X_0 = i) \\ &= \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k}\mathbb{P}(X_{n+1} = j, X_n \neq j, \ldots, X_2 \neq j \mid X_1 = k) \\ &= \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k}\mathbb{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = k) \\ &= \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} f_{k,j}^{(n)}, \qquad i,j \in \mathbb{S},\ n \geqslant 1. \end{aligned}$$

c) By summing (5.4.16) over $n \geqslant 1$, we find

$$\begin{aligned} f_{i,j} &= \sum_{n\geqslant 1} f_{i,j}^{(n)} \\ &= f_{i,j}^{(1)} + \sum_{n\geqslant 2} f_{i,j}^{(n)} \end{aligned}$$

$$
\begin{aligned}
&= P_{i,j} + \sum_{n\geqslant 1} f_{i,j}^{(n+1)} \\
&= P_{i,j} + \sum_{n\geqslant 1} \sum_{\substack{k\in\mathbb{S} \\ k\neq j}} P_{i,k} f_{k,j}^{(n)} \\
&= P_{i,j} + \sum_{\substack{k\in\mathbb{S} \\ k\neq j}} P_{i,k} f_{k,j}, \qquad i,j\in\mathbb{S}.
\end{aligned}
$$

d) Let $\tilde{f}$ denote another solution of (5.4.17). We have $\tilde{f}_{i,j} \geqslant P_{i,j} = f_{i,j}^{(1)}$, and if $\tilde{f}_{i,j} \geqslant \sum_{l=1}^n f_{i,j}^{(l)}$ then by (5.4.16) and (5.4.17) we have

$$
\begin{aligned}
\tilde{f}_{i,j} &= P_{i,j} + \sum_{\substack{k\in\mathbb{S} \\ k\neq j}} P_{i,k} \tilde{f}_{k,j} \\
&\geqslant P_{i,j} + \sum_{\substack{k\in\mathbb{S} \\ k\neq j}} P_{i,k} \sum_{l=1}^n f_{k,j}^{(l)} \\
&= P_{i,j} + \sum_{l=1}^n \sum_{\substack{k\in\mathbb{S} \\ k\neq j}} P_{i,k} f_{k,j}^{(l)} \\
&= P_{i,j} + \sum_{l=1}^n f_{i,j}^{(l+1)} \\
&= P_{i,j} + \sum_{l=2}^{n+1} f_{i,j}^{(l)} \\
&= \sum_{l=1}^{n+1} f_{i,j}^{(l)}
\end{aligned}
$$

hence by induction we obtain

$$
\tilde{f}_{i,j} \geqslant \sum_{l=1}^n f_{i,j}^{(l)}, \quad i,j\in\mathbb{S}, \quad n\geqslant 1,
$$

and letting n tend to infinity, we find

$$
\tilde{f}_{i,j} \geqslant \sum_{l=1}^\infty f_{i,j}^{(l)} = f_{i,j}, \quad i,j\in\mathbb{S}.
$$

Finally, we check that if f and g are two minimal solutions then $f \geqslant g$ and $g \geqslant f$, hence $f = g$ and the minimal solution is unique.

e) The condition $g_{i,j}^{(1)} = f_{i,j}^{(1)}$ is satisfied by construction, for $i, j \in \mathbb{S}$. Next, assuming that $g_{i,j}^{(n)} = n f_{i,j}^{(n)}$, $i, j \in \mathbb{S}$, we have

$$\begin{aligned} g_{i,j}^{(n+1)} &= f_{i,j}^{(n+1)} + n \sum_{\substack{k \in \mathbb{S} \\ k \neq j}} P_{i,k} f_{k,j}^{(n)} \\ &= f_{i,j}^{(n+1)} + n f_{k,j}^{(n+1)} \\ &= (n+1) f_{i,j}^{(n+1)}, \quad i, j \in \mathbb{S}, \quad n \geqslant 1. \end{aligned}$$

f) We have

$$\begin{aligned} h_{i,j} &= \sum_{n \geqslant 1} g_{i,j}^{(n)} \\ &= g_{i,j}^{(1)} + \sum_{n \geqslant 1} g_{i,j}^{(n+1)} \\ &= f_{i,j}^{(1)} + \sum_{n \geqslant 1} \left(f_{i,j}^{(n+1)} + n \sum_{\substack{k \in \mathbb{S} \\ k \neq j}} P_{i,k} f_{k,j}^{(n)} \right) \\ &= \sum_{n \geqslant 1} f_{i,j}^{(n)} + \sum_{\substack{k \in \mathbb{S} \\ k \neq j}} P_{i,k} \sum_{n \geqslant 1} n f_{k,j}^{(n)} \\ &= f_{i,j} + \sum_{\substack{k \in \mathbb{S} \\ k \neq j}} P_{i,k} h_{k,j}, \qquad i, j \in \mathbb{S}. \end{aligned}$$

g) By (5.4.18), for $n = 1$ we have

$$\begin{aligned} \tilde{h}_{i,j} &= f_{i,j} + \sum_{\substack{k \in \mathbb{S} \\ k \neq j}} P_{i,k} \tilde{h}_{k,j} \\ &\geqslant f_{i,j} \\ &\geqslant f_{i,j}^{(1)} \\ &= g_{i,j}^{(1)}. \end{aligned}$$

Next, assuming that

$$\tilde{h}_{i,j} \geqslant \sum_{l=1}^{n} g_{i,j}^{(l)}, \qquad i, j \in \mathbb{S},$$

holds at the rank $n \geqslant 1$, we have

$$\begin{aligned}
\tilde{h}_{i,j} &= f_{i,j} + \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k}\tilde{h}_{k,j} \\
&\geqslant f_{i,j} + \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k}\sum_{l=1}^{n} g_{k,j}^{(l)} \\
&= f_{i,j} + \sum_{l=1}^{n}\sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} g_{k,j}^{(l)} \\
&= f_{i,j} + \sum_{l=1}^{n} l \sum_{\substack{k\in\mathbb{S}\\ k\neq j}} P_{i,k} f_{k,j}^{(l)} \\
&= f_{i,j} + \sum_{l=1}^{n} l f_{i,j}^{(l+1)} \\
&= f_{i,j} + \sum_{l=2}^{n+1} (l-1) f_{i,j}^{(l)} \\
&\geqslant \sum_{l=1}^{n+1} f_{i,j}^{(l)} + \sum_{l=2}^{n+1} (l-1) f_{i,j}^{(l)} \\
&= f_{i,j}^{(1)} + \sum_{l=2}^{n+1} l f_{i,j}^{(l)} \\
&= \sum_{l=1}^{n+1} g_{i,j}^{(l)}, \qquad i,j\in\mathbb{S}.
\end{aligned}$$

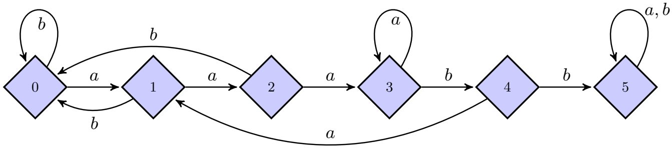
Letting n tend to infinity, we find

$$\tilde{h}_{i,j} \geqslant \sum_{l=1}^{\infty} g_{i,j}^{(l)} = h_{i,j}, \qquad i,j\in\mathbb{S},$$

proving that $h_{i,j}$ is a minimal solution to (5.4.18). Finally, we check that if f and g are two minimal solutions then $f \geqslant g$ and $g \geqslant f$, hence $f = g$ and the minimal solution is unique.

Problem 5.36 This problem is based on a simplified version of questions considered in Gusev (2014).

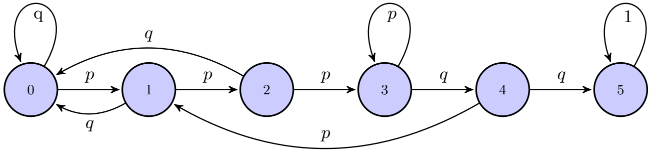
a) We find



The smallest integers are $l = 3$, $m = 2$ and the word is a^3b^2.

b) This language can be denoted by $\Sigma^* a^3 b^2 \Sigma^*$. An example of five-letter word is *aabbb*.

c) The process $(Y_k)_{k\in\mathbb{N}}$ is clearly a Markov chain because given Y_k, the distribution of $Y_{k+1} := f(X_{k+1}, Y_k)$ is independent of $Y_0, \ldots, Y_{k-1}$. The graph of the chain $(Y_k)_{k\in\mathbb{N}}$ is



The transition matrix of the chain $(Y_k)_{k\in\mathbb{N}}$ is

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 5} = \begin{bmatrix} q & p & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & 0 \\ q & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & p & q & 0 \\ 0 & p & 0 & 0 & 0 & q \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

d) Denoting by $h_5(k)$ the time it takes to reach state $\textcircled{5}$ starting from state $k = 0, 1, 2, 3, 4, 5$, we have the equations

$$\begin{cases} h_5(0) = 1 + qh_5(0) + ph_5(1) \\ h_5(1) = 1 + qh_5(0) + ph_5(2) \\ h_5(2) = 1 + qh_5(0) + ph_5(3) \\ h_5(3) = 1 + ph_5(3) + qh_5(4) \\ h_5(4) = 1 + ph_5(1) + qh_5(5) \\ h_5(5) = 0, \end{cases}$$

i.e.

$$\begin{cases} ph_5(0) = 1 + ph_5(1) \\ h_5(1) = 1 + qh_5(0) + ph_5(2) \\ h_5(2) = 1 + qh_5(0) + ph_5(3) \\ qh_5(3) = 1 + qh_5(4) = 1 + qph_5(0) \\ h_5(4) = 1 + ph_5(1) = ph_5(0) \\ h_5(5) = 0, \end{cases}$$

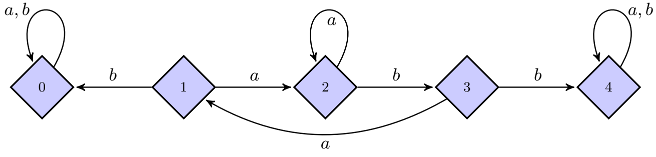
i.e.

$$\begin{cases} h_5(0) = \dfrac{1}{p} + \dfrac{1}{p^2} + \dfrac{1}{p^3} + h_5(3) \\ h_5(1) = \dfrac{1}{p^2} + \dfrac{1}{p^3} + h_5(3) \\ h_5(2) = \dfrac{1}{p^3} + h_5(3) \\ h_5(3) = \dfrac{1}{q} + ph_5(0) \\ h_5(4) = ph_5(0) \\ h_5(5) = 0, \end{cases}$$

i.e.

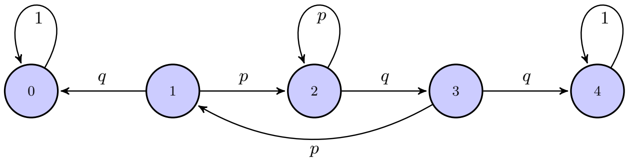
$$\begin{cases} h_5(0) = \dfrac{q(p^2+p+1)+p^3}{p^3q^2} = \dfrac{1}{p^3q^2} \\ h_5(1) = \dfrac{1}{p^3q^2} - \dfrac{1}{p} \\ h_5(2) = \dfrac{1}{p^3q^2} - \dfrac{1}{p} - \dfrac{1}{p^2} \\ h_5(3) = \dfrac{1}{q} + \dfrac{1}{p^2q^2} \\ h_5(4) = \dfrac{1}{p^2q^2} \\ h_5(5) = 0. \end{cases}$$

e) We find



f) The synchronizing words are "*abab*" and "*aabb*".

g) The graph of the chain $(Z_k)_{k\in\mathbb{N}}$ is



The transition matrix of the chain $(Z_k)_{k\in\mathbb{N}}$ is

$$[\ P_{i,j}\ ]_{0\leqslant i,j\leqslant 4} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & 0 & p & q & 0 \\ 0 & p & 0 & 0 & q \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

h) Denoting by $g_0(k)$ the probability that state ⓪ is reached first starting from state $k = 0,1,2,3,4$, we have the equations

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = qg_0(0) + pg_0(2) = q + pg_0(2) \\ g_0(2) = pg_0(2) + qg_0(3) \\ g_0(3) = pg_0(1) + qg_0(4) = pg_0(1) \\ g_0(4) = 0, \end{cases}$$

i.e.

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = q + pg_0(2) \\ g_0(2) = pg_0(2) + qg_0(3) = pg_0(2) + qpg_0(1) \\ g_0(3) = pg_0(1) \\ g_0(4) = 0, \end{cases}$$

i.e.

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = q + p^2 g_0(1) \\ g_0(2) = pg_0(1) \\ g_0(3) = pg_0(1) \\ g_0(4) = 0, \end{cases}$$

i.e.

$$\begin{cases} g_0(0) = 1 \\ g_0(1) = \dfrac{q}{1-p^2} = \dfrac{1}{1+p} \\ g_0(2) = \dfrac{pq}{1-p^2} = \dfrac{p}{1+p} \\ g_0(3) = \dfrac{pq}{1-p^2} = \dfrac{p}{1+p} \\ g_0(4) = 0. \end{cases}$$

The probability $g_0(1)$ can also be computed by pathwise analysis and a geometric series, as

$$g_0(1) = 1 - pq \sum_{k \geqslant 0} p^{2k} = 1 - \frac{pq}{1-p^2} = 1 - \frac{p}{1+p} = \frac{1}{1+p}.$$

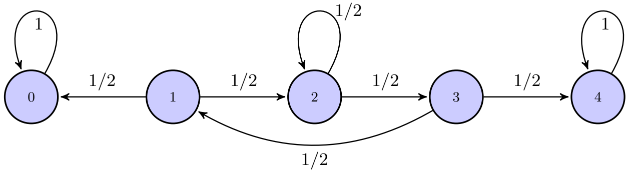
Now, starting from state ⓵ one may move directly to state ⓪ with probability q, in which case the first synchronized word is "b", not "abab". For this reason we need to subtract q from $g_0(1)$, and the probability that the first synchronized word is "abab" starting from state ⓵ is $1/(1+p) - (1-p) = p^2/(1+p)$.

Note that the above computations apply only when $p \in [0, 1)$. In case $p = 1$ the problem admits a trivial solution since the word "abab" will never occur.

Exercise 5.37

a) The word "abb" synchronizes to state ⓸ starting from states ⓵ and ⓶. However, the unique shortest word that synchronizes to state ⓸ starting from all states ⓵, ⓶ and ⓷ is "*aabb*".

b) The process $(Z_k)_{k\in\mathbb{N}}$ is a Markov chain on the state space $\{0, 1, 2, 3, 4\}$, with the following graph:



The transition matrix of the chain $(Z_k)_{k\in\mathbb{N}}$ is

$$\left[\ P_{i,j}\ \right]_{0\leqslant i,j\leqslant 4} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

c) Denoting by $g_4(k)$ the probability that state ⓸ is reached first starting from state $k = 0, 1, 2, 3, 4$, we have the equations

$$\begin{cases} g_4(0) = 0 \\ g_4(1) = \frac{1}{2}g_4(0) + \frac{1}{2}g_4(2) = \frac{1}{2}g_4(2) \\ g_4(2) = \frac{1}{2}g_4(2) + \frac{1}{2}g_4(3) \\ g_4(3) = \frac{1}{2}g_4(1) + \frac{1}{2}g_4(4) = \frac{1}{2}g_4(1) + \frac{1}{2} \\ g_4(4) = 1, \end{cases}$$

with the solution

$$\begin{cases} g_4(0) = 0 \\ g_4(1) = \frac{1}{3} \\ g_4(2) = \frac{2}{3} \\ g_4(3) = \frac{2}{3} \\ g_4(4) = 1. \end{cases}$$

Hence the probability that the first synchronized word is "aabb" when the automaton is started from state ① is 1/3.

Exercise 5.38

a) The unique shortest word that synchronizes to state ④ starting from all states ①, ② and ③ is "*aba*".
b) By the same analysis as in Exercise 5.37-(c), the probability that the first synchronized word is "aba" when the automaton is started from state ① is 1/3.

Problem 5.39

a) We have

$$\begin{aligned} V(k) &= \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\middle|\; X_0 = k\right] \\ &= \sum_{m\in \mathbb{S}} P_{k,m}\left(R(k) + \mathbb{E}\left[\sum_{n\geqslant 1} R(X_n) \;\middle|\; X_1 = m\right]\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{m\in\mathbb{S}} P_{k,m}R(k) + \sum_{m\in\mathbb{S}} P_{k,m}\mathbb{E}\left[\sum_{n\geqslant 1} R(X_n) \;\Big|\; X_1 = m\right] \\
&= R(k)\sum_{m\in\mathbb{S}} P_{k,m} + \sum_{m\in\mathbb{S}} P_{k,m}\mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \;\Big|\; X_0 = m\right] \\
&= R(k) + \sum_{m\in\mathbb{S}} P_{k,m}V(m), \qquad k\in\mathbb{S}.
\end{aligned}
$$

b) b1) We have

$$
Q^*(7,\downarrow) = 0, \quad Q^*(7,\rightarrow) = 0, \quad Q^*(6,\downarrow) = 5, \quad Q^*(6,\rightarrow) = 5,
$$

and $Q^*(3,\downarrow) = 4$. Regarding $Q^*(3,\rightarrow)$, we have

$$
Q^*(3,\rightarrow) = -1 + \mathrm{Max}\big(Q^*(3,\downarrow), Q^*(3,\rightarrow)\big),
$$

which implies

$$
Q^*(3,\rightarrow) < Q^*(3,\downarrow),
$$

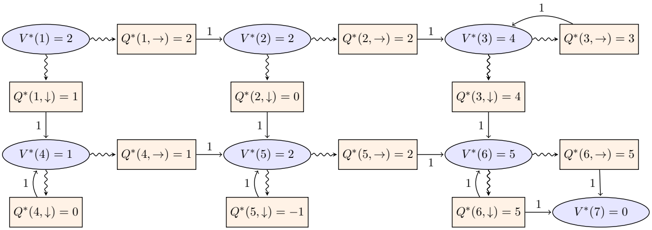
hence

$$
Q^*(3,\rightarrow) = -1 + Q^*(3,\downarrow) = 3.
$$

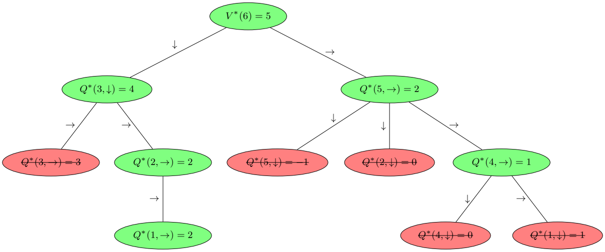
Similarly, we find

$$
\begin{cases}
Q^*(5,\downarrow) = -1, & Q^*(5,\rightarrow) = 2, \\
Q^*(2,\downarrow) = 0, & Q^*(2,\rightarrow) = 2, \\
Q^*(4,\downarrow) = 0, & Q^*(4,\rightarrow) = 1, \\
Q^*(1,\downarrow) = 1, & Q^*(1,\rightarrow) = 2.
\end{cases}
$$

The optimal action-value functional $Q^*(k,a)$ can be summarized as follows:



This solution can also be obtained by backward optimization (or dynamic programming):



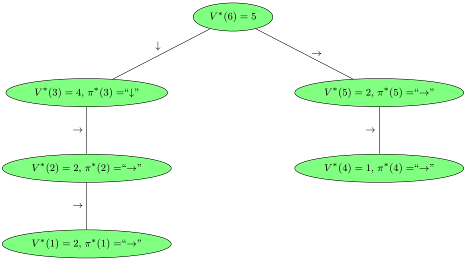
b2) At every state (k) we have

$$V^*(k) = \text{Max}\big(Q^*(k,\downarrow), Q^*(k,\rightarrow)\big),$$

hence

$$\begin{cases} V^*(7) = 0, \\ V^*(6) = 5, \\ V^*(3) = 4, \\ V^*(5) = 2, \\ V^*(2) = 2, \\ V^*(4) = 1, \\ V^*(1) = 2. \end{cases}$$

We obtain the following backward optimization tree:



b3) We find

$$\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5), \pi^*(6), \pi^*(7)) = (\rightarrow, \rightarrow, \downarrow, \rightarrow, \rightarrow, \Rsh, \Rsh),$$

which is consistent with the MDPtoolbox output:

```r
install.packages("MDPtoolbox")
library(MDPtoolbox)
P <- array(0, c(7, 7, 2))
P[,,1] <- matrix(c(0,0,0,1,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,1,0,
                   0,0,0,1,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,0,1,
                   0,0,0,0,0,0,1), nrow=7, ncol=7, byrow=TRUE)
P[,,2] <- matrix(c(0,1,0,0,0,0,0,
                   0,0,1,0,0,0,0,
                   0,0,1,0,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,1,0,
                   0,0,0,0,0,0,1,
                   0,0,0,0,0,0,1), nrow=7, ncol=7, byrow=TRUE)
R <- array(0, c(7, 2))
R[,1] <- matrix(c(0, -2, -1, -1, -3, 5, 0), nrow=1, ncol=7, byrow=TRUE)
R[,2] <- R[,1]
mdp_check(P, R)
mdp_value_iteration(P,R,discount=1)
$V
[1] 2 2 4 1 2 5 0
$policy
[1] 2 2 1 2 2 1 1
```

c) c1) We have

$$
\begin{aligned}
Q^*(k,a) &:= \operatorname*{Max}_{\pi\,:\,\pi(k)=a} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \,\Big|\, X_0 = k\right] \\
&= \operatorname*{Max}_{\pi\,:\,\pi(k)=a} \mathbb{E}\left[R(X_0) + \sum_{n\geqslant 1} R(X_n) \,\Big|\, X_0 = k\right] \\
&= \operatorname*{Max}_{\pi\,:\,\pi(k)=a} \mathbb{E}\left[R(k) + \sum_{n\geqslant 1} R(X_n) \,\Big|\, X_0 = k\right] \\
&= \sum_{l\in\mathbb{S}} P^{(a)}_{k,l}\left(R(k) + \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 1} R(X_n) \,\Big|\, X_1 = l\right]\right) \\
&= R(k) + \sum_{l\in\mathbb{S}} P^{(a)}_{k,l} \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_{n+1}) \,\Big|\, X_1 = l\right] \\
&= R(k) + \sum_{l\in\mathbb{S}} P^{(a)}_{k,l} \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \,\Big|\, X_0 = l\right] \\
&= R(k) + \sum_{l\in\mathbb{S}} P^{(a)}_{k,l} V^*(l), \qquad k \in \mathbb{S}.
\end{aligned}
$$

Similarly, the value function $Q^\pi(k,a)$ for a given policy π by setting the first action at state ⓚ to a satisfies the equation

$$Q^\pi(k,a) = R(k) + \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^\pi(l), \qquad k \in \mathbb{S}. \tag{S.44}$$

c2) The optimal policy $\pi_k^* \in \mathbb{A}$ at state $k \in \mathbb{S}$ is given by

$$\pi_k^* = \mathrm{argmax}_{a\in\mathbb{A}} Q^*(k,a) = \mathrm{argmax}_{a\in\mathbb{A}} \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^*(l).$$

c3) We have

$$\begin{aligned} V^*(k) &= \operatorname*{Max}_{\pi} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \,\Big|\, X_0 = k\right] \\ &= \operatorname*{Max}_{a\in\mathcal{A}} \operatorname*{Max}_{\pi\,:\,\pi(k)=a} \mathbb{E}\left[\sum_{n\geqslant 0} R(X_n) \,\Big|\, X_0 = k\right] \\ &= \operatorname*{Max}_{a\in\mathcal{A}} Q^*(k,a) \\ &= \operatorname*{Max}_{a\in\mathcal{A}} \left(R(k) + \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^*(l)\right) \\ &= R(k) + \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^*(l), \qquad k \in \mathbb{S}. \end{aligned}$$

This also implies

$$\begin{aligned} V^*(k) &= R(k) + \operatorname*{Max}_{a\in\mathcal{A}} \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^*(l) \\ &= \operatorname*{Max}_{a\in\mathcal{A}} \left(R(k) + \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} V^*(l)\right) \\ &= \operatorname*{Max}_{a\in\mathcal{A}} Q^*(k,a), \qquad k \in \mathbb{S}. \end{aligned} \tag{S.45}$$

Similarly, the value function $V^\pi(k)$ for a given policy satisfies the equation

$$V^\pi(k) = R(k) + \sum_{l\in\mathbb{S}} P_{k,l}^{(\pi(k))} V^\pi(l) \qquad k \in \mathbb{S}, \tag{S.46}$$

and the Bellman equation for $Q^*(k,a)$ can also be rewritten as

$$Q^*(k,a) = R(k) + \sum_{l\in\mathbb{S}} P_{k,l}^{(a)} \operatorname*{Max}_{b\in\mathcal{A}} Q^*(k,b), \qquad k \in \mathbb{S}.$$

In order to solve the MDP problem, starting from an arbitrary initial policy choice π, one can:

i) Compute V^{π} by applying (S.46) iteratively.

ii) Deduce the value of $Q^{\pi}(k, a)$ for every state $\textcircled{k}$ and action a from (S.44).

iii) For every state $\textcircled{k}$ and action a, based on (S.45), choose to update π with $\pi(k) := a$ if $Q^{\pi}(k, a) > V^{\pi}(k)$.

iv) Repeat the above iteratively.

d) d1) Similarly to Question (b1), we have

$$\begin{cases} Q^*(7, \downarrow) = 0, \quad Q^*(7, \rightarrow) = 0, \\ Q^*(6, \downarrow) = 5, \quad Q^*(6, \rightarrow) = 5, \\ Q^*(3, \downarrow) = 4, \quad Q^*(3, \rightarrow) = 3, \\ Q^*(5, \downarrow) = -1, \, Q^*(5, \rightarrow) = 2. \end{cases}$$

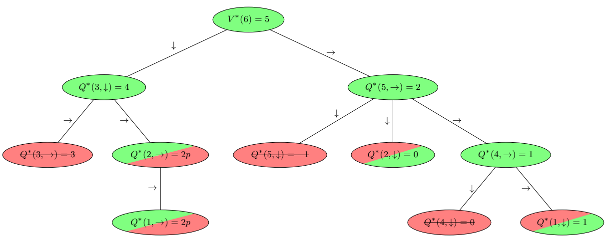
We also have $Q^*(2, \downarrow) = 0$ and

$$\begin{aligned} Q^*(2, \rightarrow) &= -2 + p \operatorname{Max}\big(Q^*(3, \downarrow), Q^*(3, \rightarrow)\big) + q \operatorname{Max}\big(Q^*(5, \rightarrow), Q^*(5, \rightarrow)\big) \\ &= -2 + pQ^*(3, \downarrow) + qQ^*(5, \rightarrow) \\ &= -2 + 4p + 2q = 2p, \end{aligned}$$

and

$$Q^*(4, \downarrow) = 0, \quad Q^*(4, \rightarrow) = 1, \quad Q^*(1, \downarrow) = 1, \quad Q^*(1, \rightarrow) = Q^*(2, \rightarrow) = 2p.$$

In other words, we have the following backward optimization (or dynamic programming) graph:



d2) At every state $\textcircled{k}$ we have

$$V^*(k) = \text{Max}\big(Q^*(k,\downarrow), Q^*(k,\rightarrow)\big),$$

hence

$$\begin{cases} V^*(7) = 0, \\ V^*(6) = 5, \\ V^*(5) = 2, \\ V^*(4) = 1, \\ V^*(3) = 4, \\ V^*(2) = 2p, \\ V^*(1) = \text{Max}(2p, 1). \end{cases}$$

d3) When $p = 0$ we find

$$\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5), \pi^*(6), \pi^*(7)) = (\downarrow, \overset{\rightarrow}{\downarrow}, \downarrow, \rightarrow, \rightarrow, \overset{\rightarrow}{\Downarrow}, \overset{\rightarrow}{\Downarrow})$$

which is consistent with the MDPtoolbox output:

```
$V
[1] 1 0 4 1 2 5 0
$policy
[1] 1 1 1 2 2 1 1
```

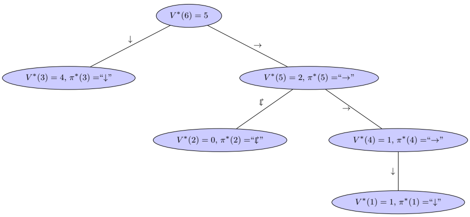


Fig. S.16: Optimal value function with $p = 0$.

When $0 < p < 1/2$ we obtain

$$\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5), \pi^*(6), \pi^*(7)) = (\downarrow, \rightarrow, \downarrow, \rightarrow, \rightarrow, \overset{\rightarrow}{\Downarrow}, \overset{\rightarrow}{\Downarrow})$$

which is consistent with the MDPtoolbox output, here with $p = 0.25$:

```
$V
[1] 1.0 0.5 4.0 1.0 2.0 5.0 0.0
$policy
[1] 1 2 1 2 2 1 1
```

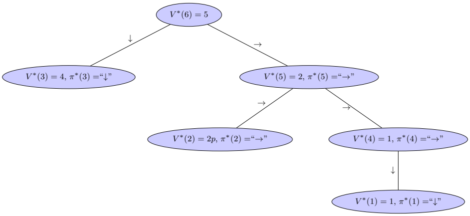


Fig. S.17: Optimal value function with $0 < p < 1/2$.

When $p = 1/2$ we find

$$\pi^* = (\pi^*(1), \pi^*(2), \pi^*(3), \pi^*(4), \pi^*(5), \pi^*(6), \pi^*(7)) = (\vec{\Downarrow}, \rightarrow, \downarrow, \rightarrow, \rightarrow, \vec{\Downarrow}, \vec{\Downarrow})$$

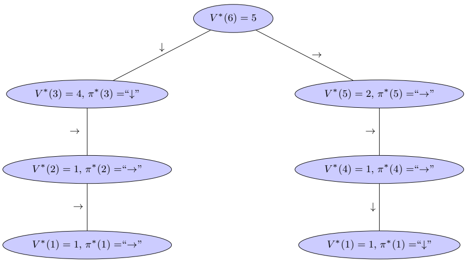


Fig. S.18: Optimal value function with $p = 1/2$.

which is consistent with the MDPtoolbox output, with $p = 0.5$:

```
$V
[1] 1 1 4 1 2 5 0
$policy
[1] 1 2 1 2 2 1 1
```

When $1/2 < p \leqslant 1$, we obtain

$$\pi^* = (\pi^*(1), \pi^*(1), \pi_3^*, \pi_4^*, \pi_5^*, \pi_6^*, \pi_7^*) = (\rightarrow, \rightarrow, \downarrow, \rightarrow, \rightarrow, \overset{\rightarrow}{\downarrow}, \overset{\rightarrow}{\downarrow}).$$

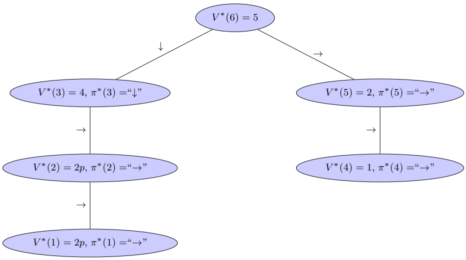


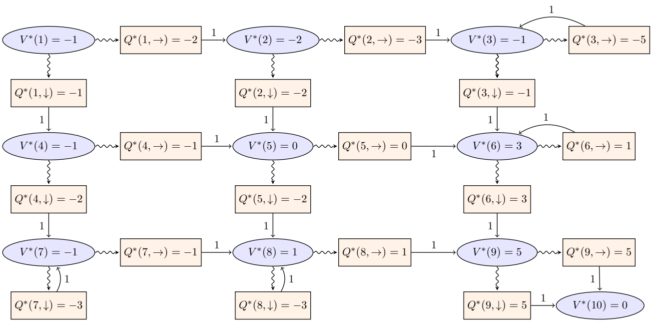
Fig. S.19: Optimal value function with $1/2 < p \leqslant 1$.

which is also consistent with the following MDPtoolbox output, here with $p = 0.75$:

```
library(MDPtoolbox)
P <- array(0, c(7, 7, 2));p=0.75;q=1-p
P[,,1] <- matrix(c(0,0,0,1,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,1,0,
                   0,0,0,1,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,0,1,
                   0,0,0,0,0,0,1), nrow=7, ncol=7, byrow=TRUE)
P[,,2] <- matrix(c(0,1,0,0,0,0,0,
                   0,0,p,0,q,0,0,
                   0,0,1,0,0,0,0,
                   0,0,0,0,1,0,0,
                   0,0,0,0,0,1,0,
                   0,0,0,0,0,0,1,
                   0,0,0,0,0,0,1), nrow=7, ncol=7, byrow=TRUE)
R <- array(0, c(7, 2))
R[,1] <- matrix(c(0, -2, -1, -1, -3, 5, 0), nrow=1, ncol=7, byrow=TRUE);R[,2] <- R[,1]
mdp_check(P, R);mdp_value_iteration(P,R,discount=1)
$V
[1] 1.5 1.5 4.0 1.0 2.0 5.0 0.0
$policy
[1] 2 2 1 2 2 1 1
```

Exercise 5.40

a) The optimal action-value functional $Q^*(k, a)$ is obtained as follows:



b) The optimal value function $V^*(k)$, $k = 1, 2, \ldots, 9$, is given in the next table.

① $V^*(1) = -1$	② $V^*(2) = -2$	③ $V^*(3) = -1$
④ $V^*(4) = -1$	⑤ $V^*(5) = 0$	⑥ $V^*(6) = +3$
⑦ $V^*(7) = -1$	⑧ $V^*(8) = 1$	⑨ $V^*(9) = +5$

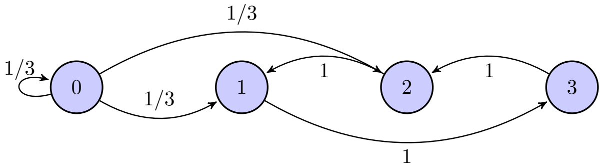
c) The optimal policy $\pi^*(k) \in \{\rightarrow, \downarrow\}$, $k = 1, 2, \ldots, 9$, is given as follows.

① $\pi^*(1) = \downarrow$	② $\pi^*(2) = \downarrow$	③ $\pi^*(3) = \downarrow$
④ $\pi^*(4) = \rightarrow$	⑤ $\pi^*(5) = \rightarrow$	⑥ $\pi^*(6) = \downarrow$
⑦ $\pi^*(7) = \rightarrow$	⑧ $\pi^*(8) = \rightarrow$	⑨ $\pi^*(9) =$ ⥮

Chapter 6 - Classification of States

Exercise 6.1

a) The graph of the chain is



This chain is reducible because its state space can be partitioned into two communicating classes as $\mathbb{S} = \{0\} \cup \{1, 2, 3\}$.

b) State ⓪ has period 1 and states ①, ②, ③ have period 3.

c) We have

$$p_{0,0} = \mathbb{P}(T_0 < \infty \mid X_0 = 0) = \mathbb{P}(T_0 = 1 \mid X_0 = 0) = \frac{1}{3},$$

and

$$\mathbb{P}(T_0 = \infty \mid X_0 = 0) = 1 - \mathbb{P}(T_0 < \infty \mid X_0 = 0) = \frac{2}{3}.$$

We also have

$$\mathbb{P}(R_0 < \infty \mid X_0 = 0) = \mathbb{P}(T_0 = \infty \mid X_0 = 0) \sum_{n \geqslant 1} (\mathbb{P}(T_0 < \infty \mid X_0 = 0))^n$$
$$= \frac{2}{3} \sum_{n \geqslant 1} \left(\frac{1}{3}\right)^n = 1.$$

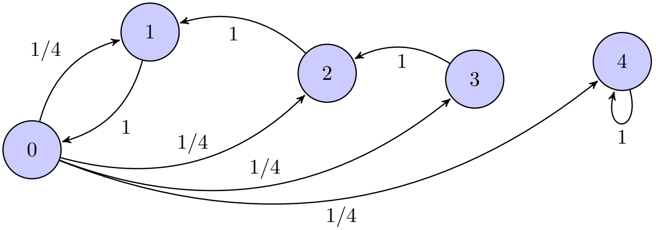
d) There are no absorbing states, state ⓪ is transient, and states ①, ②, ③ are recurrent by Corollary 6.7. State ⓪ is transient since $\mathbb{P}(R_0 < \infty \mid X_0 = 0) = 1$, as expected from (5.4.4).

Exercise 6.2

a) The chain is reducible and its communicating classes are $\{0, 1\}$ and $\{2\}$.
b) States ⓪ and ① are transient and the state ② is (positive) recurrent and absorbing.
c) All states have period 1.

Exercise 6.3

a) The chain has the following graph



b) All states ⓪, ①, ② and ③ have period 1, which can be obtained as the greatest common divisor (GCD) of $\{2, 3\}$ for states ⓪, ①, ② and $\{4, 6, 7\}$ for state ③. The chain is aperiodic.
c) State ④ is absorbing (and therefore recurrent), state ⓪ is transient because

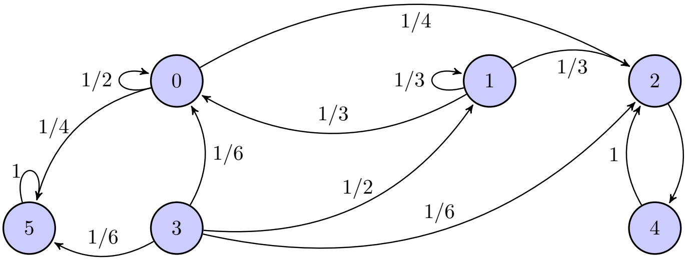
$$\mathbb{P}(T_0^r = \infty | X_0 = 0) \geqslant \frac{1}{4} > 0,$$

and the remaining states ①, ②, ③ are also transient because they communicate with the transient state ⓪, cf. Corollary 6.7. By pathwise or first step analysis, we can actually check that

$$\begin{aligned}
&\mathbb{P}(T_0^r < \infty | X_0 = 0)\\
&= \frac{1}{4}\big(\mathbb{P}(T_0^r < \infty | X_0 = 1) + \mathbb{P}(T_0^r < \infty | X_0 = 2) + \mathbb{P}(T_0^r < \infty | X_0 = 3)\big)\\
&= \frac{3}{4}.
\end{aligned}$$

d) The Markov chain is reducible because its state space $\mathbb{S} = \{0, 1, 2, 3, 4\}$ can be partitioned into two communicating classes $\{0, 1, 2, 3\}$ and $\{4\}$.

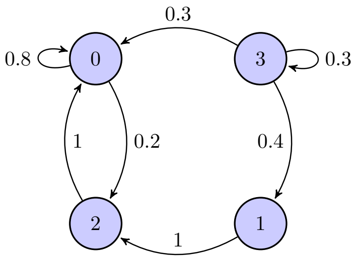
Exercise 6.4 The graph of the chain is



a) The chain is reducible and its communicating classes are $\{0\}$, $\{1\}$, $\{3\}$, $\{5\}$, and $\{2, 4\}$.
b) States ⓪, ①, ③ are transient and states ②, ④, ⑤ are recurrent.
c) State ③ has period 0, states ② and ④ have period 2, and states ⓪, ①, ⑤ are aperiodic.

Exercise 6.5

a) The graph of the chain is



This chain is reducible, with communicating classes $\{0,2\}$, $\{1\}$, and $\{3\}$.

b) States $\textcircled{0}$, $\textcircled{2}$, $\textcircled{3}$ have period 1, and state $\textcircled{1}$ has period 0. States $\textcircled{1}$ and $\textcircled{3}$ are transient, states $\textcircled{0}$ and $\textcircled{2}$ are recurrent by Theorem 6.11 and Corollary 6.7, and they are also positive recurrent since the state space is finite. There are no absorbing states.

Exercise 6.6

a) The communicating classes are $\{0\}$, $\{1,2\}$, $\{3,4,5\}$.
b) States $\textcircled{1}$, $\textcircled{2}$, $\textcircled{3}$, $\textcircled{4}$ and $\textcircled{5}$ are transient.
c) State $\textcircled{0}$ is recurrent.
d) State $\textcircled{0}$ is positive recurrent.
e) State $\textcircled{0}$ has period 1, states $\textcircled{1}$ and $\textcircled{2}$ have period 2, and states $\textcircled{3}$, $\textcircled{4}$ and $\textcircled{5}$ have period 3.

Exercise 6.7

a) This question refers to the Ehrenfest chain, cf. Example (iv) page 127. By (4.3.2) and (4.3.3) the transition matrix of this chain takes the form

$$\begin{bmatrix}
0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
1/N & 0 & (N-1)/N & 0 & \cdots & \cdots & 0 & 0 \\
0 & 2/N & 0 & (N-2)/N & \cdots & \cdots & 0 & 0 \\
0 & 0 & 3/N & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \vdots & 0 & 3/N & 0 & 0 \\
0 & 0 & \cdots & 0 & (N-2)/N & 0 & 2/N & 0 \\
0 & 0 & \cdots & 0 & 0 & (N-1)/N & 0 & 1/N \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0
\end{bmatrix}.$$

Note that the random choice made is that of a ball among N balls, not a box among two boxes.

b) All states have period 2 and are recurrent. There is no transient state.

Exercise 6.8

a) For the Markov chain of Exercise 4.10-(a) we find that state ⓪ is (positive) recurrent as it is absorbing, while every state ⓚ is transient when $k \geqslant 1$ since $\mathbb{P}(T_k^r < \infty \mid X_0 = k) = 0$, $k \geqslant 1$. State ⓪ is aperiodic, and all other states have period 0.

b) For the success runs chain of Exercise 4.10-(b), we have

$$\mathbb{P}(T_0^r < \infty \mid X_0) = q \sum_{n \geqslant 0} p^n = \frac{q}{1-p} = 1,$$

hence state ⓪ is recurrent. Since the chain is irreducible, it is recurrent by Corollary 6.7. State ⓪ is also positive recurrent as we have

$$\mathbb{E}[T_0^r \mid X_0] = q \sum_{n \geqslant 1} n p^{n-1} = \frac{q}{(1-p)^2} = \frac{1}{q}.$$

Next, by first step analysis we have

$$\begin{aligned} \mathbb{E}[T_k^r \mid X_0 = 0] &= \sum_{i=1}^{k} \mathbb{E}[T_i^r \mid X_0 = i-1] \\ &= \sum_{i=1}^{k} \big(q(1 + \mathbb{E}[T_{i-1}^r \mid X_0 = 0]) + p \big) \\ &= \sum_{i=1}^{k} \big(1 + q\mathbb{E}[T_{i-1}^r \mid X_0 = 0] \big) \\ &= k + q \sum_{i=1}^{k} \mathbb{E}[T_{i-1}^r \mid X_0 = 0], \qquad k \geqslant 1, \end{aligned}$$

which shows by induction that $\mathbb{E}[T_k^r \mid X_0 = 0] < \infty$, $k \geqslant 0$. On the other hand, we have

$$\begin{aligned} \mathbb{E}[T_k^r \mid X_0 = k] &= 1 + q\mathbb{E}[T_k^r \mid X_0 = 0] + p\mathbb{E}[T_k^r \mid X_0 = k+1] \\ &= 1 + q\mathbb{E}[T_k^r \mid X_0 = 0] + p\mathbb{E}[T_k^r \mid X_0 = k], \quad k \geqslant 0, \end{aligned}$$

since the starting point does not matter as long as we start from $k \geqslant 1$, hence

$$\mathbb{E}[T_k^r \mid X_0 = k] = \frac{1}{q} + \mathbb{E}[T_k^r \mid X_0 = 0], \qquad k \geqslant 1,$$

which shows that $\mathbb{E}[T_k^r \mid X_k = 0] < \infty$, $k \geqslant 0$, therefore the success runs Markov chain is also positive recurrent.

c) The chain is aperiodic because it is recurrent and state ⓪ has a returning loop.

Problem 6.9 By the same argument as in Exercise 3.2, by first step analysis we have

$$\begin{aligned}\mathbb{E}[T_0^r \mid X_0 = 1] &= \frac{1}{\alpha+1} \times 1 + \frac{\alpha}{\alpha+1} \times (1 + \mathbb{E}[T_0^r \mid X_0 = 2]) \\ &= \frac{1}{\alpha+1} \times 1 + \frac{\alpha}{\alpha+1} \times (1 + 2\mathbb{E}[T_0^r \mid X_0 = 1]) \\ &= 1 + \frac{2\alpha}{\alpha+1}\mathbb{E}[T_0^r \mid X_0 = 1],\end{aligned}$$

hence

$$\mathbb{E}[T_0^r \mid X_0 = 1] = \frac{1+\alpha}{1-\alpha}, \qquad \alpha < 1.$$

From the spatial homogeneity of the chain, we deduce that

$$\begin{aligned}\mathbb{E}[T_0^r \mid X_0 = k] &= \sum_{i=1}^{k} \mathbb{E}[T_{i-1}^r \mid X_0 = i] \\ &= k\mathbb{E}[T_0^r \mid X_0 = 1] \\ &= k\frac{1+\alpha}{1-\alpha}, \qquad k \geqslant 1, \quad \alpha < 1,\end{aligned}$$

and

$$\mathbb{E}[T_0^r \mid X_0 = 0] = 1 + \mathbb{E}[T_0^r \mid X_0 = 1] = 1 + \frac{1+\alpha}{1-\alpha} = \frac{2}{1-\alpha},$$

hence state ⓪ is positive recurrent if $\alpha < 1$. On the other hand, we have

$$\begin{aligned}\mathbb{E}[T_1^r \mid X_0 = 1] &= \frac{1}{\alpha+1} \times 2 + \frac{\alpha}{\alpha+1} \times (1 + \mathbb{E}[T_1^r \mid X_0 = 2]) \\ &= 1 + \frac{1}{\alpha+1} + \frac{\alpha}{\alpha+1}\mathbb{E}[T_0^r \mid X_0 = 1] \\ &= 1 + \frac{1}{\alpha+1} + \frac{\alpha}{1-\alpha} \\ &= \frac{2}{1-\alpha^2}.\end{aligned}$$

Similarly to the above, we have $\mathbb{E}[T_1^r \mid X_0 = 0] = 1$ and, for $k \geqslant 1$,

$$\begin{aligned}\mathbb{E}[T_{k+1}^r \mid X_0 = k] &= \frac{\alpha}{\alpha+1} \times 1 + \frac{1}{\alpha+1} \times (1 + \mathbb{E}[T_{k+1}^r \mid X_0 = k-1]) \\ &= 1 + \frac{1}{\alpha+1} \times (\mathbb{E}[T_k^r \mid X_0 = k-1] + \mathbb{E}[T_{k+1}^r \mid X_0 = k]),\end{aligned}$$

hence

$$\mathbb{E}[T^r_{k+1} \mid X_0 = k] = \frac{\alpha + 1 + \mathbb{E}[T^r_k \mid X_0 = k-1]}{\alpha}, \qquad k \geqslant 1,$$

which can be solved as

$$\mathbb{E}[T^r_{k+1} \mid X_0 = k] = \frac{2\alpha^{-k} - \alpha - 1}{1 - \alpha}, \qquad k \geqslant 0,$$

see *e.g.* the command

RSolve[f[k+1]==1+1/a+f[k]/a,f[0] == 1, f[k], k],

which shows that

$$\begin{aligned}
&\mathbb{E}[T^r_k \mid X_0 = k] \\
&= \frac{1}{\alpha+1}\left(1 + \mathbb{E}[T^r_k \mid X_0 = k-1]\right) + \frac{\alpha}{\alpha+1}\left(1 + \mathbb{E}[T^r_k \mid X_0 = k+1]\right) \\
&= 1 + \frac{1}{\alpha+1}\mathbb{E}[T^r_k \mid X_0 = k-1] + \frac{\alpha}{\alpha+1}\mathbb{E}[T^r_k \mid X_0 = k+1] \\
&= 1 + \frac{1}{\alpha+1}\mathbb{E}[T^r_k \mid X_0 = k-1] + \frac{\alpha}{1-\alpha} \\
&= \frac{1}{1-\alpha} + \frac{1}{\alpha+1}\mathbb{E}[T^r_k \mid X_0 = k-1] \\
&= \frac{2\alpha^{-k}}{1-\alpha^2},
\end{aligned}$$

therefore the chain is positive recurrent if and only if $\alpha < 1$.

Exercise 6.10

a) The number of cookies present in the considered region is kL.
b) The number of time steps is kL.
c) Let N denote the average number of time steps needed. From the relation $N(\tilde{p} - \tilde{q}) = L$ we deduce $N = L/(\tilde{p} - \tilde{q})$.
d) The condition is $kL \leqslant N = L/(\tilde{p} - \tilde{q})$, or $k \leqslant 1/(\tilde{p} - \tilde{q})$, which yields

$$\frac{1}{2} < \tilde{p} \leqslant \frac{1}{2}\left(1 + \frac{1}{k}\right).$$

e) Under the condition

$$\tilde{p} > \frac{1}{2}\left(1 + \frac{1}{k}\right)$$

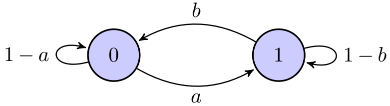
the amount of cookies consumed will remain strictly lower than the number of available cookies, thus ensuring the transience of the random walk.

Chapter 7 - Long-Run Behavior of Markov Chains

Exercise 7.1 Writing the condition $\pi P = \pi$ leads to the equations

$$\begin{cases} \dfrac{\pi_0}{3} + 2\dfrac{\pi_1}{3} = \pi_0 \\ \\ 2\dfrac{\pi_0}{3} + \dfrac{\pi_1}{3} = \pi_1 \end{cases}$$

i.e. $\pi_0 = \pi_1$. Combining this relation with the condition $\pi_0 + \pi_1 = 1$ shows that $\pi_0 = \pi_1 = 1/2$.



Using the general relation

$$[\pi_0, \pi_1] = \left[\frac{b}{a+b}, \frac{a}{a+b}\right],$$

with $(a,b) \neq (0,0)$ and $(a,b) \neq (1,1)$ for the two-state chain with transition matrix

$$P = \begin{bmatrix} 1-a & a \\ \\ b & 1-b \end{bmatrix}$$

yields the same answer

$$[\pi_0, \pi_1] = \left[\frac{1}{2}, \frac{1}{2}\right]$$

when $a = b$, in which case the matrix P is also *column-stochastic*. .

```
install.packages("igraph");install.packages("markovchain")
library("igraph");library(markovchain)
P<-matrix(c(1/3,2/3,2/3,1/3),nrow=2,byrow=TRUE);MC
      <-new("markovchain",transitionMatrix=P)
graph <- as(MC, "igraph")
plot(graph,vertex.size=50,edge.label.cex=2,edge.label=E(graph)$prob,edge.color='black',
      vertex.color='dodgerblue',vertex.label.cex=3)
steadyStates(object = MC)
      1 2
[1,] 0.5 0.5
```

Exercise 7.2

a) The chain is reducible, state ⓪ is aperiodic and all other states have period 0. On the other hand, we check that state ⓪ is positive recurrent and all other states are transient, cf. Exercise 6.8-(a). Therefore, the assumptions of Theorems 7.2, 7.8 and 7.10 are not satisfied.

b) Writing $\pi = \pi P$, *i.e.*

$$[\pi_0, \pi_1, \pi_2, \pi_3, \cdots] = [\pi_0, \pi_1, \pi_2, \pi_3, \cdots] \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ q & 0 & 0 & p & 0 & \cdots \\ q & 0 & 0 & 0 & p & \cdots \\ q & 0 & 0 & 0 & 0 & \cdots \\ \vdots & 0 & 0 & 0 & 0 & \ddots \end{bmatrix},$$

shows that

$$\begin{cases} \pi_0 = \pi_0 + q(\pi_1 + \pi_2 + \cdots) = \pi_0 + q(1 - \pi_0) = q + p\pi_0, \\ \pi_1 = 0, \\ \pi_2 = p\pi_1, \\ \pi_3 = p\pi_2, \\ \vdots \end{cases}$$

hence $\pi_k = 0$, $k \geqslant 1$, and the stationary distribution is given by

$$\pi = [1, 0, 0, 0, \ldots].$$

c) Taking the limit as n tends to infinity in (S.32), we find

$$\lim_{n \to \infty} P^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

hence the limiting distribution coincides with the stationary distribution π obtained in (b).

Exercise 7.3

a) The chain irreducible, aperiodic, recurrent and positive recurrent by Exercise 6.8-(c). Therefore, the assumptions of Theorems 7.2, 7.8 and 7.10 are all satisfied and the chain admits a limiting distribution which coincides with its stationary distribution.

b) Writing $\pi = \pi P$, *i.e.*

$$[\pi_0, \pi_1, \pi_2, \pi_3, \cdots] = [\pi_0, \pi_1, \pi_2, \pi_3, \cdots] \times \begin{bmatrix} q & p & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ q & 0 & 0 & p & 0 & \cdots \\ q & 0 & 0 & 0 & p & \cdots \\ q & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

shows that

$$\begin{cases} \pi_1 = p\pi_0, \\ \pi_2 = p\pi_1, \\ \pi_3 = p\pi_2, \\ \vdots \end{cases}$$

hence by induction we get $\pi_k = p^k \pi_0$, $k \geqslant 1$. By the condition $\sum_{k\geqslant 0} \pi_k = 1$ we find

$$1 = \sum_{k\geqslant 0} \pi_k = \pi_0 \sum_{k\geqslant 0} p^k = \frac{\pi_0}{1-p},$$

which shows that $\pi_0 = q$. We conclude that π is the geometric distribution

$$\pi_k = (1-p)p^k, \qquad k \geqslant 0.$$

c) Taking the limit as n tends to infinity in (S.33), we find, when $p < 1$,

$$\lim_{n\to\infty} P^n = \begin{bmatrix} q & qp & qp^2 & qp^3 & qp^4 & \cdots \\ q & qp & qp^2 & qp^3 & qp^4 & \cdots \\ q & qp & qp^2 & qp^3 & qp^4 & \cdots \\ q & qp & qp^2 & qp^3 & qp^4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

hence the chain admits the geometric distribution with parameter p as a limiting distribution, which coincides with the stationary distribution π obtained in (b).

Exercise 7.4

a) The Ehrenfest chain is irreducible and positive recurrent, and has period 2, therefore it satisfies the conditions of Theorem 7.10 and it admits a stationary distribution.
b) The stationary distribution of the Ehrenfest chain satisfies the equation $\pi P = \pi$, which reads

$$[\pi_0, \pi_1, \pi_2, \ldots] \times \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 1/N & 0 & (N-1)/N & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 2/N & 0 & (N-2)/N & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 3/N & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \vdots & 0 & 3/N & 0 & 0 \\ 0 & 0 & \cdots & 0 & (N-2)/N & 0 & 2/N & 0 \\ 0 & 0 & \cdots & 0 & 0 & (N-1)/N & 0 & 1/N \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$= [\pi_0, \pi_1, \pi_2, \ldots] = \pi,$$

cf. Exercise 6.7. This yields

$$\begin{cases} \pi_0 = \dfrac{\pi_1}{N} \\ \pi_1 = \pi_0 + \dfrac{2}{N}\pi_2, \\ \pi_2 = \pi_1 \dfrac{N-1}{N} + \dfrac{3}{N}\pi_3, \\ \pi_3 = \pi_2 \dfrac{N-2}{N} + \dfrac{4}{N}\pi_4, \\ \vdots \end{cases}$$

from which we deduce

$$\begin{cases} \pi_0 = \dfrac{\pi_1}{N} \\ \pi_1 = \dfrac{2}{N-1}\pi_2, \\ \pi_2 = \dfrac{3}{N-2}\pi_3, \\ \pi_3 = \dfrac{4}{N-3}\pi_4, \\ \vdots \end{cases}$$

hence

$$\pi_0 = \frac{1}{N} \times \frac{2}{N-1} \times \frac{3}{N-2} \times \cdots \times \frac{k}{N-k+1}\pi_k = \frac{k!(N-k)!}{N!}\pi_k,$$

or

$$\pi_k = \binom{N}{k}\pi_0, \qquad k = 0, \ldots, N.$$

In addition we have

$$1 = \sum_{k=0}^{N} \pi_k = \pi_0 \sum_{k=0}^{N} \binom{N}{k} = (1+1)^n \pi_0 = 2^N \pi_0,$$

hence $\pi_0 = 2^{-N}$ and $(\pi_k)_{k=0,1,\ldots,N}$ is the binomial distribution with parameter $(N, 1/2)$, *i.e.*

$$\pi_k = \binom{N}{k} \left(\frac{1}{2}\right)^N, \qquad k = 0, 1, \ldots, N. \tag{S.47}$$

Conversely, from (S.47) we can recover

$$\begin{aligned}
\sum_{l=0}^{N} \pi_l P_{l,k} &= \pi_{k-1} \frac{N-(k-1)}{N} + \pi_{k+1} \frac{k+1}{N} \\
&= \binom{N}{k-1} \left(\frac{1}{2}\right)^N \frac{N-k+1}{N} + \binom{N}{k+1} \left(\frac{1}{2}\right)^N \frac{k+1}{N} \\
&= \left(\frac{1}{2}\right)^N \frac{(N-1)!k}{(N-k)!k!} + \left(\frac{1}{2}\right)^N \frac{(N-1)!(N-k)}{(N-k)!k!} \\
&= \left(\frac{1}{2}\right)^N \binom{N}{k} \\
&= \pi_k, \qquad k = 1, 2, \ldots, N-1,
\end{aligned}$$

and similarly for $k = 0$ and $k = N$, showing that $\pi = \pi P$.

c) The powers of the transition matrix of the Ehrenfest chain can be computed using the following R code which shows that $(P^n)_{n \geqslant 1}$ admits two subsequences converging to two distinct limits, therefore $\lim_{n\to\infty} P^n$ does not exist and this chain does not admit a limiting distribution. Note that the conditions of Theorems 7.8 and 7.10 are not satisfied here, as this chain has period 2.

```
install.packages("expm");library(expm)
Ehrenfest <- function(n) {States <- c(0, seq(1,n))
  TPM <- matrix(0,nrow=length(States),ncol=length(States),dimnames=
                list(seq(0,n),seq(0,n)))
  tran_prob <- function(i,n) {
    tranRow <- rep(0,n+1)
    if(i==0) tranRow[2] <- 1
    if(i==n) tranRow[(n+1)-1] <- 1
    if(i!=0 & i!=n) {
      j=i+1
      tranRow[j-1] <- i/(n)
      tranRow[j+1] <- 1-i/(n)}
    return(tranRow)}
  for(j in 0:(n))TPM[j+1,] <- tran_prob(j,n)
  return(TPM)}
P=Ehrenfest(4)
P%^%1000
P%^%1001
P%^%1002
```

Exercise 7.5

a) The Bernoulli-Laplace chain is irreducible, aperiodic and positive recurrent. Therefore it satisfies all the assumptions of Theorems 7.2, 7.8 and 7.10, the chain admits a limiting distribution which coincides with its stationary distribution.

b) The equation $\pi = \pi P$ reads

$$\begin{cases} \pi_0 = \dfrac{\pi_1}{N^2} \\[2ex] \pi_1 = \pi_0 + \dfrac{2(N-1)}{N^2}\pi_1 + \dfrac{2^2}{N^2}\pi_2, \\[2ex] \pi_2 = \dfrac{(N-1)^2}{N^2}\pi_1 + \dfrac{4(N-2)}{N^2}\pi_2 + \dfrac{3^2}{N^2}\pi_3, \\[2ex] \pi_3 = \dfrac{(N-2)^2}{N^2}\pi_1 + \dfrac{6(N-3)}{N^2}\pi_2 + \dfrac{3^2}{N^2}\pi_3, \\ \vdots \end{cases}$$

from which we deduce

$$\begin{cases} \pi_0 = \dfrac{\pi_1}{N^2} \\ \\ \pi_1 = \dfrac{2^2}{(N-1)^2}\pi_2, \\ \\ \pi_2 = \dfrac{3^2}{(N-2)^2}\pi_3, \\ \\ \pi_3 = \dfrac{4^2}{(N-3)^2}\pi_4, \\ \vdots \end{cases}$$

hence

$$\pi_0 = \frac{1}{N^2} \times \frac{2^2}{(N-1)^2} \times \frac{3^2}{(N-2)^2} \times \cdots \times \frac{k^2}{(N-k+1)^2}\pi_k = \frac{k!^2(N-k)!^2}{N!^2}\pi_k,$$

or

$$\pi_k = \binom{N}{k}^2 \pi_0, \qquad k = 0, \ldots, N.$$

In addition we have

$$1 = \sum_{k=0}^{N} \pi_k = \pi_0 \sum_{k=0}^{N} \binom{N}{k}^2 = \binom{2N}{N}\pi_0,$$

hence $\pi_0 = 1/\binom{2N}{N}$ and $(\pi_k)_{k=0,1,\ldots,N}$ is the probability distribution given by

$$\pi_k = \frac{\binom{N}{k}^2}{\binom{2N}{N}}, \qquad k = 0, 1, \ldots, N.$$

c) The Bernoulli-Laplace chain admits a limiting distribution which coincides with its stationary distribution by Theorem 7.8.

Exercise 7.6

a) We have

$$\begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}.$$

b) By first step analysis, we have

$$\begin{cases} \mu_0(0) = 1 + \frac{1}{2}\mu_0(1) + \frac{1}{2}\mu_0(3) \\ \\ \mu_0(1) = 1 + \frac{1}{2}\mu_0(2) \\ \\ \mu_0(2) = 1 + \frac{1}{2}\mu_0(1) + \frac{1}{2}\mu_0(3) \\ \\ \mu_0(3) = 1 + \frac{1}{2}\mu_0(2). \end{cases}$$

By symmetry of the maze we have $\mu_0(1) = \mu_0(3)$, hence

$$\begin{cases} \mu_0(0) = \mu_0(2) = 1 + \mu_0(1) \\ \\ \mu_0(1) = \mu_0(3) = 1 + \frac{1}{2}\mu_0(2). \end{cases}$$

and

$$\mu_0(0) = 4, \quad \mu_0(1) = 3, \quad \mu_0(2) = 4, \quad \mu_0(3) = 3.$$

The symmetry of the problem shows that we have, $\mu_0(1) = \mu_0(3)$, which greatly simplifies the calculations. One has to pay attention to the fact that $\mu_0(0)$ is a return time, *not* a hitting time, so that $\mu_0(0)$ cannot be equal to zero.

c) Clearly, the probability distribution

$$(\pi_0, \pi_1, \pi_2, \pi_3) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

is invariant and satisfies the condition $\pi = \pi P$, see also Exercise 7.15.

Recall that the family $(\pi_0, \pi_1, \pi_2, \pi_3)$ is a *(probability) distribution* **if and only if**

(i) $\pi_0 \geqslant 0, \quad \pi_1 \geqslant 0, \quad \pi_2 \geqslant 0, \quad \pi_3 \geqslant 0,$

and

(ii)

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = \mathbf{1}.$$

In the present situation we have

$$\pi_0 = \pi_1 = \pi_2 = \pi_3$$

and

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = \mathbf{1} = 4\pi_0,$$

which necessarily yields

$$\pi_0 = \pi_1 = \pi_2 = \pi_3 = \frac{1}{4} = \frac{1}{\mu_0(0)}.$$

Exercise 7.7

a) Clearly, the transition from the current state to the next state depends only on the current state on the chain, hence the process is Markov. The transition matrix of the chain on the state space $\mathbb{S} = (D, N)$ is

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{bmatrix}.$$

b) The stationary distribution $\pi = (\pi_D, \pi_N)$ is solution of $\pi = \pi P$, *i.e.*

$$\begin{cases} \pi_D = (1-a)\pi_D + b\pi_N = \frac{1}{4}\pi_D + \frac{1}{4}\pi_N \\ \\ \pi_N = a\pi_D + (1-b)\pi_N = \frac{3}{4}\pi_D + \frac{3}{4}\pi_N \end{cases}$$

under the condition $\pi_D + \pi_N = 1$, which yields $\pi_D = b/(a+b) = 1/4$ and $\pi_N = a/(a+b) = 3/4$.

c) In the long run, by the Ergodic Theorem 7.12 we find that the fraction of distorted signals is $\pi_D = 1/4 = 25\%$.

d) The average time $h_N(D) = \mu_N(D)$ to reach state $\textcircled{N}$ starting from state $\textcircled{D}$ satisfies

$$\mu_N(D) = (1-a)(1 + \mu_N(D)) + a \tag{S.48}$$

hence $\mu_N(D) = 1/a = 4/3$.

Additional comments:

(i) The stationary distribution $[\pi_D, \pi_N]$ satisfies

$$[\pi_D, \pi_N] = \left[\frac{\mu_N(D)}{\mu_D(N) + \mu_N(D)}, \frac{\mu_D(N)}{\mu_D(N) + \mu_N(D)}\right],$$

as the ratios of times spent in states D and N, since $\mu_N(D) = 1/a$ and $\mu_D(N) = 1/b$.

(ii) The value of $\mu_N(D)$ may also be recovered as

$$\mu_N(D) = \frac{3}{4}\sum_{k\geqslant 1} k\left(\frac{1}{4}\right)^{k-1} = \frac{3}{4}\frac{1}{(1-1/4)^2} = \frac{4}{3},$$

cf. (A.4).

e) The average time $\mu_D(N)$ to reach state $\textcircled{D}$ starting from state $\textcircled{N}$ satisfies

$$\mu_D(N) = (1-b)(1+\mu_D(N)) + b \tag{S.49}$$

hence $\mu_N(D) = 1/b = 4$.

Additional comments:

(i) The value of $\mu_D(N)$ may also be recovered as

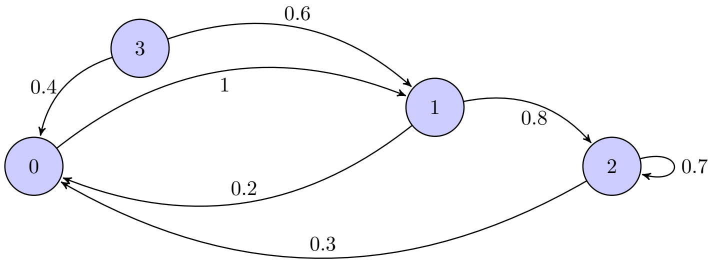
$$\mu_D(N) = \frac{1}{4}\sum_{k\geqslant 1} k\left(\frac{3}{4}\right)^{k-1} = \frac{1}{4}\frac{1}{(1-3/4)^2} = 4,$$

cf. (A.4).

(ii) The values of μ_D^D and μ_N^N can also be computed from $\mu_D^D = 1/\pi_D$ and $\mu_N^N = 1/\pi_N$.

Exercise 7.8

a) The chain has the following graph:



The chain is reducible and its communicating classes are $\{0,1,2\}$ and $\{3\}$.

b) State $\textcircled{3}$ is transient because $\mathbb{P}(T_3 = \infty \mid X_0 = 3) = 0.4 + 0.6 = 1$, cf. (6.3.1), and states $\textcircled{0}$, $\textcircled{1}$, $\textcircled{2}$ are recurrent by Theorem 6.11 and Corollary 6.7.

c) It suffices to consider the subchain on $\{0,1,2\}$ with transition matrix

$$\widetilde{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0.2 & 0 & 0.8 \\ 0.3 & 0 & 0.7 \end{bmatrix},$$

and to solve $\pi = \pi\widetilde{P}$, *i.e.*

$$\begin{cases} \pi_0 = 0.2\pi_1 + 0.3\pi_2 \\ \pi_1 = \pi_0 \\ \pi_2 = 0.8\pi_1 + 0.7\pi_2 \end{cases}$$

which yields $\pi_1 = \pi_0$ and $0.3\pi_2 = 0.8\pi_1 = 0.8\pi_0$, with

$$1 = \pi_0 + \pi_1 + \pi_2 = 2\pi_0 + \frac{8}{3}\pi_0,$$

i.e.

$$\pi_0 = \frac{3}{14}, \quad \pi_1 = \frac{3}{14}, \quad \pi_2 = \frac{4}{7},$$

and the fraction of time spent at state ⓪ in the long run is $3/14 \simeq 0.214$ as the limiting and stationary distributions coincide.

d) Letting $h_0(k)$ denote the mean hitting time of state ⓪ starting from state ⓚ, we have

$$\begin{cases} h_0(0) = 0 \\ h_0(1) = 0.2(1 + h_0(0)) + 0.8(1 + h_0(2)) \\ h_0(2) = 0.3(1 + h_0(0)) + 0.7(1 + h_0(2)) \\ h_0(3) = 0.4(1 + h_0(0)) + 0.6(1 + h_0(1)), \end{cases}$$

i.e.

$$\begin{cases} h_0(0) = 0, \\ \\ h_0(1) = 1 + 0.8h_0(2), \\ \\ h_0(2) = 1 + 0.7h_0(2), \\ \\ h_0(3) = 1 + 0.6h_0(1), \end{cases} \quad \text{or} \quad \begin{cases} h_0(0) = 0, \\ \\ h_0(1) = 1 + 0.8h_0(2), \\ \\ 0.3h_0(2) = 1, \\ \\ h_0(3) = 1 + 0.6h_0(1), \end{cases}$$

hence

$$h_0(0) = 0, \quad h_0(1) = \frac{11}{3}, \quad h_0(2) = \frac{10}{3}, \quad h_0(3) = \frac{16}{5},$$

and the mean time to reach state ⓪ starting from state ② is found to be equal to $h_0(2) = 10/3$, which can also be recovered by pathwise analysis and the geometric series

$$h_0(2) = 0.3\sum_{k\geqslant 1} k(0.7)^{k-1} = \frac{0.3}{(1-0.7)^2} = \frac{10}{3}.$$

Note that the value of $h_0(2)$ could also be computed by restriction to the sub-chain $\{0, 1, 2\}$, by solving

$$\begin{cases} h_0(0) = 0 \\ h_0(1) = 0.2(1 + h_0(0)) + 0.8(1 + h_0(2)) \\ h_0(2) = 0.3(1 + h_0(0)) + 0.7(1 + h_0(2)). \end{cases}$$

Exercise 7.9

a) First, we note that the chain has finite state space and it is irreducible, positive recurrent and aperiodic, hence by Theorem 7.8 its limiting distribution coincides with its stationary distribution which is the unique solution of $\pi = \pi P$. After calculations, this equation can be solved as

$$\pi_0 = c \times 161, \qquad \pi_1 = c \times 460, \qquad \pi_2 = c \times 320, \qquad \pi_3 = c \times 170,$$

see *here*. The condition

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3 = c \times 161 + c \times 460 + c \times 320 + c \times 170$$

shows that

$$c = \frac{1}{161 + 460 + 320 + 170} = \frac{1}{1111},$$

hence

$$\pi_0 = \frac{161}{1111}, \qquad \pi_1 = \frac{460}{1111}, \qquad \pi_2 = \frac{320}{1111}, \qquad \pi_3 = \frac{170}{1111}. \tag{S.50}$$

b) We choose to solve this problem using mean return times since $\mu_0(i) = h_0(i)$, $i = 1, 2, 3$, however it could also be solved using mean hitting times $h_i(j)$. We have

$$\begin{cases} \mu_0(0) = 1 + \mu_0(1), \\ \\ \mu_0(1) = 0.1 + 0.4(1 + \mu_0(1)) + 0.2(1 + \mu_0(2)) + 0.3(1 + \mu_0(3)) \\ \qquad\quad = 1 + 0.4\mu_0(1) + 0.2\mu_0(2) + 0.3\mu_0(3), \\ \\ \mu_0(2) = 0.2 + 0.2(1 + \mu_0(1)) + 0.5(1 + \mu_0(2)) + 0.1(1 + \mu_0(3)) \\ \qquad\quad = 1 + 0.2\mu_0(1) + 0.5\mu_0(2) + 0.1\mu_0(3), \\ \\ \mu_0(3) = 1 + 0.3\mu_0(1) + 0.4\mu_0(2), \end{cases}$$

hence

$$\mu_0(1) = \frac{950}{161}, \qquad \mu_0(2) = \frac{860}{161}, \qquad \mu_0(3) = \frac{790}{161},$$

see *here* or *here*. Note that the data of the first row in the transition matrix is not needed in order to compute the mean return times.

c) We find

$$\mu_0(0) = 1 + \mu_0(1) = 1 + \frac{950}{161} = \frac{161 + 950}{161} = \frac{1111}{161},$$

hence the relation $\pi_0 = 1/\mu_0(0)$ is satisfied from (S.50).

Exercise 7.10

a) All states of this chain have period 2.
b) The chain is irreducible and it has a finite state space, hence it is positive recurrent from Theorem 6.13. By Proposition 6.16, all states have period 2 hence the chain is not aperiodic, and for this reason Theorem 7.2 and Theorem 7.8 cannot be used and the chain actually has no limiting distribution. Nevertheless, Theorem 7.10 applies and shows that the equation $\pi = \pi P$ characterizes the stationary distribution. This equation reads

$$\begin{cases} \pi_0 = \frac{1}{4}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_1 = \frac{1}{2}\pi_0 + \frac{1}{3}\pi_2 \\ \pi_2 = \frac{3}{4}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_3 = \frac{1}{2}\pi_0 + \frac{2}{3}\pi_2, \end{cases}$$

hence

$$\pi_1 = \pi_0, \quad \pi_2 = 3\pi_0/2, \quad \pi_3 = 3\pi_0/2,$$

see *here*, under the condition

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1,$$

which yields the solution

$$\pi_0 = 2c, \quad \pi_1 = 2c, \quad \pi_2 = 3c, \quad \pi_3 = 3c,$$

and $c = 1/10$, *i.e*

$$\pi_0 = 20\%, \quad \pi_1 = 20\%, \quad \pi_2 = 30\%, \quad \pi_3 = 30\%.$$

Exercise 7.11 We choose to model the problem on the state space $\{1, 2, 3, 4\}$, meaning that the replacement of a component is immediate upon failure. Let X_n denote the remaining active time of the component at time n. Given that at time n there remains $X_n = k \geqslant 2$ units of time until failure, we know with certainty that at the next time step $n+1$ there will remain $X_{n+1} = k - 1 \geqslant 1$ units of time until failure. Hence at any time $n \geqslant 1$ we have

$$X_n = 4 \quad \Longrightarrow X_{n+1} = 3 \quad \Longrightarrow X_{n+2} = 2 \quad \Longrightarrow X_{n+3} = 1,$$

whereas when $X_n = 1$ the component will become inactive at the next time step and will be immediately replaced by a new component of random lifetime $T \in \{1, 2, 3, 4\}$. Hence we have

$$\mathbb{P}(X_{n+1} = k \mid X_n = 1) = \mathbb{P}(T = k), \qquad k = 1, 2, 3, 4,$$

and the process $(X_n)_{n\in\mathbb{N}}$ is a Markov chain on $\mathbb{S} = \{1, 2, 3, 4\}$, with transition matrix

$$P = \begin{bmatrix} \mathbb{P}(Y=1) & \mathbb{P}(Y=2) & \mathbb{P}(Y=3) & \mathbb{P}(Y=4) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

We now look for the limit

$$\lim_{n\to\infty} \mathbb{P}(X_n = 1).$$

Since the chain is irreducible, aperiodic (all states are checked to have period one) and its state space is finite, we know by Theorem 7.8 that

$$\pi_1 = \lim_{n\to\infty} \mathbb{P}(X_n = 1),$$

where $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ is the stationary distribution π uniquely determined from the equation $\pi = \pi P$, as follows:

$$\begin{cases} \pi_1 = 0.1\pi_1 + \pi_2 \\ \\ \pi_2 = 0.2\pi_1 + \pi_3 \\ \\ \pi_3 = 0.3\pi_1 + \pi_4 \\ \\ \pi_4 = 0.4\pi_1. \end{cases}$$

hence

$$\pi_2 = 0.9\pi_1, \qquad \pi_3 = 0.7\pi_1, \qquad \pi_4 = 0.4\pi_1,$$

under the condition

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1,$$

i.e.

$$\pi_1 + 0.9\pi_1 + 0.7\pi_1 + 0.4\pi_1 = 1,$$

which yields

$$\pi_1 = \frac{1}{3}, \quad \pi_2 = \frac{9}{30}, \quad \pi_3 = \frac{7}{30}, \quad \pi_4 = \frac{4}{30}.$$

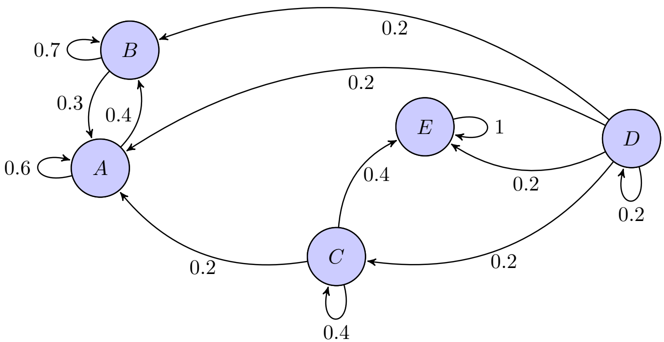
This result can be confirmed by computing the limit of matrix powers $(P^n)_{n\in\mathbb{N}}$ as n tends to infinity using the following Matlab/Octave commands:

```
P = [0.1,0.2,0.3,0.4;
1,0,0,0;
0,1,0,0;
0,0,1,0;]
mpower(P,1000)
```

showing that

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0.33333 & 0.30000 & 0.23333 & 0.13333 \\ 0.33333 & 0.30000 & 0.23333 & 0.13333 \\ 0.33333 & 0.30000 & 0.23333 & 0.13333 \\ 0.33333 & 0.30000 & 0.23333 & 0.13333 \end{bmatrix}.$$

Exercise 7.12 The graph of the chain is as follows:



We note that the chain is reducible, and that its state space $\mathbb{S}$ can be partitioned into 4 communicating classes:

$$\mathbb{S} = \{A, B\} \cup \{C\} \cup \{D\} \cup \{E\},$$

where A, B are recurrent, E is absorbing, and C, D are transient.

Starting from state $\textcircled{C}$, one can only return to $\textcircled{C}$ or end up in one of the absorbing classes $\{A, B\}$ or $\{E\}$. Let us denote by

$$T_{\{A,B\}} = \inf\{n \geqslant 0 \ : \ X_n \in \{A, B\}\}$$

the hitting time of $\{A, B\}$. We start by computing $\mathbb{P}(T_{\{A,B\}} < \infty \mid X_0 = C)$. By first step analysis, we find that this probability satisfies

$$\mathbb{P}(T_{\{A,B\}} < \infty \mid X_0 = C) = 0.2 + 0.4 \times \mathbb{P}(T_{\{A,B\}} < \infty \mid X_0 = C) + 0.4 \times 0,$$

hence

$$\mathbb{P}(T_{\{A,B\}} < \infty \mid X_0 = C) = \frac{1}{3},$$

which can also be recovered using a geometric series as

$$\mathbb{P}(T_{\{A,B\}} < \infty \mid X_0 = C) = 0.2 \sum_{n \geqslant 0} (0.4)^n = \frac{0.2}{1 - 0.4} = \frac{1}{3}.$$

On the other hand, $\{A, B\}$ is a closed two-state chain with transition matrix

$$\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix},$$

hence, starting from any state within $\{A, B\}$, the long run probability of being in A is given by

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 \in \{A,B\}) \\
&\quad = \lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 = A) \frac{\mathbb{P}(X_0 = A)}{\mathbb{P}(X_0 \in \{A,B\})} \\
&\qquad + \lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 = B) \frac{\mathbb{P}(X_0 = B)}{\mathbb{P}(X_0 \in \{A,B\})} \\
&\quad = \frac{b}{a+b} \left(\frac{\mathbb{P}(X_0 = A)}{\mathbb{P}(X_0 \in \{A,B\})} + \frac{\mathbb{P}(X_0 = B)}{\mathbb{P}(X_0 \in \{A,B\})} \right) \\
&\quad = \frac{0.3}{0.3 + 0.4} = \frac{3}{7}.
\end{aligned}$$

Since

$$\{X_n = A\} \subset \{T_{\{A,B\}} \leqslant n\} \subset \{T_{\{A,B\}} < \infty\}, \qquad n \geqslant 0,$$

we conclude that

$$\begin{aligned}
\alpha &:= \lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 = C) \\
&= \lim_{n\to\infty} \mathbb{P}\big(T_{\{A,B\}} < \infty \text{ and } X_n = A \mid X_0 = C\big) \\
&= \lim_{n\to\infty} \mathbb{P}\big(T_{\{A,B\}} < \infty \text{ and } X_{n+T_{\{A,B\}}} = A \mid X_0 = C\big) \\
&= \mathbb{P}\big(T_{\{A,B\}} < \infty \mid X_0 = C\big) \lim_{n\to\infty} \mathbb{P}\big(X_{n+T_{\{A,B\}}} = A \mid T_{\{A,B\}} < \infty \text{ and } X_0 = C\big) \\
&= \mathbb{P}\big(T_{\{A,B\}} < \infty \mid X_0 = C\big) \\
&\qquad \times \lim_{n\to\infty} \mathbb{P}\big(X_{n+T_{\{A,B\}}} = A \mid T_{\{A,B\}} < \infty,\ X_{T_{\{A,B\}}} \in \{A,B\},\ X_0 = C\big) \\
&= \mathbb{P}\big(T_{\{A,B\}} < \infty \mid X_0 = C\big) \lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 \in \{A,B\}) \\
&= \frac{1}{3} \times \frac{3}{7}
\end{aligned}$$

$$= \frac{1}{7},$$

where we used the strong Markov property, cf. Exercise 5.9.

The value α of this probability could also be obtained in a faster way by the following more intuitive first step analysis:

$$\alpha = 0.4 \times \alpha + 0.2 \times \frac{0.3}{0.7},$$

which recovers $\alpha = 1/7$, although this way of reasoning is a bit too fast to be truly recommended.

On the other hand, the numerical computation of the $20th$ power of P shows that

$$P^{20} = \begin{bmatrix} 0.4285 & 0.5714 & 0 & 0 & 0 \\ 0.4285 & 0.5714 & 0 & 0 & 0 \\ 0.1428 & 0.1904 & 1.099510^{-8} & 0 & 0.6666 \\ 0.2500 & 0.3333 & 1.099510^{-8} & 1.048510^{-14} & 0.4166 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

which recovers

$$\lim_{n\to\infty} \mathbb{P}(X_n = A \mid X_0 = C) = \frac{1}{7} \simeq 0.142857$$

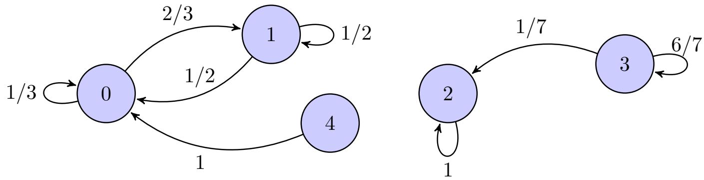
up to 6 decimals, and one can reasonably conjecture that the limiting distribution of the chain is given by

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 3/7 & 4/7 & 0 & 0 & 0 \\ 3/7 & 4/7 & 0 & 0 & 0 \\ 1/7 & 4/21 & 0 & 0 & 2/3 \\ 1/4 & 1/3 & 0 & 0 & 5/12 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \text{(S.51)}$$

which could also be recovered term by term using the above method. From this matrix one also sees clearly that C and D are transient states since they correspond to vanishing columns in the above matrix. From (S.51) we also note that $\lim_{n\to\infty} \mathbb{P}(X_n = i \mid X_0 = j)$ *is* dependent on the initial state ⓙ, due to the fact that the chain is not irreducible and Theorems 7.2 and 7.8 cannot be applied except if the initial state belongs to the communicating class $\{A, B\}$, for which $(3/7, 4/7)$ is a limiting distribution independent of the initial state. We also check that every limiting distribution in (S.51) is a stationary distribution solution of $\pi = \pi P$, cf. Proposition 7.7.

Exercise 7.13

a) The chain has the following graph:



b) The communicating classes are $\{0,1\}$, $\{2\}$, $\{3\}$, and $\{4\}$.
c) States ③ and ④ are transient, states ⓪ and ① are recurrent, and state ② is absorbing (hence it is recurrent).
d) By (4.5.7) we have

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 4) = \lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 0) = \frac{1/2}{2/3+1/2} = \frac{3}{7},$$

cf. also the Table 8.1.

Exercise 7.14

a) The transition matrix P of the chain on the state space $\mathbb{S} = (C, T)$ is given by

$$\begin{bmatrix} 4/5 & 1/5 \\ 3/4 & 1/4 \end{bmatrix}.$$

b) The stationary distribution $\pi = (\pi_C, \pi_T)$ is solution of $\pi = \pi P$, *i.e.*

$$\begin{cases} \pi_C = \dfrac{4}{5}\pi_C + \dfrac{3}{4}\pi_T \\ \\ \pi_T = \dfrac{1}{5}\pi_C + \dfrac{1}{4}\pi_T \end{cases}$$

under the condition $\pi_C + \pi_T = 1$, which yields $\pi_C = 15/19$ and $\pi_T = 4/19$.
c) In the long run, applying the Ergodic Theorem 7.12 we find that 4 out of 19 vehicles are trucks.
d) Let $\mu_T(C)$ and $\mu_T(T)$ denote the mean return times to state Ⓣ starting from Ⓒ and Ⓣ, respectively. By first step analysis, we have

$$\begin{cases} \mu_T(C) = 1 + \dfrac{4}{5}\mu_T(C) \\ \\ \mu_T(T) = 1 + \dfrac{3}{4}\mu_T(C) \end{cases}$$

which has for solution $\mu_T(C) = 5$ and $\mu_T(T) = 19/4$. Consequently, it takes on average $19/4 = 4.75$ vehicles after a truck until the next truck is seen under the bridge, and we check that the relation $\pi_D = 1/\mu_T(T)$ holds.

Exercise 7.15

a) We note that, due to the symmetry of P, the vector

$$\pi = [\pi_1, \pi_2, \ldots, \pi_N] := [1/N, \ldots, 1/N]$$

is an invariant probability distribution for the chain, since we can check the balance condition

$$\begin{aligned}[\pi P]_j &= \sum_{i=1}^{N} \pi_i P_{i,j} \\ &= \frac{1}{N} \sum_{i=1}^{N} P_{j,i} \\ &= \frac{1}{N} = \pi_j, \qquad j = 1, 2, \ldots, N.\end{aligned}$$

b) When $N = 2$ the chain can only have the transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which clearly has period 2. When $N \geqslant 3$ the chain is aperiodic because it is irreducible and one can travel from ⓪ to ⓪ in 2 steps via the path ⓪ → ① → ⓪, and in 3 steps via the path ⓪ → ① → ② → ⓪.

Exercise 7.16 (Problem 5.36-(d) continued).

a) The chain is reducible and its communicating classes are $\{0, 1, 2, 3, 4\}$ and $\{5\}$.
b) The limiting distribution is $(0, 0, 0, 0, 0, 1)$ independently of the initial state because the states $\{0, 1, 2, 3, 4\}$ are transient (cf. Proposition 7.4) and state ⑤ is absorbing. This means that

$$\lim_{n \to \infty} P^n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

which would be difficult to recover by a direct computation of P^n.

For the stationary distribution, the equation $\pi = \pi P$ reads

$$\begin{cases} \pi_0 = q\pi_0 + q\pi_1 + q\pi_2 \\ \pi_1 = p\pi_0 + p\pi_4 \\ \pi_2 = p\pi_1 \\ \pi_3 = p\pi_2 + p\pi_3 \\ \pi_4 = p\pi_3 \\ \pi_5 = q\pi_4 + \pi_5, \end{cases}$$

i.e.

$$\begin{cases} p\pi_0 = q\pi_1 + q\pi_2 \\ \pi_1 = p\pi_0 + p\pi_4 \\ \pi_2 = p\pi_1 \\ q\pi_3 = p\pi_2 \\ \pi_4 = p\pi_3 \\ \pi_4 = 0, \end{cases}$$

hence $(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5) = (0, 0, 0, 0, 0, 1)$, which coincides with the limiting distribution.

Note that the relation $\pi_i = 1/\mu_i(i)$ still holds for $i = 0, 1, 2, 3, 4, 5$, although not all of the assumptions of Theorems 7.2, 7.8 and 7.8 (notably the irreducibility condition) are satisfied here.

Exercise 7.17

a) We solve the system of equations

$$\begin{cases} \pi_0 = q(\pi_0 + \pi_1 + \pi_2 + \pi_3) + \pi_4 = q + p\pi_4 \\ \\ \pi_1 = p\pi_0 \\ \\ \pi_2 = p\pi_1 = p^2\pi_0 \\ \\ \pi_3 = p\pi_2 = p^3\pi_0 \\ \\ \pi_4 = p\pi_3 = p^4\pi_0, \end{cases}$$

which yields

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 = \pi_0(1 + p + p^2 + p^3 + p^4),$$

and

$$\begin{cases} \pi_0 = \dfrac{1}{1+p+p^2+p^3+p^4} \\ \pi_1 = \dfrac{p}{1+p+p^2+p^3+p^4} \\ \pi_2 = \dfrac{p^2}{1+p+p^2+p^3+p^4} \\ \pi_3 = \dfrac{p^3}{1+p+p^2+p^3+p^4} \\ \pi_4 = \dfrac{p^4}{1+p+p^2+p^3+p^4}. \end{cases}$$

b) Since the chain is irreducible and aperiodic with finite state space, its limiting distribution coincides with its stationary distribution.

Exercise 7.18

a) The transition matrix P is given by

$$P = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

b) The chain is aperiodic, irreducible, and has finite state space hence we can apply Theorem 7.8 or Theorem 7.10. The equation $\pi P = \pi$ reads

$$\pi P = [\pi_A, \pi_B, \pi_C, \pi_D] \times \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}\pi_B + \frac{1}{2}\pi_D \\ \frac{1}{2}\pi_A + \pi_C + \frac{1}{2}\pi_D \\ \frac{1}{3}\pi_B \\ \frac{1}{2}\pi_A + \frac{1}{3}\pi_B \end{bmatrix}$$

$$= [\pi_A, \pi_B, \pi_C, \pi_D],$$

i.e. $\pi_A = \pi_D = 2\pi_C$ and $\pi_B = 3\pi_C$, which, under the condition $\pi_A + \pi_B + \pi_C + \pi_D = 1$, gives $\pi_A = 1/4$, $\pi_B = 3/8$, $\pi_C = 1/8$, $\pi_D = 1/4$.

c) We solve the system

$$\begin{cases} \mu_D(A) = \dfrac{1}{2} + \dfrac{1}{2}(1 + \mu_D(B)) = 1 + \dfrac{1}{2}\mu_D(B) \\ \mu_D(B) = \dfrac{1}{3} + \dfrac{1}{3}(1 + \mu_D(A)) + \dfrac{1}{3}(1 + \mu_D(C)) = 1 + \dfrac{1}{3}(\mu_D(A) + \mu_D(C)) \\ \mu_D(C) = 1 + \mu_D(B) \\ \mu_D(D) = \dfrac{1}{2}(1 + \mu_D(A)) + \dfrac{1}{2}(1 + \mu_D(B)) = 1 + \dfrac{1}{2}(\mu_D(A) + \mu_D(B)), \end{cases}$$

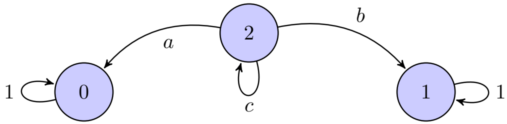
which has for solution

$$\mu_D(A) = \frac{8}{3}, \quad \mu_D(B) = \frac{10}{3}, \quad \mu_D(C) = \frac{13}{3}, \quad \mu_D(D) = 4.$$

On average, player D has to wait $\mu_D(D) = 4$ time units before recovering the token.

d) This probability is $\pi_D = 0.25$, and we check that the relation $\mu(D) = 1/\pi_D = 4$ is satisfied.

Exercise 7.19 Clearly we may assume that $c < 1$, as the case $c = 1$ corresponds to the identity matrix, or to constant a chain. On the other hand, we cannot directly apply Theorem 7.8 since the chain is reducible. The chain has the following graph:



a) By observation of

$$P^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a(1+c) & b(1+c) & c^2 \end{bmatrix}$$

and

$$P^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a(1+c+c^2) & b(1+c+c^2) & c^3 \end{bmatrix}$$

we infer that P^n takes the general form

$$P^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_n & b_n & c_n \end{bmatrix},$$

where a_n, b_n and c_n are coefficients to be determined by the following induction argument. Writing down the relation $P^{n+1} = P \times P^n$ as

$$P^{n+1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{n+1} & b_{n+1} & c_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_n & b_n & c_n \end{bmatrix}$$

shows that we have the recurrence relations

$$\begin{cases} a_{n+1} = a + c a_n, \\ b_{n+1} = b + c b_n, \\ c_{n+1} = c \times c_n, \end{cases} \quad \text{which yield} \quad \begin{cases} a_n = a + ac + \cdots + ac^{n-1} = a\dfrac{1-c^n}{1-c}, \\ b_n = b + bc + \cdots + bc^{n-1} = b\dfrac{1-c^n}{1-c}, \\ c_{n+1} = c^n, \end{cases}$$

hence

$$P^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a\dfrac{1-c^n}{1-c} & b\dfrac{1-c^n}{1-c} & c^n \end{bmatrix}.$$

b) From the structure of P^n it follows that the chain admits a limiting distribution

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{a}{1-c} & \dfrac{b}{1-c} & 0 \end{bmatrix}.$$

which is dependent of the initial state, provided that $c < 1$. The limiting probabilities

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_0 = 2) = \frac{a}{1-c},$$

resp.

$$\lim_{n\to\infty} \mathbb{P}(X_n = 1 \mid X_0 = 2) = \frac{b}{1-c},$$

correspond to the probability of moving to state ⓪, resp. the probability of moving to state ①, given one does not return to state ②.

In addition we have $\mathbb{P}(T_2^r = \infty \mid X_0 = 2) = a + b > 0$, hence state ② is transient and the chain is not recurrent.

c) By solving the equation $\pi = \pi P$ we find that the chain admits an infinity of stationary distributions of the form $(\pi_0, \pi_1, 0)$ with $\pi_0 + \pi_1 = 1$ when $c < 1$. We also note that here, all limiting distributions obtained in Question (b) are also stationary distributions on every row. Here again, Theorem 7.8 does not apply for the same reasons as in Question (b) and the limiting and stationary distributions may differ.

Exercise 7.20

a) The process $(X_n)_{n\geqslant 0}$ is a two-state Markov chain on $\{0, 1\}$ with transition matrix

$$\begin{bmatrix} \alpha & \beta \\ p & q \end{bmatrix}$$

and $\alpha = 1 - \beta$. The entries on the second line are easily obtained. Concerning the first line we note that $\mathbb{P}(N = 1) = \beta$ is the probability of switching from 0 to 1 in one time step, while the equality $\mathbb{P}(N \geqslant 2) = 1 - \beta$ shows that the probability of remaining at 0 for one time step is $1 - \beta$.

b) This probability is given from the stationary distribution (π_0, π_1) as $\pi_1 = \dfrac{\beta}{p+\beta}$.

Exercise 7.21

a) Since the chain is irreducible and aperiodic with finite state space it admits a unique stationary distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$ with

$$\pi_i = \frac{1}{\mu_i(i)}, \qquad i = 1, 2, \ldots, N,$$

by Corollary 7.9. Since $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$ is a probability distribution we have $\pi_1 + \pi_2 + \cdots + \pi_N = 1$ hence there exists $i \in \{1, 2, \ldots, N\}$ with $\pi_i = 1/\mu_i(i) \geqslant 1/N$, *i.e.* $\mu_i(i) \leqslant N$, otherwise the sum $\pi_1 + \pi_2 + \cdots + \pi_N$ would be lower than one.

b) Similarly, the condition $\pi_1 + \pi_2 + \cdots + \pi_N = 1$ shows that there must exist $i \in \{1, 2, \ldots, N\}$ with $\pi_i = 1/\mu_i(i) \leqslant 1/N$, *i.e.* $\mu_i(i) \geqslant N$, otherwise the sum $\pi_1 + \pi_2 + \cdots + \pi_N$ would be larger than one.

Exercise 7.22

a) The $N \times N$ transition matrix of the chain is

$$P = \begin{bmatrix} q & p & 0 & \cdots & \cdots & 0 & 0 & 0 \\ q & 0 & p & \cdots & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & \ddots & \ddots & 0 & 0 & 0 \\ 0 & 0 & q & \ddots & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & 0 & p & 0 \\ 0 & 0 & 0 & \cdots & \cdots & q & 0 & p \\ 0 & 0 & 0 & \cdots & \cdots & 0 & q & p \end{bmatrix}.$$

b) The chain is irreducible if $p \in (0, 1)$, and reducible if $p = 0$ or $p = 1$.

c) If $p \in (0, 1)$ there are no absorbing states and all states are positive recurrent. If $p = 0$, state ① is absorbing and all other states are transient. If $p = 1$, state N is absorbing and all other states are transient.

d) The equation $\pi = \pi P$ yields

$$\pi_2 = \frac{p}{q}\pi_1 \quad \text{and} \quad \pi_N = \frac{p}{q}\pi_{N-1}, \qquad k = 1, 2, \ldots, N-1,$$

and

$$p(\pi_k - \pi_{k-1}) = q(\pi_{k+1} - \pi_k) \qquad k = 2, 3, \ldots, N-1.$$

We check that π_k given by

$$\pi_k = \frac{p^{k-1}}{q^{k-1}} \pi_1, \qquad k = 1, 2, \ldots, N,$$

satisfies the above conditions. The normalization condition

$$1 = \sum_{k=1}^{N} \pi_k = \pi_1 \sum_{k=1}^{N} \frac{p^{k-1}}{q^{k-1}} = \pi_1 \sum_{k=0}^{N-1} \left(\frac{p}{q}\right)^k = \pi_1 \frac{1-(p/q)^N}{1-p/q}$$

shows that

$$\pi_k = \frac{1-p/q}{1-(p/q)^N} \frac{p^{k-1}}{q^{k-1}}, \qquad k = 1, 2, \ldots, N,$$

provided that $0 < p \neq q < 1$. When $p = q = 1/2$ we find that the uniform distribution

$$\pi_k = \frac{1}{N}, \qquad k = 1, 2, \ldots, N,$$

is stationary. When $p = 0$ the stationary distribution is $\mathbb{1}_{\{0\}} = [1, 0, \ldots, 0, 0]$, and when $p = 1$ it is $\mathbb{1}_{\{N\}} = [0, 0, \ldots, 0, 1]$.

e) The chain has finite state space and when $p \in (0, 1)$ it is irreducible and aperiodic, hence its limiting distribution coincides with its stationary distribution. It can by easily checked that this coincidence also occurs here for $p = 0$ and $p = 1$, although in those cases the chain is not irreducible and not aperiodic.

Exercise 7.23 (Problem 6.9 continued). We note that the chain is irreducible and has period 2.

a) By Problem 6.9 the chain is positive recurrent if and only $\alpha < 1$, hence by Theorem 7.10 it admits a stationary distribution π in this case. The relation $\pi = \pi P$ reads

$$\pi = [\pi_0, \pi_1, \pi_2, \ldots] = [\pi_0, \pi_1, \pi_2, \ldots] \times \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{1}{\alpha+1} & 0 & \frac{\alpha}{\alpha+1} & 0 & \cdots \\ 0 & \frac{1}{\alpha+1} & 0 & \frac{\alpha}{\alpha+1} & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix},$$

hence

$$\pi_0 = \frac{\pi_1}{\alpha+1}, \qquad \pi_1 = \pi_0 + \frac{\pi_2}{\alpha+1}, \qquad \pi_2 = \frac{\alpha\pi_1}{\alpha+1} + \frac{\pi_3}{\alpha+1}, \ldots$$

i.e.

$$\pi_0 = \frac{\pi_2}{\alpha(\alpha+1)}, \qquad \pi_1 = \frac{\pi_2}{\alpha}, \qquad \pi_2 = \frac{\pi_3}{\alpha}, \ldots$$

from which we can deduce that

$$\pi_k = \pi_0 \alpha^{k-1}(1+\alpha), \qquad k \geqslant 1.$$

From the condition $\sum_{k\geqslant 0} \pi_k = 1$ we find

$$1 = \pi_0 + \pi_0(1+\alpha)\sum_{k\geqslant 1}\alpha^{k-1} = \pi_0 + \pi_0\frac{1+\alpha}{1-\alpha},$$

hence

$$\pi_0 = \frac{1-\alpha}{2} \text{ and } \pi_k = \alpha^{k-1}\frac{1-\alpha^2}{2}, \qquad k \geqslant 1.$$

More generally, from the relation $\pi = \pi P$ we can also write

$$\pi_k = \frac{\alpha\pi_{k-1}}{\alpha+1} + \frac{\pi_{k+1}}{\alpha+1}, \qquad k \geqslant 2,$$

which can be written as

$$\pi_k = q\pi_{k-1} + p\pi_{k+1}, \qquad k \geqslant 2, \tag{S.52}$$

where $p = 1/(\alpha+1)$ and $q = \alpha/(\alpha+1)$. We look for a solution of (S.52) of the form $\pi_k = C(q/p)^k = C\alpha^k$ as in (2.2.15), $k \geqslant 1$, which clearly satisfies (S.52) written as

$$\frac{\alpha\alpha^{k-1}}{\alpha+1} + \frac{\alpha^{k+1}}{\alpha+1} = \alpha^k, \qquad k \geqslant 2.$$

When $\alpha < 1$ the value of C can be found from the condition

$$\begin{aligned} 1 &= \sum_{k\geqslant 0}\pi_k \\ &= \pi_0 + \sum_{k\geqslant 2}\pi_k \\ &= \frac{\pi_2}{\alpha(\alpha+1)} + \sum_{k\geqslant 1}\pi_k \\ &= C\frac{\alpha}{\alpha+1} + C\frac{\alpha}{1-\alpha} \\ &= C\frac{2\alpha}{1-\alpha^2} \end{aligned}$$

hence

$$C = \frac{1-\alpha^2}{2\alpha}$$

and

$$\pi_0 = \frac{1-\alpha}{2}, \quad \pi_k = \alpha^{k-1}\frac{1-\alpha^2}{2}, \qquad k \geqslant 1.$$

b) When $\alpha > 1$ the series $\sum_{k\geqslant 1} \pi_k$ does not converge so there cannot be a stationary distribution. In addition, π_0 would be negative if $\alpha > 1$. When $\alpha = 1$ we find $\pi_k = 0$, $k \in \mathbb{N}$, which cannot be a probability distribution as well (note that from Question (a) the solution of $\pi = \pi P$ is unique).

c) From the solution of Problem 6.9, we check that

$$\mathbb{E}[T_0 \mid X_0 = 0] = \frac{1}{1-\alpha} = \frac{1}{\pi_0},$$

and

$$\mathbb{E}[T_k \mid X_0 = k] = \frac{2\alpha^{k-1}}{1-\alpha^2} = \frac{1}{\pi_k} < \infty, \qquad k \geqslant 1,$$

as expected from Theorem 7.10.

Exercise 7.24

a) We have

$$\begin{aligned}\mathbb{P}(X_n = x) &= \mathbb{P}(\{X_n = x\} \cap \{\tau \leqslant n\}) + \mathbb{P}(\{X_n = x\} \cap \{\tau > n\}) \\ &= \mathbb{P}(\{Y_n = x\} \cap \{\tau \leqslant n\}) + \mathbb{P}(\{X_n = x\} \cap \{\tau > n\}) \\ &\leqslant \mathbb{P}(Y_n = x) + \mathbb{P}(\tau > n).\end{aligned}$$

b) Similarly to Question (a), we have

$$\begin{aligned}\mathbb{P}(Y_n = x) &= \mathbb{P}(\{Y_n = x\} \cap \{\tau \leqslant n\}) + \mathbb{P}(\{Y_n = x\} \cap \{\tau > n\}) \\ &= \mathbb{P}(\{X_n = x\} \cap \{\tau \leqslant n\}) + \mathbb{P}(\{Y_n = x\} \cap \{\tau > n\}) \\ &\leqslant \mathbb{P}(X_n = x) + \mathbb{P}(\tau > n),\end{aligned}$$

hence

$$-\mathbb{P}(\tau > n) \leqslant \mathbb{P}(X_n = x) - \mathbb{P}(Y_n = x) \leqslant \mathbb{P}(\tau > n),$$

which leads to the conclusion.

Exercise 7.25

a) We have

$$\begin{aligned}&\mathbb{P}(X_n \in A) = \mathbb{P}(X_n \in A \text{ and } \tau \leqslant n) + \mathbb{P}(X_n \in A \text{ and } \tau > n) \\ &= \mathbb{P}(X_n \in A \mid \tau \leqslant n)\mathbb{P}(\tau \leqslant n) + \mathbb{P}(X_n \in A \mid \tau > n)\mathbb{P}(\tau > n) \\ &= \pi(A)\mathbb{P}(\tau \leqslant n) + \mathbb{P}(X_n \in A \mid \tau > n)\mathbb{P}(\tau > n) \\ &= \pi(A) + (\mathbb{P}(X_n \in A \mid \tau > n) - \pi(A))\mathbb{P}(\tau > n).\end{aligned}$$

b) We have

$$\begin{aligned}| \mathbb{P}(X_n \in A) - \pi(A) | &= | (\mathbb{P}(X_n \in A \mid \tau > n) - \pi(A)) | \mathbb{P}(\tau > n) \\ &\leqslant \mathbb{P}(\tau > n),\end{aligned}$$

since for any $a, b \in [0,1]$ we have $|a-b| \leqslant 1$ due to the inequalities

$$-1 \leqslant a - 1 \leqslant a - b \leqslant 1 - b \leqslant 1.$$

c) Such an example can be constructed as the hitting time τ of a domain inside $\mathbb{S}$, by freezing $X_n = X_{\min(\tau,n)}$ at time τ.

Exercise 7.26

a) Since M has positive entries and is column-stochastic, $P := M^\top$ is the transition probability matrix of an aperiodic irreducible Markov chain with finite state space $\mathbb{S} = \{1, 2, \ldots, n\}$. By Corollary 7.9, the chain admits a unique stationary distribution π such that $\pi = \pi P$, *i.e.* $\pi^\top = (\pi P)^\top = P^\top \pi^\top = M\pi^\top$, *i.e.* π^T is the only eigenvector of M with eigenvalue 1 under the normalization condition $\|\pi\|_1 = 1$.

b) The first statement follows as in Question (a) above from Corollary 7.9, by letting $\pi = q^\top$. The second statement also follows from Corollary 7.9, which states that

$$q = \pi^\top = \lim_{k\to\infty} (e_j P^k)^\top = \lim_{k\to\infty} (P^\top)^k e_j^\top = \lim_{k\to\infty} M^k e_j^\top = \lim_{k\to\infty} [M^k]_{\cdot,j}$$

for any $e_j = \mathbf{1}_{\{j\}}$, $j \in \mathbb{S}$. Therefore, decomposing x_0 as $x_0 = \sum_{j\in\mathbb{S}} x_0^j e_j^\top$, we have

$$q = q \sum_{j\in\mathbb{S}} x_0^j = \sum_{j\in\mathbb{S}} x_0^j \lim_{k\to\infty} M^k e_j^\top = \lim_{k\to\infty} M^k \sum_{j\in\mathbb{S}} x_0^j e_j^\top = \lim_{k\to\infty} M^k x_0.$$

Exercise 7.27

a) We have

$$\begin{aligned}\lim_{n\to\infty} \frac{\mathbb{E}[R_n^i]}{n} &= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\sum_{j=1}^n \mathbb{1}_{\{X_j = i\}}\right] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[\mathbb{1}_{\{X_j = i\}}] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n \mathbb{P}(X_j = i)\end{aligned}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \sum_{k\in\mathbb{S}} \mathbb{P}(X_j = i \mid X_0 = k)\mathbb{P}(X_0 = k)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \sum_{k\in\mathbb{S}} [P^j]_{k,i}\mathbb{P}(X_0 = k)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k\in\mathbb{S}} [P^{j+1}]_{k,i}\mathbb{P}(X_0 = k)$$
$$= \sum_{l\in\mathbb{S}} P_{i,l} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \sum_{k\in\mathbb{S}} [P^j]_{k,i}\mathbb{P}(X_0 = k)$$
$$= \sum_{l\in\mathbb{S}} P_{i,l} \lim_{n\to\infty} \frac{\mathbb{E}[R_n^l]}{n},$$

hence $\eta_i := \lim_{n\to\infty} \mathbb{E}[R_n^l]/n$, $i \in \mathbb{S}$, satisfies the equation $\eta = \eta P$ and we conclude by uniqueness of the stationary distribution $(\pi_i)_{i\in\mathbb{S}}$ as the solution to that equation.

b) Letting $\tau_x^{(0)} := 0$ and letting $\tau_x^{(k)}$ denote the time of the kth visit to state x, the sequence $\big(\tau_x^{(k+1)} - \tau_x^{(k)}\big)_{k\geqslant 0}$, resp. $\big(R^y_{\tau_x^{(k+1)}} - R^y_{\tau_x^{(k)}}\big)_{k\geqslant 0}$, is made of independent random variables, $i \in \mathbb{S}$, hence by the law of large numbers for renewal processes, see Corollary 14 page 106 of Serfozo (2009), we have

$$\pi_y = \lim_{n\to\infty} \frac{\mathbb{E}[R_n^y]}{n} = \frac{\mathbb{E}\big[R^y_{\tau_x^{(1)}} \mid X_0 = x\big]}{\mathbb{E}\big[\tau_x^{(1)} \mid X_0 = x\big]} = \frac{\mathbb{E}[N_{x,y} \mid X_0 = x]}{\mathbb{E}[\tau_x \mid X_0 = x]}, \quad x, y \in \mathbb{S}.$$

c) We have

$$\mathbb{P}(N_{x,y} = 0 \mid X_0 = x) = 1 - \mathbb{P}(N_{x,y} \geqslant 1 \mid X_0 = x) = 1 - \alpha_{x,y}$$

and

$$\begin{aligned}
&\mathbb{P}(N_{x,y} = k \mid X_0 = x) \\
&= \mathbb{P}(N_{x,y} \geqslant 1 \mid X_0 = x)\big(\mathbb{P}(N_{y,x} = 0 \mid X_0 = y)\big)^{k-1}\mathbb{P}(N_{y,x} \geqslant 1 \mid X_0 = y) \\
&= \alpha_{x,y}(1 - \alpha_{y,x})^{k-1}\alpha_{y,x}, \qquad k \geqslant 1,
\end{aligned}$$

and we check that

$$\begin{aligned}
\mathbb{P}(N_{x,y} \geqslant 0 \mid X_0 = x) &= \mathbb{P}(N_{x,y} = 0 \mid X_0 = x) + \mathbb{P}(N_{x,y} \geqslant 1 \mid X_0 = x) \\
&= 1 - \alpha_{x,y} + \sum_{k\geqslant 1} \mathbb{P}(N_{x,y} = k \mid X_0 = x) \\
&= 1 - \alpha_{x,y} + \alpha_{x,y}\alpha_{y,x} \sum_{k\geqslant 1} (1 - \alpha_{y,x})^{k-1}
\end{aligned}$$

$$= 1, \qquad x, y \in \mathbb{S}.$$

d) We have

$$\begin{aligned}
\frac{\pi_y}{\pi_x} &= \pi_y \mathbb{E}[\tau_x \mid X_0 = x] \\
&= \mathbb{E}[N_{x,y} \mid X_0 = x] \\
&= \sum_{k=1}^{\infty} k \mathbb{P}(N_{x,y} = k \mid X_0 = x) \\
&= \alpha_{x,y}\alpha_{y,x} \sum_{k=1}^{\infty} k(1-\alpha_{y,x})^{k-1} \\
&= \frac{\alpha_{x,y}\alpha_{y,x}}{\alpha_{y,x}^2} \\
&= \frac{\alpha_{x,y}}{\alpha_{y,x}}, \qquad x, y \in \mathbb{S}.
\end{aligned}$$

Problem 7.28

a) The computation of eigenvalues shows that the two eigenvalues are $\lambda = 1 - a - b$ and 1.
b) Solving the equation $\pi = \pi P$ for π shows that the stationary distribution is given by $(\pi_0, \pi_1) = (b/(a+b), a/(a+b))$.
c) The relation is clearly verified for $n = 0$. Next, assuming that it holds at the rank n, we have

$$\begin{aligned}
&\begin{bmatrix} \mathbb{E}\big[\exp\big(t\sum_{k=1}^{n+1} X_k\big) \mid X_0 = 0\big] \\ \mathbb{E}\big[\exp\big(t\sum_{k=1}^{n+1} X_k\big) \mid X_0 = 1\big] \end{bmatrix} \\
&= \begin{bmatrix} (1-a)\mathbb{E}\big[\mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 0\big] + ae^t\mathbb{E}\big[\mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 1\big] \\ b\mathbb{E}\big[\mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 0\big] + (1-b)e^t\mathbb{E}\big[\mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 1\big] \end{bmatrix} \\
&= \begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} \begin{bmatrix} \mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 0\big] \\ \mathbb{E}\big[\exp\big(t\sum_{k=2}^{n+1} X_k\big) \mid X_1 = 1\big] \end{bmatrix} \\
&= \begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} \begin{bmatrix} \mathbb{E}\big[\exp\big(t\sum_{k=1}^{n} X_k\big) \mid X_0 = 0\big] \\ \mathbb{E}\big[\exp\big(t\sum_{k=1}^{n} X_k\big) \mid X_0 = 1\big] \end{bmatrix} \\
&= \begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} \left(\begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix}\right)^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\end{aligned}$$

$$= \left(\begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} \right)^{n+1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad t \in \mathbb{R}.$$

d) By diagonalizing P as

$$\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\pi_0}} & 0 \\ 0 & \frac{1}{\sqrt{\pi_1}} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & -\sqrt{\pi_1} \\ \sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & \sqrt{\pi_1} \\ -\sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & 0 \\ 0 & \sqrt{\pi_1} \end{bmatrix},$$

we have

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] = [\pi_0, \pi_1] \begin{bmatrix} \mathbb{E}\big[\exp\left(t\sum_{k=1}^n X_k\right) \bigm| X_0 = 0\big] \\ \mathbb{E}\big[\exp\left(t\sum_{k=1}^n X_k\right) \bigm| X_0 = 1\big] \end{bmatrix}$$

$$= [\pi_0, \pi_1] \left(\begin{bmatrix} 1-a & ae^t \\ b & (1-b)e^t \end{bmatrix} \right)^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [\pi_0, \pi_1] \left(\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \right)^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [\pi_0, \pi_1] \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix}$$

$$\times \left(\begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \right)^{n-1} \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [\pi_0, \pi_1] \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix}$$

$$\times \left(\begin{bmatrix} \frac{1}{\sqrt{\pi_0}} & 0 \\ 0 & \frac{e^{t/2}}{\sqrt{\pi_1}} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & -\sqrt{\pi_1} \\ \sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & \sqrt{\pi_1} \\ -\sqrt{\pi_1} & \sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & 0 \\ 0 & e^{t/2}\sqrt{\pi_1} \end{bmatrix} \right)^{n-1}$$

$$\times \begin{bmatrix} 1 & 0 \\ 0 & e^{t/2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [\pi_0, \pi_1 e^{t/2}]$$

$$\times \left(\begin{bmatrix} \frac{1}{\sqrt{\pi_0}} & 0 \\ 0 & \frac{1}{\sqrt{\pi_1}} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & -\sqrt{\pi_1} \\ e^{t/2}\sqrt{\pi_1} & e^{t/2}\sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & e^{t/2}\sqrt{\pi_1} \\ -\sqrt{\pi_1} & e^{t/2}\sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & 0 \\ 0 & \sqrt{\pi_1} \end{bmatrix} \right)^{n-1}$$

$$
\begin{aligned}
&\quad \times \begin{bmatrix} 1 \\ e^{t/2} \end{bmatrix} \\
&= [\pi_0, \pi_1 e^{t/2}] \begin{bmatrix} \frac{1}{\sqrt{\pi_0}} & 0 \\ 0 & \frac{1}{\sqrt{\pi_1}} \end{bmatrix} \left(\begin{bmatrix} \sqrt{\pi_0} & -\sqrt{\pi_1} \\ e^{t/2}\sqrt{\pi_1} & e^{t/2}\sqrt{\pi_0} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \sqrt{\pi_0} & e^{t/2}\sqrt{\pi_1} \\ -\sqrt{\pi_1} & e^{t/2}\sqrt{\pi_0} \end{bmatrix} \right)^{n-1} \\
&\quad \times \begin{bmatrix} \sqrt{\pi_0} & 0 \\ 0 & \sqrt{\pi_1} \end{bmatrix} \begin{bmatrix} 1 \\ e^{t/2} \end{bmatrix} \\
&= [\sqrt{\pi_0}, \sqrt{\pi_1} e^{t/2}] \left(\begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & (\lambda + (1-\lambda)\pi_1)e^t \end{bmatrix} \right)^{n-1} \begin{bmatrix} \sqrt{\pi_0} \\ \sqrt{\pi_1} e^{t/2} \end{bmatrix},
\end{aligned}
$$

$t \in \mathbb{R}$.

e) Taking

$$
\begin{aligned}
M(t) &= \begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & (\lambda + (1-\lambda)\pi_1)e^t \end{bmatrix} \\
&= \begin{bmatrix} \pi_0 + \lambda\pi_1 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & e^t(\pi_1 + \lambda\pi_0) \end{bmatrix},
\end{aligned}
$$

We have

$$
\mu(t) = \frac{1}{2}\left(\mathrm{Tr}(M(t)) + \sqrt{(\mathrm{Tr}(M(t)))^2 - 4\lambda e^t}\right),
$$

where

$$
\mathrm{Tr}(M(t)) = \lambda + (1-\lambda)\pi_0 + (\lambda + (1-\lambda)\pi_1)e^t
$$

f) Since the matrix $M(t)$ is symmetric, by Proposition 9 in Foucart (2010) we have

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] \\
&\leqslant \|[\sqrt{\pi_0}, \sqrt{\pi_1}e^{t/2}]\|_2 \left\| \left(\begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} \\ (1-\lambda)e^{t/2}\sqrt{\pi_0\pi_1} & (\lambda + (1-\lambda)\pi_1)e^t \end{bmatrix} \right)^{n-1} \right\|_2 \left\| \begin{bmatrix} \sqrt{\pi_0} \\ \sqrt{\pi_1}e^{t/2} \end{bmatrix} \right. \\
&= (\mu(t))^{n-1} \|[\sqrt{\pi_0}, \sqrt{\pi_1}e^{t/2}]\|_2^2 \\
&= (\pi_0 + \pi_1 e^t)(\mu(t))^{n-1}.
\end{aligned}
$$

Next, applying again Proposition 9 in Foucart (2010) to $A := \sqrt{M(t)}$, we have

$$\mu(t) \geqslant \frac{1}{\|[\sqrt{\pi_0}, e^{t/2}\sqrt{\pi_1}]\|_2^2} \left\| \sqrt{M(t)}[\sqrt{\pi_0}, e^{t/2}\sqrt{\pi_1}]^\top \right\|_2^2$$
$$= \frac{1}{\pi_0 + \pi_1 e^t} \langle [\sqrt{\pi_0}, e^{t/2}\sqrt{\pi_1}], M(t)[\sqrt{\pi_0}, e^{t/2}\sqrt{\pi_1}]^\top \rangle$$
$$= \frac{1}{\pi_0 + \pi_1 e^t} \left\langle [\sqrt{\pi_0}, e^{t/2}\sqrt{\pi_1}], \begin{bmatrix} \pi_0\sqrt{\pi_0} + \lambda\pi_1\sqrt{\pi_0} + (1-\lambda)e^t\pi_1\sqrt{\pi_0} \\ (1-\lambda)e^{t/2}\pi_0\sqrt{\pi_1} + e^{3t/2}\pi_1\sqrt{\pi_1} + \lambda e^{3t/2}\pi_0\sqrt{\pi_1} \end{bmatrix} \right\rangle$$
$$= \frac{\pi_0^2 + 2e^t\pi_0\pi_1 + e^{2t}\pi_1^2 + \lambda(\pi_0\pi_1 - 2e^t\pi_0\pi_1 + e^{2t}\pi_0\pi_1)}{\pi_0 + \pi_1 e^t}$$
$$= \pi_0 + \pi_1 e^t + \lambda \frac{(\pi_0 - e^t\pi_1)^2}{\pi_0 + \pi_1 e^t}$$
$$\geqslant \pi_0 + \pi_1 e^t$$

since $\lambda \geqslant 0$, which shows that

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] \leqslant (\mu(t))^n, \qquad t \in \mathbb{R}_+.$$

g) By the classical Markov or Chernoff bound argument, we have

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n (X_k - \pi_1) \geqslant z\right) &= \mathbb{P}\left(\exp\left(t\sum_{k=1}^n X_k\right) \geqslant e^{ntz+nt\pi_1}\right) \\ &= e^{-ntz-nt\pi_1}\mathbb{E}\left[\exp\left(t\sum_{k=1}^n X_k\right)\right] \\ &= e^{-ntz-nt\pi_1}(\mu(t))^n \\ &= \mathrm{e}^{-n(t(\pi_1+z)-\log\mu(t))}, \qquad t > 0.\end{aligned}$$

h) This section only sketches the solution argument, see Appendices A and B in Léon and Perron (2004) for the full proof details. By differentiating

$$\begin{aligned} t &\mapsto xt - \log\mu(t) \\ &= xt - \log\left(\frac{1}{2}(\lambda + (1-\lambda)\pi_0 + (\lambda + (1-\lambda)\pi_1)e^t \right. \\ &\qquad \left. + \sqrt{(\lambda + (1-\lambda)\pi_0 + (\lambda + (1-\lambda)\pi_1)e^t)^2 - 4\lambda e^t}\right)\end{aligned}$$

with respect to $t > 0$, we find that the maximizing value $t(x)$ satisfies

$$x = \frac{\mu'(t)}{\mu(t)}$$

$$= \frac{\mathrm{Tr}(M'(t)) + (2\mathrm{Tr}(M'(t))\mathrm{Tr}(M(t)) - 4\lambda e^t)/2/\sqrt{(\mathrm{Tr}(M(t)))^2 - 4\lambda e^t}}{\mathrm{Tr}(M(t)) + \sqrt{(\mathrm{Tr}(M(t)))^2 - 4\lambda e^t}},$$

After multiplying the numerator and denominator by $\mathrm{Tr}(M(t)) - \sqrt{(\mathrm{Tr}(M(t)))^2 - 4\lambda e^t}$ and simplifying, we obtain

$$(2x-1)\sqrt{(\mathrm{Tr}(M(t)))^2 - 4\lambda e^t} = (\pi_1 + \lambda\pi_0)e^t - (\pi_0 + \lambda\pi_1).$$

This relation can be used to derive a quadratic equation for $\mathrm{e}^{t(x)}$, with solution

$$\mathrm{e}^{t(x)} = \frac{(\pi_0 + \lambda\pi_1)\big(2x - 1 + \sqrt{\Delta(x)}\big)}{(\pi_1 + \lambda\pi_0)\big(1 - 2x + \sqrt{\Delta(x)}\big)},$$

where

$$\Delta(x) := 1 + \frac{4\lambda(1-x)x}{\pi_0\pi_1(1-\lambda)^2},$$

which yields

$$\mu(t(x)) = \frac{(\pi_0 + \lambda\pi_1)(1 + \sqrt{\Delta})}{1 - 2x + \sqrt{\Delta}}.$$

Letting

$$g(x) := \frac{xt(x) - \log\mu(t(x))}{(x - \pi_1)^2}, \qquad x \in (0,1),$$

we check that $g'(\pi_0) = 0$ and $g(x)$ admits a global minimum at $x = \pi_0$. Then, we have

$$\Delta(\pi_0) := 1 + \frac{4\lambda}{(1-\lambda)^2} = \frac{(1+\lambda)^2}{(1-\lambda)^2},$$

$$t(\pi_0) = \log\frac{(\pi_0 + \lambda\pi_1)\big(\pi_0 - \pi_1 + \frac{1+\lambda}{1-\lambda}\big)}{(\pi_1 + \lambda\pi_0)\big(\pi_1 - \pi_0 + \frac{1+\lambda}{1-\lambda}\big)},$$

$$\mu(t(\pi_0)) = \frac{(\pi_0 + \lambda\pi_1)(1 + \frac{1+\lambda}{1-\lambda})}{\pi_1 - \pi_0 + \frac{1+\lambda}{1-\lambda}},$$

and letting $r := (b-a)/(2-a-b)$, we have

$$\begin{aligned} g(\pi_0) &= \frac{1}{\pi_0 - \pi_1}\log\frac{1-(1-\lambda)\pi_1}{1-(1-\lambda)\pi_0} \\ &= \frac{a+b}{b-a}\log\frac{1-a}{1-b} \\ &= \frac{1-\lambda}{1+\lambda}\frac{1}{r}\log\frac{1+r}{1-r} \\ &= \frac{1-\lambda}{1+\lambda}\frac{1}{r}(\log(1+r) - \log(1-r)) \end{aligned}$$

$$= \frac{1-\lambda}{1+\lambda}\frac{1}{r}\left(\sum_{n\geqslant 1}(-1)^{n+1}\frac{r^n}{n} + \sum_{n\geqslant 1}\frac{r^n}{n}\right)$$
$$= \frac{1-\lambda}{1+\lambda}\frac{1}{r}\sum_{n\geqslant 0}\frac{r^{2n+1}}{2n+1}$$
$$\geqslant 2\frac{1-\lambda}{1+\lambda},$$

hence for $z \in [0, 1-\pi_1]$ we have

$$\log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \pi_1) \geqslant z\right) \leqslant -nz^2 g(\pi_1 + z) \leqslant -nz^2 g(\pi_0) \leqslant -2nz^2\frac{1-\lambda}{1+\lambda},$$

while

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \pi_1) \geqslant z\right) = 0$$

for $z > 1 - \pi_1$.

Problem 7.29

a) We have

$$\lim_{n\to\infty}\frac{\mathbb{E}[R_n^i]}{n} = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n}\mathbb{1}_{\{X_j=i\}}\right]$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\mathbb{1}_{\{X_j=i\}}]$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}(X_j = i)$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\sum_{k\in\mathbb{S}}\mathbb{P}(X_j = i \mid X_0 = k)\mathbb{P}(X_0 = k)$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\sum_{k\in\mathbb{S}}[P^j]_{k,i}\mathbb{P}(X_0 = k)$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\sum_{k\in\mathbb{S}}[P^{j+1}]_{k,i}\mathbb{P}(X_0 = k)$$
$$= \sum_{l\in\mathbb{S}}P_{i,l}\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\sum_{k\in\mathbb{S}}[P^j]_{k,i}\mathbb{P}(X_0 = k)$$

$$= \sum_{l\in\mathbb{S}} P_{i,l} \lim_{n\to\infty} \frac{\mathbb{E}[R_n^l]}{n},$$

hence $\eta_i := \lim_{n\to\infty} \mathbb{E}[R_n^l]/n$, $i \in \mathbb{S}$, satisfies the equation $\eta = \eta P$ and we conclude by uniqueness of the stationary distribution $(\pi_i)_{i\in\mathbb{S}}$ as the solution to that equation.

b) Theorem 31 page 15 of Freedman (1983) shows that letting $\tau_0 := 0$, the sequence $(\tau_{k+1} - 1 - \tau_k)_{k\geqslant 0}$, resp. $(R^i_{\tau_{k+1}-1} - R^i_{\tau_k})_{k\geqslant 0}$, is made of independent random variables, $i \in \mathbb{S}$, hence by the law of large numbers for renewal processes, see Corollary 14 page 106 of Serfozo (2009), we have

$$\pi_i = \lim_{n\to\infty} \frac{\mathbb{E}[R_n^i]}{n} = \frac{\mathbb{E}[R^i_{\tau_1-1}]}{\mathbb{E}[\tau_1 - 1]}.$$

c) By the Wald identity, see e.g. Theorem 2 of Chewi (2017), we have

$$\mathbb{E}[T-1] = \mathbb{E}[\tau_1 - 1]\mathbb{E}[\kappa]$$

and

$$\mathbb{E}\left[\sum_{j=1}^{T-1} \mathbb{1}_{\{X_j=i\}}\right] = \mathbb{E}\left[\sum_{j=1}^{\tau_1-1} \mathbb{1}_{\{X_j=i\}}\right]\mathbb{E}[\kappa],$$

hence

$$\pi_i = \frac{\mathbb{E}\left[\sum_{j=1}^{\tau_1-1} \mathbb{1}_{\{X_j=i\}}\right]}{\mathbb{E}[\tau_1-1]} = \frac{\mathbb{E}\left[\sum_{j=1}^{T-1} \mathbb{1}_{\{X_j=i\}}\right]}{\mathbb{E}[T-1]}, \qquad i \in \mathbb{S}.$$

Problem 7.30

a) Bounded regret.

i) Define the sequence $(\tau_k)_{k\geqslant 1}$ recursively as

$$\tau_1 := \inf\{l > 1 \ : \ X_l = X_1\},$$

and

$$\tau_k := \inf\{l > \tau_{k-1} \ : \ X_l = X_1\}, \qquad k \geqslant 2,$$

and let

$$T := \inf\{l > \tau \ : \ X_l = X_1\}.$$

By Question (c) of Problem 7.29, we have

$$\pi_1^{(i)}\mathbb{E}[T-1] = \mathbb{E}\big[R^{(i)}_{T-1}\big], \qquad i \in \mathbb{S}.$$

Hence we have

$$R_{T-1}^{(i)} - (T-\tau) \leqslant R_T^{(i)} - (T-\tau) \leqslant R_\tau^{(i)} \leqslant R_{T-1}^{(i)}$$

and

$$\pi_1^{(i)} \mathbb{E}[T-1] - \mathbb{E}[T-\tau] \leqslant \mathbb{E}\big[R_\tau^{(i)}\big] \leqslant \mathbb{E}\big[R_{T-1}^{(i)}\big] = \pi_1^{(i)} \mathbb{E}[T-1]$$

or

$$\pi_1^{(i)} \mathbb{E}[T-1] - \mathbb{E}[T-\tau] \leqslant \mathbb{E}\big[R_\tau^{(i)}\big] \leqslant \pi_1^{(i)} \mathbb{E}[\tau] + \mathbb{E}[T-\tau]$$

hence

$$\pi_1^{(i)} \mathbb{E}[\tau] - \mathbb{E}[T-\tau] \leqslant \mathbb{E}\big[R_\tau^{(i)}\big] \leqslant \pi_1^{(i)} \mathbb{E}[\tau] + \mathbb{E}[T-\tau],$$

and therefore

$$\big|\mathbb{E}\big[R_\tau^{(i)}\big] - \pi_1^{(i)} \mathbb{E}[\tau]\big| \leqslant \mathbb{E}[T-\tau]. \tag{S.53}$$

ii) We have

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{i=1}^N \sum_{k=1}^{T_n^{(i,\alpha)}} X_k^{(i)} - \sum_{i=1}^N \pi_1^{(i)} T_n^{(i,\alpha)}\right]\right| \leqslant \sum_{i=1}^N \left|\mathbb{E}\left[\sum_{k=1}^{T_n^{(i,\alpha)}} X_k^{(i)} - \pi_1^{(i)} T_n^{(i,\alpha)}\right]\right| \\
&\leqslant \sum_{i=1}^N \left|\mathbb{E}\big[R_{T_n^{(i,\alpha)}}^{(i)} - \pi_1^{(i)} T_n^{(i,\alpha)}\big]\right| \\
&\leqslant \sum_{i=1}^N \mathbb{E}\big[\tau_\kappa^{(i)} - T_n^{(i,\alpha)}\big] \\
&= \sum_{i=1}^N \sum_{l,j\in\{0,1\}} \mathbb{E}\big[\tau_\kappa^{(i)} - T_n^{(i,\alpha)} \mid X_{\tau_\kappa^{(i)}}^{(i)} = l,\ X_{T_n^{(i,\alpha)}}^{(i)} = j\big] \mathbb{P}\big(X_{\tau_\kappa^{(i)}}^{(i)} = l,\ X_{T_n^{(i,\alpha)}}^{(i)} = j\big) \\
&\leqslant C, \quad n > N,
\end{aligned}$$

for some constant $C > 0$ independent of $n > N$, where we applied (S.53), see also Anantharam et al. (1987).

Remark 14.1. *Note that in general we do not have*

$$\mathbb{E}\big[\tau_\kappa^{(i)} - \tau\big] \leqslant \operatorname*{Max}_{j\in\mathbb{S}} \mu_j^{(i)}(j)$$

for any stopping time τ. For example, if τ is the first hitting time of state 0 by the two-state chain with transition matrix $P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$, we have

$$\mathbb{E}\big[\tau_\kappa^{(i)} - \tau\big] = \mu_0(0)\mathbb{P}(X_1^{(i)} = 0) + \mu_1(0)\mathbb{P}(X_1^{(i)} = 1)$$

$$
\begin{aligned}
&= \mathbb{P}(X_1^{(i)} = 0)\left(1 + \frac{a}{b}\right) + \frac{1}{a}\mathbb{P}(X_1^{(i)} = 1) \\
&= \big((1-a)\mathbb{P}(X_0^{(i)} = 0) + b\mathbb{P}(X_0^{(i)} = 1)\big)\left(1 + \frac{a}{b}\right) \\
&\quad + \frac{1}{a}\big(a\mathbb{P}(X_0^{(i)} = 0) + (1-b)\mathbb{P}(X_0^{(i)} = 1)\big).
\end{aligned}
$$

In particular, when $a = b$ we find

$$
\begin{aligned}
\mathbb{E}\big[\tau_\kappa^{(i)} - \tau\big] &= 2\big((1-a)\mathbb{P}(X_0^{(i)} = 0) + a\mathbb{P}(X_0^{(i)} = 1)\big) \\
&\quad + \mathbb{P}(X_0^{(i)} = 0) + \frac{1-a}{a}\mathbb{P}(X_0^{(i)} = 1),
\end{aligned}
$$

which does not remain bounded as a tends to zero, whereas in this case

$$
\operatorname*{Max}_{j\in\mathbb{S}} \mu_j^{(i)}(j) = \operatorname{Max}\left(\frac{a+b}{a}, \frac{a+b}{b}\right) = 2.
$$

iii) Letting

$$
K := 2\sum_{i=1}^N \operatorname*{Max}_{l,j\in\mathbb{S}} \mu_l^{(i)}(j),
$$

we have

$$
\begin{aligned}
\mathcal{R}_n^\alpha &= n\pi_1^{(N)} - \mathbb{E}\left[\sum_{k=1}^n X_k^{(\alpha_k)}\right] \\
&\leqslant K + n\pi_1^{(N)} - \sum_{i=1}^N \pi_1^{(i)}\mathbb{E}[T_n^{(i,\alpha)}], \qquad n > N.
\end{aligned}
$$

b) Bounding the modified regret.

i) If none of the stated conditions, hold, *i.e.* if

$$
\widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log n}{T_{n-1}^{(N,\alpha^*)}}} > \pi_1^{(N)}, \ \ \widehat{m}_{n-1}^{(i,\alpha^*)} \leqslant \pi_1^{(i)} + \sqrt{\frac{L\log n}{T_{n-1}^{(i,\alpha^*)}}}, \ \ T_{n-1}^{(i,\alpha^*)} \geqslant \frac{4L\log n}{(\pi_1^{(N)} - \pi_1^{(i)})^2}.
$$

then we have

$$
\begin{aligned}
\widehat{m}_{n-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log n}{T_{n-1}^{(N,\alpha^*)}}} &> \pi_1^{(N)} \\
&= \pi_1^{(i)} + \pi_1^{(N)} - \pi_1^{(i)} \\
&\geqslant \pi_1^{(i)} + 2\sqrt{\frac{L\log n}{T_{n-1}^{(i,\alpha^*)}}}
\end{aligned}
$$

$$\geqslant \widehat{m}_{n-1}^{(i,\alpha^*)} + \sqrt{\frac{L\log n}{T_{n-1}^{(i,\alpha^*)}}},$$

which implies $\alpha_n^* \neq i$.

ii) We have

$$\begin{aligned}
T_n^{(i,\alpha^*)} &= \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \\
&= \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)}<\widehat{n}_i\}} + \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)}\geqslant\widehat{n}_i\}} \\
&= \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k}^{(i,\alpha^*)}\leqslant\widehat{n}_i\}} + \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)}\geqslant\widehat{n}_i\}} \\
&\leqslant \widehat{n}_i + \sum_{k=1}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)}\geqslant\widehat{n}_i\}} \\
&\leqslant \widehat{n}_i + \sum_{k>\widehat{n}_i}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\{T_{k-1}^{(i,\alpha^*)}\geqslant\widehat{n}_i\}} \\
&\leqslant \widehat{n}_i + \sum_{k>\widehat{n}_i}^n \mathbb{1}_{\{\alpha_k^*=i\}} \mathbb{1}_{\left\{T_{k-1}^{(i,\alpha^*)}\geqslant\frac{4L\log k}{(\pi_1^{(N)}-\pi_1^{(i)})^2}\right\}} \\
&\leqslant \widehat{n}_i + \sum_{k=1+\widehat{n}_i}^n \mathbb{1}_{\left\{\widehat{m}_{k-1}^{(N,\alpha^*)}+\sqrt{(L\log k)/T_{k-1}^{(N,\alpha^*)}}\leqslant\pi_1^{(N)}\right\}} \\
&\quad + \sum_{k=1+\widehat{n}_i}^n \mathbb{1}_{\left\{\widehat{m}_{k-1}^{(N,\alpha^*)}>\pi_1^{(i)}+\sqrt{(L\log k)/T_{k-1}^{(i,\alpha^*)}}\right\}},
\end{aligned}$$

hence

$$\begin{aligned}
\mathbb{E}\big[T_n^{(i,\alpha^*)}\big] &\leqslant \widehat{n}_i + \sum_{k=\widehat{n}_i+1}^n \mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant \pi_1^{(N)}\right) \\
&\quad + \sum_{k=\widehat{n}_i+1}^n \mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} > \pi_1^{(i)} + \sqrt{\frac{L\log k}{T_{k-1}^{(i,\alpha^*)}}}\right),
\end{aligned}$$

see § 2.2 of Bubeck and Cesa-Bianchi (2012).

iii) By Question (h) of Problem 7.28, we have

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(N,\alpha^*)} + \sqrt{\frac{L\log k}{T_{k-1}^{(N,\alpha^*)}}} \leqslant \pi_1^{(N)}\right)$$

$$\leqslant \mathbb{P}\left(\exists l \in \{1,\ldots,k\} \ : \ \frac{1}{l}\sum_{j=1}^{l}\big(X_j^{(N)} - \pi_1^{(N)}\big) + \sqrt{\frac{L\log k}{l}} \leqslant \pi_1^{(N)}\right)$$
$$\leqslant \sum_{l=1}^{k}\mathbb{P}\left(\frac{1}{l}\sum_{j=1}^{l}\big(X_j^{(N)} - \pi_1^{(N)}\big) + \sqrt{\frac{L\log k}{l}} \leqslant \pi_1^{(N)}\right)$$
$$\leqslant \sum_{l=1}^{k}\mathbb{P}\left(\frac{1}{l}\sum_{j=1}^{l}\big(1 - X_j^{(N)} - \big(1 - \pi_1^{(N)}\big)\big) \geqslant \sqrt{\frac{L\log k}{l}}\right)$$
$$\leqslant \sum_{l=1}^{k} e^{-2(1-\lambda_N)(L\log k)/(1+\lambda_N)}$$
$$= \sum_{l=1}^{k}\frac{1}{k^{2L(1-\lambda)/(1+\lambda)}}$$
$$= \frac{1}{k^{2L(1-\lambda)/(1+\lambda)-1}},$$

and similarly

$$\mathbb{P}\left(\widehat{m}_{k-1}^{(i,\alpha^*)} > \pi_1^{(i)} + \sqrt{\frac{L\log k}{T_{k-1}^{(i,\alpha^*)}}}\right)$$
$$\leqslant \mathbb{P}\left(\exists l \in \{1,\ldots,k\} \ : \ \frac{1}{l}\sum_{j=1}^{l} X_j^{(N)} > \pi_1^{(N)} + \sqrt{\frac{L\log k}{l}}\right)$$
$$\leqslant \sum_{l=1}^{k}\mathbb{P}\left(\frac{1}{l}\sum_{j=1}^{l}\big(X_j^{(N)} - \pi_1^{(N)}\big) > \sqrt{\frac{L\log k}{l}}\right)$$
$$\leqslant \sum_{l=1}^{k} e^{-2L(1-\lambda)(\log k)/(1+\lambda)}$$
$$= \frac{1}{k^{2L(1-\lambda)/(1+\lambda)-1}}.$$

iv) We have

$$\begin{aligned}\mathbb{E}\big[T_n^{(i,\alpha^*)}\big] &\leqslant \frac{4L\log n}{(\pi_1^{(N)} - \pi_1^{(i)})^2} + \sum_{k=1}^{n}\frac{2}{k^{2L(1-\lambda)/(1+\lambda)-1}}\\ &\leqslant \frac{4L\log n}{(\pi_1^{(N)} - \pi_1^{(i)})^2} + \int_1^n \frac{2}{t^{2L(1-\lambda)/(1+\lambda)-1}}dt\\ &\leqslant \frac{4L\log n}{(\pi_1^{(N)} - \pi_1^{(i)})^2} + \frac{1}{L(1-\lambda)/(1+\lambda)-1}\left(1 - \frac{1}{n^{2L(1-\lambda)/(1+\lambda)-2}}\right),\end{aligned}$$

hence

$$\begin{aligned}\overline{\mathcal{R}}n_n^{\alpha^*} &= n\pi_1^{(N)} - \mathbb{E}\left[\sum_{k=1}^n \pi_{\alpha_k^*}\right] \\ &= \sum_{k=1}^n \mathbb{E}\big[\pi_1^{(N)} - \pi_1^{(\alpha_k^*)}\big] \\ &= n\pi_1^{(N)} - \sum_{i=1}^N \pi_1^{(i)}\mathbb{E}[T_n^{(i,\alpha^*)}] \\ &= \sum_{i=1}^N \big(\pi_1^{(N)} - \pi_1^{(i)}\big)\mathbb{E}[T_n^{(i,\alpha^*)}] \\ &\leqslant (\log n)\sum_{i=1}^{N-1} \frac{4L}{\pi_1^{(N)} - \pi_1^{(i)}} + \sum_{i=1}^N \frac{\pi_1^{(N)} - \pi_1^{(i)}}{L(1-\lambda)/(1+\lambda)-1},\end{aligned}$$

provided that $L > (1+\lambda)/(1-\lambda)$.

Problem 7.31

a) We have

$$\mathbb{P}(T_k - T_{k-1} = m) = \frac{k}{N}\left(1-\frac{k}{N}\right)^{m-1}, \qquad m \geqslant 1, \quad k = 1,\ldots,N-1,$$

i.e. $T_k - T_{k-1}$ has a geometric distribution started at 1, with parameter $p_l := 1 - l/N$, $l = 1,\ldots,N-1$.

b) We have

$$\mathbb{E}[T_k] = \sum_{l=1}^k \mathbb{E}[T_l - T_{l-1}] = \sum_{l=1}^k \frac{N}{l},$$

and in particular

$$\mathbb{E}[T_{N-1}] = \sum_{l=1}^{N-1} \frac{N}{l}.$$

c) We have

$$\mathrm{Var}[T_k] = \sum_{l=1}^k \mathrm{Var}[T_l - T_{l-1}] = \sum_{l=1}^k \frac{p_l}{(1-p_l)^2} = \sum_{l=1}^k \frac{N^2}{l^2}\left(1-\frac{l}{N}\right),$$

and in particular

$$\mathrm{Var}[T_{N-1}] = \sum_{l=1}^{N-1} \frac{N^2}{l^2}\left(1-\frac{l}{N}\right) \leqslant CN^2.$$

d) Since

$$\mathbb{E}[T_{N-1}] = \sum_{k=1}^{N-1} \frac{N}{k} \leqslant N(1+\log N),$$

we have, for N large enough,

$$\begin{aligned}
\mathbb{P}(T_{N-1} > (1+a)N\log N) &= \mathbb{P}(T_{N-1} - \mathbb{E}[T_{N-1}] > (1+a)N\log N - \mathbb{E}[T_{N-1}]) \\
&\leqslant \mathbb{P}(T_{N-1} - \mathbb{E}[T_{N-1}] > (1+a)N\log N - N(1+\log N)) \\
&\leqslant \mathbb{P}(T_{N-1} - \mathbb{E}[T_{N-1}] > aN\log N - N) \\
&\leqslant \frac{\mathrm{Var}[T_{N-1}]}{(aN\log N - N)^2} \\
&\leqslant \frac{CN^2}{(N(-1+a\log N))^2} \\
&= \frac{C}{(-1+a\log N)^2}.
\end{aligned}$$

e) The distribution of X_n given that $1+T_{N-1} \leqslant n$ is uniform on $\mathbb{S}$, because at time $1+T_{N-1}$ all cards have been uniformly displaced, including the original bottom card after it reached the top position at time T_{N-1}.

f) Since X_n has the uniform distribution π given that $1+T_{N-1} \leqslant n$, from the answers to Questions (b) and (d), for N large enough we find the convergence rate in total variation to the uniform distribution

$$\begin{aligned}
\big\|\mathbb{P}\big(X_{1+(1+a)N\log N} \in \cdot\big) - \pi\big\|_{\mathrm{TV}} &= \operatorname*{Sup}_{A\subset \mathbb{S}} \big|\mathbb{P}\big(X_{1+(1+a)N\log N} \in A\big) - \pi(A)\big| \\
&\leqslant \mathbb{P}\big(1+T_{N-1} > 1+(1+a)N\log N\big) \\
&\leqslant \frac{C}{(-1+a\log N)^2},
\end{aligned}$$

provided that $a > 0$.

Remark. It can also be shown that

$$\liminf_{N\to\infty} \big\|\mathbb{P}\big(X_{(1+a)N\log N} \in \cdot\big) - \pi\big\|_{\mathrm{TV}} > 0,$$

for all $a \in (-1,0)$, which shows that the speed $N\log N$ is optimal for the convergence of the random shuffling $(X_n)_{n\geqslant 0}$ to the uniform distribution on $\mathbb{S}$ in total variation distance as N tends to infinity.

In addition to the top-to-random shuffle, other types of shuffling include the random transpositions shuffle, the transposing neighbors shuffle, the overhand shuffle, the riffle shuffle, etc.

Problem 7.32

a) We have

$$\mathbb{P}(Y_{n+1}=j \mid Y_n=i) = \frac{\mathbb{P}(Y_{n+1}=j \text{ and } Y_n=i)}{\mathbb{P}(Y_n=i)}$$
$$= \frac{\mathbb{P}(Y_{n+1}=j)}{\mathbb{P}(Y_n=i)} \frac{\mathbb{P}(Y_n=i \text{ and } Y_{n+1}=j)}{\mathbb{P}(Y_{n+1}=j)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i)} \frac{\mathbb{P}(X_{N-n}=i \text{ and } X_{N-n-1}=j)}{\mathbb{P}(X_{N-n-1}=j)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i)} \mathbb{P}(X_{N-n}=i \mid X_{N-n-1}=j) = \frac{\pi_j}{\pi_i} P_{j,i}.$$

On the other hand, we have

$$\mathbb{P}(Y_{n+1}=j \mid Y_n=i_n,\ldots,Y_0=i_0) = \frac{\mathbb{P}(Y_{n+1}=j, Y_n=i_n,\ldots,Y_0=i_0)}{\mathbb{P}(Y_n=i_n,\ldots,Y_0=i_0)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j, X_{N-n}=i_n,\ldots,X_N=i_0)}{\mathbb{P}(X_{N-n}=i_n,\ldots,X_N=i_0)}$$
$$= \mathbb{P}(X_{N-n-1}=j \text{ and } X_{N-n}=i_n)$$
$$\times \frac{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0 \mid X_{N-n-1}=j, X_{N-n}=i_n)}{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i_n)} \frac{\mathbb{P}(X_{N-n-1}=j \text{ and } X_{N-n}=i_n)}{\mathbb{P}(X_{N-n-1}=j)}$$
$$\times \frac{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0 \mid X_{N-n-1}=j, X_{N-n}=i_n)}{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0 \mid X_{N-n}=i_n)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i_n)} \mathbb{P}(X_{N-n}=i_n \mid X_{N-n-1}=j,)$$
$$\times \frac{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0 \mid X_{N-n-1}=j)}{\mathbb{P}(X_{N-n+1}=i_{n-1},\ldots,X_N=i_0 \mid X_{N-n}=i_n)}$$
$$= \frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i_n)} \mathbb{P}(X_{N-n}=i_n \mid X_{N-n-1}=j) = \frac{\pi_j}{\pi_{i_n}} P_{j,i_n},$$

and this shows that

$$\mathbb{P}(Y_{n+1}=j \mid Y_n=i_n,\ldots,Y_0=i_0) = \mathbb{P}(Y_{n+1}=j \mid Y_n=i_n) = \frac{\pi_j}{\pi_{i_n}} P_{j,i_n},$$

i.e. the time-reversed process $(Y_n)_{n=0,1,\ldots,N}$ has the Markov property.

b) We find

$$P_{i,j} = \frac{\pi_i}{\pi_j} P_{j,i}, \qquad \text{(S.54)}$$

i.e.

$$\pi_i P_{i,j} = \pi_j P_{j,i},$$

which is the *detailed balance condition* with respect to the probability distribution $\pi = (\pi_i)_{i \in S}$.

c) We have

$$\pi_j = \sum_i \pi_j P_{j,i} = \sum_i \pi_i P_{i,j} = [\pi P]_j.$$

d) According to the detailed balance condition (S.54), we have

$$\begin{aligned} P_{k_1,k_2} P_{k_2,k_3} \cdots P_{k_n,k_1} &= P_{k_n,k_1} \prod_{i=1}^{n-1} P_{k_i,k_{i+1}} = P_{k_n,k_1} \prod_{i=1}^{n-1} \frac{\pi_{k_{i+1}}}{\pi_{k_i}} P_{k_{i+1},k_i} \\ &= \frac{\pi_{k_n}}{\pi_{k_1}} P_{k_n,k_1} \prod_{i=1}^{n-1} P_{k_{i+1},k_i} = P_{k_1,k_n} \prod_{i=1}^{n-1} P_{k_{i+1},k_i}, \end{aligned}$$

holds for all sequences $\{k_1, k_2, \ldots, k_n\}$ of states and $n \geqslant 2$.

e) If the Markov chain satisfies

$$P_{i,k_2} P_{k_2,k_3} \cdots P_{k_{n-1},j} P_{j,i} = P_{i,j} P_{j,k_{n-1}} \cdots P_{k_3,k_2} P_{k_2,i}$$

then by summation over the indices $k_2, k_3, \ldots, k_{n-1}$, using the matrix power relation

$$[P^{n-1}]_{i,j} = \sum_{k_2,\ldots,k_{n-1}} P_{i,k_2} P_{k_2,k_3} \cdots P_{k_{n-1},j},$$

we get

$$\begin{aligned} [P^{n-1}]_{i,j} P_{j,i} &= P_{j,i} \sum_{k_2,\ldots,k_{n-1}} P_{i,k_2} P_{k_2,k_3} \cdots P_{k_{n-1},j} \\ &= P_{i,j} \sum_{k_2,\ldots,k_{n-1}} P_{j,k_{n-1}} \cdots P_{k_3,k_2} P_{k_2,i} \\ &= P_{i,j} [P^{n-1}]_{j,i}. \end{aligned}$$

On the other hand, taking the limit as n goes to infinity, Theorem 7.8 shows that

$$\lim_{n \to \infty} [P^{n-1}]_{j,i} = \lim_{n \to \infty} [P^n]_{j,i} = \pi_i, \qquad i, j \in \mathbb{S},$$

since the limiting and stationary distributions coincide, and we get

$$\pi_j P_{j,i} = P_{i,j} \pi_i, \qquad i, j \in \mathbb{S},$$

which is the detailed balance condition.

f) The detailed balance condition reads

$$\pi_i P_{i,i+1} = \pi_i \left(\frac{1}{2} - \frac{i}{2M}\right) = \pi_{i+1} P_{i+1,i} = \pi_{i+1} \frac{i+1}{2M},$$

hence

$$\frac{\pi_{i+1}}{\pi_i} = \frac{1 - i/M}{(i+1)/M} = \frac{M-i}{i+1},$$

which shows that

$$\pi_i = \frac{(M-i+1)}{i} \frac{(M-i+2)}{(i-1)} \cdots \frac{(M-1)}{2} \frac{M}{1} \pi_0 = \frac{M!}{i!(M-i)!} \pi_0 = \pi_0 \binom{M}{i},$$

$i = 0, 1, \ldots, M$, where the constant $\pi_0 > 0$ is given by

$$1 = \sum_i \pi_i = \pi_0 \sum_{i=0}^{M} \binom{M}{i} = \pi_0 2^M,$$

hence $\pi_0 = 2^{-M}$ and

$$\pi_i = \frac{1}{2^M} \binom{M}{i}, \qquad i = 0, 1, \ldots, M.$$

g) We have

$$\begin{aligned}
[\pi P]_i &= P_{i+1,i}\pi_{i+1} + P_{i,i}\pi_i + P_{i-1,i}\pi_{i-1} \\
&= \frac{1}{2^M} \frac{i+1}{2M} \binom{M}{i+1} + \frac{1}{2^M} \left(\frac{1}{2} - \frac{i-1}{2M}\right) \binom{M}{i-1} + \frac{1}{2^M} \times \frac{1}{2} \binom{M}{i} \\
&= \frac{1}{2^M} \frac{i+1}{2M} \frac{M!}{(i+1)!(M-i-1)!} + \frac{1}{2^M} \frac{M-i+1}{2M} \frac{M!}{(i-1)!(M-i+1)!} \\
&\quad + \frac{1}{2^M \times 2} \frac{M!}{i!(M-i)!} \\
&= \frac{1}{2} \left(\frac{1}{2^M} \frac{(M-1)!}{i!(M-i-1)!} + \frac{1}{2^M} \frac{(M-1)!}{(i-1)!(M-i)!} \right) + \frac{1}{2^{M+1}} \binom{M}{i} \\
&= \frac{1}{2^{M+1}} \binom{M-1}{i} + \frac{1}{2^{M+1}} \binom{M-1}{i-1} + \frac{1}{2^{M+1}} \binom{M}{i} \\
&= \frac{1}{2^M} \binom{M}{i},
\end{aligned}$$

which is also known as Pascal's triangle.

h) The chain is positive recurrent, irreducible and aperiodic, therefore by Theorem 7.8 it admits a limiting distribution equal to π.

i) We have

$$\lim_{n\to\infty} \mathbb{E}[X_n \mid X_0 = i] = \lim_{n\to\infty} \sum_{j=0}^{M} j \mathbb{P}(X_n = j \mid X_0 = i)$$

$$= \sum_{j=0}^{M} j \lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{j=0}^{M} j\pi_j$$
$$= \frac{1}{2^M}\sum_{j=0}^{M} j\binom{M}{j} = \frac{1}{2^M}\sum_{j=0}^{M} j\frac{M!}{j!(M-j)!}$$
$$= \frac{M}{2^M}\sum_{j=1}^{M}\frac{(M-1)!}{(j-1)!(M-j)!} = \frac{M}{2^M}\sum_{j=0}^{M-1}\frac{(M-1)!}{j!(M-1-j)!} = \frac{M}{2},$$

independently of $i = 0, 1, \ldots, M$.

j) Clearly, the relation

$$\mathbb{E}\left[X_0 - \frac{M}{2} \;\middle|\; X_0 = i\right] = \left(i - \frac{M}{2}\right)$$

holds when n =0. Next, assuming that the relation holds at the rank $n \geqslant 0$, we have

$$h_{n+1}(i) = \mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_0 = i\right]$$
$$= P_{i,i+1}\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i+1\right] + P_{i,i}\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i\right]$$
$$+P_{i,i-1}\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i-1\right]$$
$$= \left(\frac{1}{2} - \frac{i}{2M}\right)\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i+1\right] + \frac{1}{2}\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i\right]$$
$$+\frac{i}{2M}\mathbb{E}\left[X_{n+1} - \frac{M}{2} \;\middle|\; X_1 = i-1\right]$$
$$= \left(\frac{1}{2} - \frac{i}{2M}\right)\mathbb{E}\left[X_n - \frac{M}{2} \;\middle|\; X_0 = i+1\right] + \frac{1}{2}\mathbb{E}\left[X_n - \frac{M}{2} \;\middle|\; X_0 = i\right]$$
$$+\frac{i}{2M}\mathbb{E}\left[X_n - \frac{M}{2} \;\middle|\; X_0 = i-1\right]$$
$$= \left(\frac{1}{2} - \frac{i}{2M}\right)\left(i + 1 - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^n + \frac{1}{2}\left(i - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^n$$
$$+\frac{i}{2M}\left(i - 1 - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^n = \left(i - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^{n+1}, \qquad n \geqslant 0,$$

for all $i = 0, 1, \ldots, M$.

Taking the limit as n goes to infinity, we get

$$\lim_{n\to\infty}\mathbb{E}\left[X_n - \frac{M}{2} \;\middle|\; X_0 = i\right] = \lim_{n\to\infty}\left(i - \frac{M}{2}\right)\left(1 - \frac{1}{M}\right)^n = 0,$$

hence

$$\lim_{n\to\infty} \mathbb{E}[X_n \mid X_0 = i] = \frac{M}{2},$$

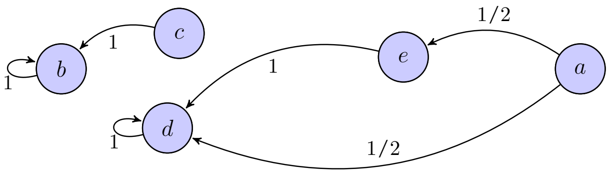
for all $i = 0, 1, \ldots, M$, which recovers the result of Question (i).

Problem 7.33

a) The transition matrix of the chain $(X_n)_{n\geqslant 0}$ is given as follows:

$$P = \begin{bmatrix} 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

b) The chain $(X_n)_{n\in\mathbb{N}}$ admits the following graph, and is clearly reducible:



c) Starting from state ⓐ, ⓓ or ⓔ, the limiting distribution is $(0, 0, 0, 1, 0)$, starting from state ⓑ or ⓒ, the limiting distribution is $(0, 1, 0, 0, 0)$, so that although the chain admits limiting distributions, it does *not* admit a limiting distribution independent of the initial state. More precisely, it can be checked that the powers P^n of the transition matrix P take the form

$$P^n = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{for all } n \geqslant 2, \text{ hence } \quad \lim_{n\to\infty} P^n = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

d) The equation $\pi = \pi P$ is satisfied by any probability distribution of the form

$$\pi = [\pi_a, \pi_b, \pi_c, \pi_d, \pi_e] = [0, p, 0, 1-p, 0],$$

with $p \in [0, 1]$. The stationary distribution is not unique here because the chain is reducible.

e) All rows in the matrix $\widetilde{P}$ clearly add up to 1, so $\widetilde{P}$ is a Markov transition matrix. On the other hand, all states become accessible from each other so that the new chain is irreducible and all states have period 1.

f) Since the chain is irreducible, aperiodic and has a finite state space, we know by Corollary 7.9 that it admits a unique stationary distribution $\tilde{\pi}$. The equation $\tilde{\pi} = \tilde{\pi}\widetilde{P}$ reads

$$\begin{aligned}\tilde{\pi} &= \tilde{\pi}\widetilde{P} \\ &= \frac{\varepsilon}{n}\tilde{\pi}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} + (1-\varepsilon)\tilde{\pi}P \\ &= \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P.\end{aligned}$$

g) The equation

$$\tilde{\pi} = \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P$$

reads

$$[\pi_a, \pi_b, \pi_c, \pi_d, \pi_e] = \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}\begin{bmatrix} 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

which admits the solution

$$\begin{cases} \pi_a = \dfrac{\varepsilon}{5}, \\[2mm] \pi_b = \dfrac{2-\varepsilon}{5}, \\[2mm] \pi_c = \dfrac{\varepsilon}{5}, \\[2mm] \pi_d = \dfrac{(2-\varepsilon)(3-\varepsilon)}{10}, \\[2mm] \pi_e = \dfrac{(3-\varepsilon)\varepsilon}{10}. \end{cases} \tag{S.55}$$

h) We note that

$$\pi_a = \pi_c < \pi_e < \pi_b < \pi_d,$$

hence we will rank the states as

Rank	State
1	d
2	b
3	e
4	$a \simeq c$

based on the idea that the most visited states should rank higher. In the graph of Figure S.20 the stationary distribution is plotted as a function of $\varepsilon \in [0,1]$.

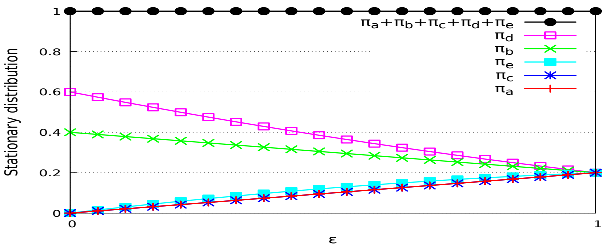


Fig. S.20: Stationary distribution as a function of $\varepsilon \in [0,1]$.

We note that the ranking of states is clearer for smaller values of ε. On the other hand, ε cannot be be chosen too large, for example taking $\varepsilon = 1$ makes all mean return times and equal and corresponds to a uniform stationary distribution.

```
install.packages("igraph")
install.packages("markovchain")
library("igraph")
library(markovchain)
P<-matrix(c(0,0,0,0.5,0.5,0,1,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,0),nrow=5,byrow=TRUE)
MC <-new("markovchain",transitionMatrix=P,states=c("a","b","c","d","e"))
graph <- as(MC, "igraph")
plot(graph,vertex.size=50,edge.label.cex=2,edge.label=E(graph)$prob,edge.color='black',
      vertex.color='dodgerblue',vertex.label.cex=3)
page_rank(graph,damping=0.97)
$vector
       a b c d e
0.00600 0.39400 0.00600 0.58509 0.00891
```

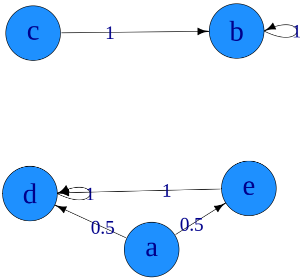


Fig. S.21: Markovchain package output.

i) By Corollary 7.9, we find

$$\begin{cases} \mu_a(a) = \dfrac{5}{\varepsilon} \\ \\ \mu_b(b) = \dfrac{5}{2-\varepsilon} \\ \\ \mu_c(c) = \dfrac{5}{\varepsilon} \\ \\ \mu_d(d) = \dfrac{10}{(2-\varepsilon)(3-\varepsilon)} \\ \\ \mu_e(e) = \dfrac{10}{\varepsilon(3-\varepsilon)}. \end{cases}$$

In the graph of Figure S.22 the mean return times are plotted as a function of $\varepsilon \in [0,1]$. A commonly used value in the literature is $\varepsilon = 1/7$.

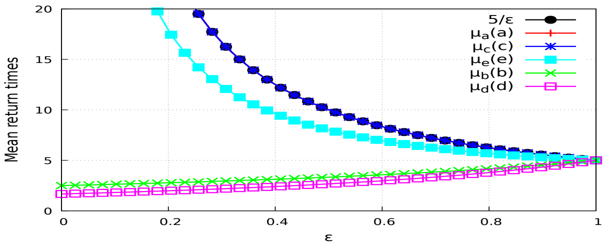


Fig. S.22: Mean return times as functions of $\varepsilon \in [0,1]$.

For small values of ε the mean return times can be higher, and therefore the simulations may take a longer time.

Problem 7.34

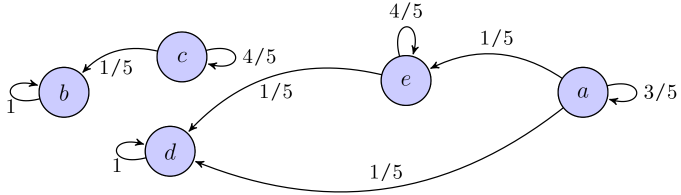
a) The ranking table is completed as follows:

$\preceq$	a	b	c	d	e
a	$=$	$\npreceq$	$\npreceq$	$\preceq$	$\preceq$
b	$\npreceq$	$=$	$\succeq$	$\npreceq$	$\npreceq$
c	$\npreceq$	$\preceq$	$=$	$\npreceq$	$\npreceq$
d	$\succeq$	$\npreceq$	$\npreceq$	$=$	$\succeq$
e	$\succeq$	$\npreceq$	$\npreceq$	$\preceq$	$=$

b) The state space of the chain $(X_n)_{n\in\mathbb{N}}$ is (a,b,c,d,e) and its transition matrix is

$$P = \begin{bmatrix} 3/5 & 0 & 0 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/5 & 4/5 \end{bmatrix}.$$

c) The chain $(X_n)_{n\in\mathbb{N}}$ admits the following graph, and is clearly reducible:



d) Starting from state ⓐ, ⓓ or ⓔ, the limiting distribution is $(0,0,0,1,0)$, starting from state ⓑ or ⓒ, the limiting distribution is $(0,1,0,0,0)$, so that although the chain admits limiting distributions, it does *not* admit a limiting distribution independent of the initial state.

More precisely, it can be checked that the power P^n of order $n \geqslant 1$ of the transition matrix P takes the form

$$P^n = \begin{bmatrix} (3/5)^n & 0 & 0 & 1-(4/5)^n & a_n \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1-(4/5)^n & (4/5)^n & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1-(4/5)^n & (4/5)^n \end{bmatrix}$$

where $a_n = (4/5)^n - (3/5)^n$ since the sum of probabilities over the rows of P^n is equal to 1, hence

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

e) The equation $\pi = \pi P$ reads

$$\begin{aligned} \pi &= [\pi_a, \pi_b, \pi_c, \pi_d, \pi_e] \\ &= \pi \begin{bmatrix} 3/5 & 0 & 0 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/5 & 4/5 \end{bmatrix} \\ &= [3\pi_a/5, \pi_b + \pi_c/5, 4\pi_c/5, \pi_a/5 + \pi_d + \pi_e/5, \pi_a/5 + 4\pi_e/5], \end{aligned}$$

i.e.

$$[0,0,0,0,0] = [-2\pi_a/5, \pi_c/5, -\pi_c/5, \pi_a/5 + \pi_e/5, \pi_a/5 - \pi_e/5],$$

or

$$\begin{cases} \pi_a = 0 \\ \pi_c = 0 \\ \pi_e = 0. \end{cases}$$

As a consequence, any probability distribution of the form

$$\pi = [\pi_a, \pi_b, \pi_c, \pi_d, \pi_e] = [0, p, 0, 1-p, 0],$$

with $p \in [0,1]$, will be a stationary distribution for the chain with matrix P. The stationary distribution is not unique here because the chain is reducible.

f) All rows in the matrix $\widetilde{P}$ clearly add up to 1, so $\widetilde{P}$ is a Markov transition matrix. On the other hand, all states become accessible from each other so that the new chain is irreducible and all states have period 1.

g) Since the chain is irreducible, aperiodic and has a finite state space, we know by Corollary 7.9 that it admits a unique stationary distribution $\tilde{\pi}$. The equation $\tilde{\pi} = \tilde{\pi}\widetilde{P}$ reads

$$\begin{aligned}\tilde{\pi} &= \tilde{\pi}\widetilde{P} \\ &= \frac{\varepsilon}{n}\tilde{\pi}\begin{bmatrix} 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \end{bmatrix} + (1-\varepsilon)\tilde{\pi}P \\ &= \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P.\end{aligned}$$

h) The equation

$$\tilde{\pi} = \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}P$$

reads

$$\begin{aligned}[\pi_a, \pi_b, \pi_c, \pi_d, \pi_e] &= \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] + (1-\varepsilon)\tilde{\pi}\begin{bmatrix} 3/5 & 0 & 0 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/5 & 4/5 \end{bmatrix} \\ &= \left[\frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}, \frac{\varepsilon}{5}\right] \\ &\quad + (1-\varepsilon)\left[3\pi_a/5, \pi_b + \pi_c/5, 4\pi_c/5, \pi_a/5 + \pi_d + \pi_e/5, \pi_a/5 + 4\pi_e/5\right],\end{aligned}$$

i.e.

$$\begin{aligned}&[0,0,0,0,0] \\ &= [\varepsilon + 3(1-\varepsilon)\pi_a - 5\pi_a, \varepsilon + 5(1-\varepsilon)\pi_b - 5\pi_b + (1-\varepsilon)\pi_c, \varepsilon + 4(1-\varepsilon)\pi_c - 5\pi_c, \\ &\quad \varepsilon + (1-\varepsilon)\pi_a + 5(1-\varepsilon)\pi_d - 5\pi_d + (1-\varepsilon)\pi_e, \varepsilon + (1-\varepsilon)\pi_a + 4(1-\varepsilon)\pi_e - 5\pi_e],\end{aligned}$$

i.e.

$$\begin{cases} \varepsilon - (2+3\varepsilon)\pi_a = 0 \\ \varepsilon - 5\varepsilon\pi_b + (1-\varepsilon)\pi_c = 0 \\ \varepsilon - \pi_c(1+4\varepsilon) = 0 \\ \varepsilon + (1-\varepsilon)\pi_a - 5\varepsilon\pi_d + (1-\varepsilon)\pi_e = 0 \\ \varepsilon + (1-\varepsilon)\pi_a - (1+4\varepsilon)\pi_e = 0, \end{cases}$$

i.e.

$$\begin{cases} \pi_a = \dfrac{\varepsilon}{2+3\varepsilon}, \\ \pi_b = \dfrac{2+3\varepsilon}{5(1+4\varepsilon)}, \\ \pi_c = \dfrac{\varepsilon}{1+4\varepsilon}, \\ \pi_d = \dfrac{3+2\varepsilon}{5(1+4\varepsilon)}, \\ \pi_e = \dfrac{\varepsilon(3+2\varepsilon)}{(1+4\varepsilon)(2+3\varepsilon)}. \end{cases}$$

We also check that

$$\begin{aligned} &\pi_a + \pi_b + \pi_c + \pi_d + \pi_e \\ &= \frac{\varepsilon}{2+3\varepsilon} + \frac{2+3\varepsilon}{5(1+4\varepsilon)} + \frac{\varepsilon}{1+4\varepsilon} + \frac{3+2\varepsilon}{5(1+4\varepsilon)} + \frac{\varepsilon(3+2\varepsilon)}{(1+4\varepsilon)(2+3\varepsilon)} \\ &= \frac{5\varepsilon(1+4\varepsilon)}{5(2+3\varepsilon)(1+4\varepsilon)} + \frac{(2+3\varepsilon)^2}{5(1+4\varepsilon)(2+3\varepsilon)} + \frac{5\varepsilon(2+3\varepsilon)}{5(1+4\varepsilon)(2+3\varepsilon)} \\ &\quad + \frac{(3+2\varepsilon)(2+3\varepsilon)}{5(2+3\varepsilon)(1+4\varepsilon)} + \frac{5\varepsilon(3+2\varepsilon)}{5(1+4\varepsilon)(2+3\varepsilon)} \\ &= \frac{5\varepsilon(1+4\varepsilon) + (2+3\varepsilon)(5+10\varepsilon) + 5\varepsilon(3+2\varepsilon)}{5(1+4\varepsilon)(2+3\varepsilon)} \\ &= 1. \end{aligned}$$

i) We note that

$$\pi_a < \pi_c < \pi_e < \pi_b < \pi_d,$$

hence we will rank the states as

Rank	State
1	d
2	b
3	e
4	c
5	a

based on the idea that the most visited states should rank higher. In the graph of Figure S.23 the stationary distribution is plotted as a function of $\varepsilon \in [0, 1]$.

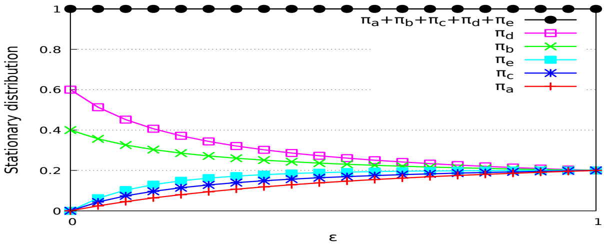


Fig. S.23: Stationary distribution as a function of $\varepsilon \in [0,1]$.

j) By Corollary 7.9, we find

$$\begin{cases} \mu_a(a) = 3 + \dfrac{2}{\varepsilon} \\ \\ \mu_b(b) = \dfrac{5(1+4\varepsilon)}{2+3\varepsilon} \\ \\ \mu_c(c) = 4 + \dfrac{1}{\varepsilon} \\ \\ \mu_d(d) = \dfrac{5(1+4\varepsilon)}{3+2\varepsilon} \\ \\ \mu_e(e) = \dfrac{(1+4\varepsilon)(2+3\varepsilon)}{\varepsilon(3+2\varepsilon)}. \end{cases}$$

In the graph of Figure S.24 the mean return times are plotted as a function of $\varepsilon \in [0,1]$. A commonly used value in the literature is $\varepsilon = 1/7$.

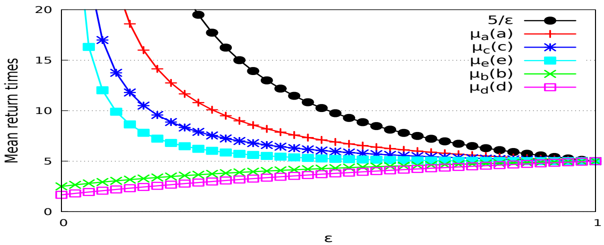


Fig. S.24: Mean return times as functions of $\varepsilon \in [0,1]$.

We note that the ranking of states is clearer for smaller values of ε. In particular ε cannot be be chosen too large, for example taking $\varepsilon = 1$ makes all mean return times and equal and corresponds to a uniform stationary distribution. However the mean return times can be higher and hence the simulations can take a longer time for small values of ε.

Problem 7.35 (cf. Levin et al. (2009)-§ 4.3-4.5)

a) For any two probability distributions $\mu = [\mu_1, \mu_2, \ldots, \mu_N]$ and $\nu = [\nu_1, \nu_2, \ldots, \nu_N]$ on $\{1, 2, \ldots, N\}$ we have

$$\begin{aligned}\|\mu - \nu\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{k=1}^{N} |\mu_k - \nu_k| \\ &\leqslant \frac{1}{2}\sum_{k=1}^{N} (\mu_k + \nu_k) \\ &= \frac{1}{2}\sum_{k=1}^{N} \mu_k + \frac{1}{2}\sum_{k=1}^{N} \nu_k \\ &= 1.\end{aligned}$$

b) We have

$$\begin{aligned}\|\mu P - \nu P\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{j=1}^{N} |[\mu P]_j - [\nu P]_j| \\ &= \frac{1}{2}\sum_{j=1}^{N} \left|\sum_{i=1}^{n} \mu_i P_{i,j} - \sum_{i=1}^{n} \nu_i P_{i,j}\right| \\ &\leqslant \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{n} P_{i,j}|\mu_i - \nu_i| \\ &= \frac{1}{2}\sum_{i=1}^{n} |\mu_i - \nu_i| \sum_{j=1}^{N} P_{i,j} \\ &= \frac{1}{2}\sum_{i=1}^{n} |\mu_i - \nu_i|.\end{aligned}$$

c) Replacing μ and ν with μP^n and π in the result of Question (b) we find

$$\begin{aligned}\|\mu P^{n+1} - \pi\|_{\mathrm{TV}} &= \|(\mu P^n)P - \pi P\|_{\mathrm{TV}} \\ &\leqslant \|\mu P^n - \pi\|_{\mathrm{TV}}.\end{aligned}$$

d) Letting $k \in \{1, 2, \ldots, N\}$ and taking

$$\mu := (0, \ldots, 0, \underset{\substack{\uparrow \\ k}}{1}, 0, \ldots, 0)$$

we have $\mu P^{n+1} = [P^{n+1}]_{k,\cdot}$ and by Question (c) we find

$$\begin{aligned}\|[P^{n+1}]_{k,\cdot} - \pi\|_{\mathrm{TV}} &= \|\mu P^{n+1} - \pi P\|_{\mathrm{TV}} \\ &\leqslant \|\mu P^n - \pi\|_{\mathrm{TV}} \\ &= \|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}}.\end{aligned}$$

Taking the maximum over $k = 1, 2, \ldots, N$ in the above inequality yields

$$d(n+1) = \operatorname*{Max}_{k=1,2,\ldots,N} \|[P^{n+1}]_{k,\cdot} - \pi\|_{\mathrm{TV}} \leqslant \operatorname*{Max}_{k=1,2,\ldots,N} \|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} = d(n),$$

$n \in \mathbb{N}$.

e) The chain is irreducible because all states can communicate in one time step since $P_{i,j} > 0$, $1 \leqslant i, j \leqslant N$. In addition the chain is aperiodic as all states have period one, given that $P_{i,i} > 0$, $i = 1, 2, \ldots, N$. Since the state space is finite, Corollary 6.2 shows that all states are positive recurrent, hence by Corollary 7.9 the chain admits a limiting and a stationary distribution that are equal.

f) We note that Q_θ can be written as

$$\begin{aligned} Q_\theta &= [\ [Q_\theta]_{i,j}\ ]_{1 \leqslant i,j \leqslant N} \\ &= \begin{bmatrix} [Q_\theta]_{1,1} & [Q_\theta]_{1,2} & \cdots & [Q_\theta]_{1,N} \\ [Q_\theta]_{2,1} & [Q_\theta]_{2,2} & \cdots & [Q_\theta]_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ [Q_\theta]_{N,1} & [Q_\theta]_{N,2} & \cdots & [Q_\theta]_{N,N} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{1-\theta}(P_{1,1} - \theta\pi_1) & \frac{1}{1-\theta}(P_{1,2} - \theta\pi_2) & \cdots & \frac{1}{1-\theta}(P_{1,N} - \theta\pi_N) \\ \frac{1}{1-\theta}(P_{2,1} - \theta\pi_1) & \frac{1}{1-\theta}(P_{2,2} - \theta\pi_2) & \cdots & \frac{1}{1-\theta}(P_{2,N} - \theta\pi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{1-\theta}(P_{N,1} - \theta\pi_1) & \frac{1}{1-\theta}(P_{N,2} - \theta\pi_2) & \cdots & \frac{1}{1-\theta}(P_{N,N} - \theta\pi_N) \end{bmatrix} \end{aligned}$$

Clearly, all entries of Q_θ are nonnegative due to the condition

$$P_{i,j} \geqslant \theta\pi_j, \qquad i, j = 1, 2, \ldots, N.$$

In addition, for all $i = 1, 2, \ldots, N$ we have

$$\begin{aligned}\sum_{j=1}^{N}[Q_\theta]_{i,j} &= \frac{1}{1-\theta}\sum_{j=1}^{N}(P_{i,j}-\theta\Pi_{i,j})\\ &= \frac{1}{1-\theta}\sum_{j=1}^{N}(P_{i,j}-\theta\pi_j)\\ &= \frac{1}{1-\theta}\sum_{j=1}^{N}P_{i,j}-\frac{\theta}{1-\theta}\sum_{j=1}^{N}\pi_j\\ &= \frac{1}{1-\theta}-\frac{\theta}{1-\theta}\\ &= 1, \qquad 0<\theta<1,\end{aligned}$$

and we conclude that Q_θ is a Markov transition matrix.

g) Clearly, the property holds for $n=1$ by the definition of Q_θ. Next, assume that

$$P^n = \Pi + (1-\theta)^n\left(Q_\theta^n - \Pi\right)$$

for some $n \geqslant 1$. Noting that the condition $\pi P = \pi$ implies $\Pi P = \Pi$, we have

$$\begin{aligned}P^{n+1} &= \big(\Pi + (1-\theta)^n(Q_\theta^n-\Pi)\big)P\\ &= \Pi P + (1-\theta)^n Q_\theta^n P - (1-\theta)^n \Pi P\\ &= \Pi + (1-\theta)^n Q_\theta^n P - (1-\theta)^n\Pi\\ &= \Pi + (1-\theta)^n Q_\theta^n\big(\Pi + (1-\theta)(Q_\theta - \Pi)\big) - (1-\theta)^n\Pi\\ &= \Pi + \theta(1-\theta)^n Q_\theta^n\Pi + (1-\theta)^{n+1}Q_\theta^{n+1} - (1-\theta)^n\Pi.\end{aligned}$$

Next, we note that since Q_θ is a Markov transition matrix by Question (f) we have $Q_\theta\Pi = \Pi$, in other words we have $P\Pi = \Pi^2 = \Pi$, and

$$Q_\theta\Pi = \frac{1}{1-\theta}\left(P-\theta\Pi\right)\Pi = \frac{1}{1-\theta}\left(P\Pi - \theta\Pi^2\right) = \frac{1}{1-\theta}\left(\Pi - \theta\Pi\right) = \Pi,$$

and more generally $Q_\theta^n\Pi = \Pi$, $n \geqslant 1$, hence

$$\begin{aligned}P^{n+1} &= \Pi + \theta(1-\theta)^n Q_\theta^n\Pi + (1-\theta)^{n+1}Q_\theta^{n+1} - (1-\theta)^n\Pi\\ &= \Pi + \theta(1-\theta)^n\Pi + (1-\theta)^{n+1}Q_\theta^{n+1} - (1-\theta)^n\Pi\\ &= \Pi + (1-\theta)^{n+1}Q_\theta^{n+1} - (1-\theta)^{n+1}\Pi\\ &= \Pi + (1-\theta)^{n+1}\big(Q_\theta^{n+1}-\Pi\big).\end{aligned}$$

h) Let $k \in \{1,2,\ldots,N\}$. By Question (g) we have

$$\|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} = \|[P^n]_{k,\cdot} - \Pi_{k,\cdot}\|_{\mathrm{TV}}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{j=1}^{N} |[P^n]_{k,j} - \pi_j| \\
&= \frac{1}{2}\sum_{j=1}^{N} |(1-\theta)^n [Q_\theta^n]_{k,j} - (1-\theta)^n \pi_j| \\
&= \frac{(1-\theta)^n}{2}\sum_{j=1}^{N} |[Q_\theta^n]_{k,j} - \pi_j| \\
&= (1-\theta)^n \|[Q_\theta^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} \\
&\leqslant (1-\theta)^n, \qquad n \geqslant 0,
\end{aligned}
$$

where we applied the result of Question (a), since $\Pi_{k,\cdot} = \pi$ is a probability distribution and the same holds for $[Q_\theta^n]_{k,\cdot}$ for all $k = 1, 2, \ldots, N$ by Question (f).

The relation

$$
\|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} = (1-\theta)^n \|[Q_\theta^n]_{k,\cdot} - \pi\|_{\mathrm{TV}}, \qquad n \in \mathbb{N},
$$

also shows that, in total variation distance, at each time step the chain associated to P converges faster (by a factor $1-\theta$) to π than the chain associated to Q_θ.

Finally, we find

$$
d(n) = \operatorname*{Max}_{k=1,2,\ldots,N} \|[P^n]_{k,\cdot} - \pi\|_{\mathrm{TV}} \leqslant (1-\theta)^n, \qquad n \geqslant 0.
$$

i) If $t_{\mathrm{mix}} = 0$ the inequality is clearly satisfied, so that we can suppose that $t_{\mathrm{mix}} \geqslant 1$. By the definition of t_{mix} and the result of Question (h) we have

$$
\frac{1}{4} < d(t_{\mathrm{mix}} - 1) \leqslant (1-\theta)^{t_{\mathrm{mix}}-1},
$$

hence

$$
\log\frac{1}{4} < \log d(t_{\mathrm{mix}} - 1) \leqslant \log\left((1-\theta)^{t_{\mathrm{mix}}-1}\right) = (t_{\mathrm{mix}} - 1)\log(1-\theta),
$$

and

$$
t_{\mathrm{mix}} - 1 \leqslant \frac{\log d(t_{\mathrm{mix}} - 1)}{\log(1-\theta)} < \frac{\log 1/4}{\log(1-\theta)}.
$$

Hence we have

$$
t_{\mathrm{mix}} < 1 + \frac{\log 1/4}{\log(1-\theta)},
$$

which yields

$$t_{\text{mix}} < 1 + \left\lceil \frac{\log 1/4}{\log(1-\theta)} \right\rceil,$$

and finally

$$t_{\text{mix}} \leqslant \left\lceil \frac{\log 1/4}{\log(1-\theta)} \right\rceil.$$

j) Given the transition matrix

$$P = \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

and its stationary distribution

$$\pi = [\pi_1, \pi_2, \pi_3] = [11/24, 9/24, 4/24],$$

we check that in order to satisfy all nine conditions $P_{i,j} \geqslant \theta \pi_j$, $i, j = 1, 2, 3$, the value of θ should be in the range $[0, 4/11]$. The optimal value of θ is the one that minimizes the bound $\left\lceil \frac{\log 1/4}{\log(1-\theta)} \right\rceil$, *i.e.* $\theta = 4/11$, and

$$t_{\text{mix}} \leqslant \left\lceil \frac{\log 1/4}{\log 7/11} \right\rceil = \lceil 3.067 \rceil = 4.$$

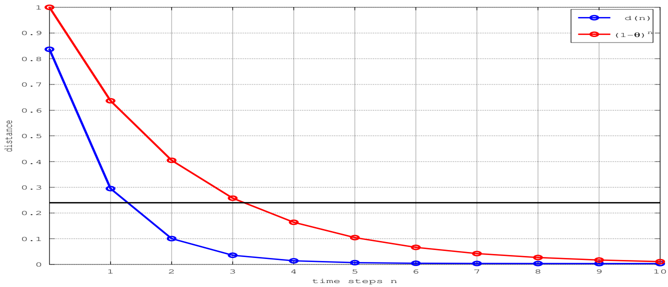


Fig. S.25: Graphs of distance to stationarity $d(n)$ and upper bound $(1-\theta)^n$.

We check from the above graph that the actual value of the mixing time is $t_{\text{mix}} = 2$. The value of $d(0)$ is the maximum distance between π and all deterministic initial distributions starting from states $k = 1, 2, \ldots, N$.

Remark. We have shown that the conditions $\pi P = \pi$ and $P_{i,j} \geqslant \theta \pi_j$, $i, j = 1, 2, \ldots, N$, for some $\theta \in (0, 1)$, define a unique (stationary) distribution π which is also a limiting distribution independent of the initial state. This is the case in particular when $P_{i,j} > 0$, $i,j = 1,2,\ldots,N$, in which case the chain is irreducible and aperiodic, and admits a unique limiting and stationary distribution. More generally, the result holds when P is *regular*, *i.e.* when there exists $n \geqslant 1$ such that $[P^n]_{i,j} > 0$ for all $i,j = 1,2,\ldots,N$, cf. § 4.3-4.5 of Levin et al. (2009).

Below is the Matlab/Octave code used to generate Figure S.25.

```
P = [2/3,1/6,1/6;
        1/3,1/2,1/6;
        1/6,2/3,1/6;]
pi = [11/24,9/24,4/25]
theta = 4/11
for n = 1:11
y(n)=n-1;
u(n)=0.25;
z(n)=(1-theta)^(n-1);
distance(n) = 0;
for k = 1:3
d = mpower(P,n-1)(k,1:3) - pi;
dist=0;
for i = 1:3
dist = dist + 0.5*abs(d(i));
end
distance(n) = max(distance(n) ,dist);
end
end
graphics_toolkit("gnuplot");
plot(y,distance,'-bo','LineWidth',8,y,z,'-ro','LineWidth',8,y,u,'-k',
        'LineWidth',8)
legend('d(n)','(1-\theta)^n')
set (gca, 'xtick', 1:10)
set (gca, 'ytick', 0:0.1:1)
grid on
xlabel('time steps n')
ylabel('distance')
pause
```

Problem 7.36

a) The cardinality of $\mathbb{S}$ is 2^N.
b) We note that $z_k(z_{k-1}+z_{k+1})/2$ can only take the three possible values $-1, 0, +1$, and treat all cases separately.
c) We have

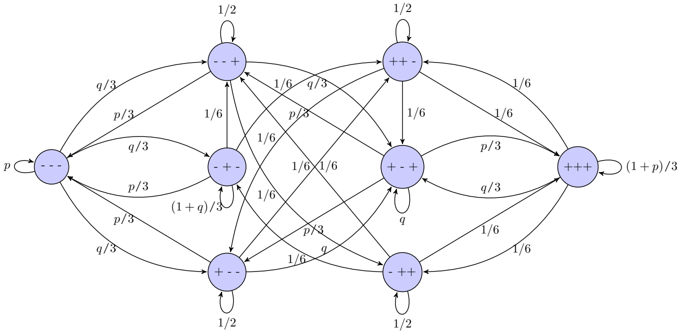
$$\begin{aligned}\mathbb{P}(Z_1 = z \mid Z_0 = z) &= 1 - \sum_{k=1}^{N} \mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z) \\ &= 1 - \frac{1}{N}\sum_{k=1}^{N} \frac{1}{1+(p/q)^{z_k(z_{k-1}+z_{k+1})/2}}, \qquad z \in \mathbb{S}.\end{aligned}$$

d) The transition probability matrix P of $(Z_n)_{n\in\mathbb{N}}$ is given by

$$
P = \begin{array}{c} --- \\ --+ \\ -+- \\ -++ \\ +-- \\ +-+ \\ ++- \\ +++ \end{array}
\overset{\begin{array}{cccccccc} --- & --+ & -+- & -++ & +-- & +-+ & ++- & +++ \end{array}}{\begin{bmatrix}
p & q/3 & q/3 & 0 & q/3 & 0 & 0 & 0 \\
p/3 & 1/2 & 0 & 1/6 & 0 & q/3 & 0 & 0 \\
p/3 & 0 & (1+q)/3 & 1/6 & 0 & 0 & 1/6 & 0 \\
0 & 1/6 & 1/6 & 1/2 & 0 & 0 & 0 & 1/6 \\
p/3 & 0 & 0 & 0 & 1/2 & q/3 & 1/6 & 0 \\
0 & p/3 & 0 & 0 & p/3 & q & 0 & p/3 \\
0 & 0 & 1/6 & 0 & 1/6 & 0 & 1/2 & 1/6 \\
0 & 0 & 0 & 1/6 & 0 & q/3 & 1/6 & (1+p)/3
\end{bmatrix}}.
$$

The chain is irreducible because starting from any configuration $z = (z_k)_{1 \leqslant k \leqslant N} \in \mathbb{S}$ we can reach any other configuration $\hat{z} = (\hat{z}_k)_{1 \leqslant k \leqslant N} \in \mathbb{S}$ in a finite number of time steps. For this, count the number of spins in $z = (z_k)_{1 \leqslant k \leqslant N}$ that differ from the spins in $\hat{z} = (\hat{z}_k)_{1 \leqslant k \leqslant N}$ and flip them one by one until we reach $\hat{z} = (\hat{z}_k)_{1 \leqslant k \leqslant N}$.

Alternatively, we could also enumerate all possible 2^N configurations by flipping one spin at a time, starting from $z = (+1, +1, \ldots, +1)$ until we reach $z = (-1, -1, \ldots, -1)$. When $N = 3$ the chain has the following graph:



From (7.3.7) we note the property

$$
\begin{aligned}
& \mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z) + \mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k) \\
&= \frac{1}{N(1 + (p/q)^{z_k(z_{k-1}+z_{k+1})/2})} + \frac{1}{N(1 + (p/q)^{-z_k(z_{k-1}+z_{k+1})/2})}, \\
&= \frac{1}{N}, \quad k = 1, 2, \ldots, N.
\end{aligned}
$$

e) The chain is irreducible by Question (d), and it has a finite state space of cardinality 2^N by Question (a), hence it is positive recurrent by Corollary 6.14. In addition it is aperiodic since every state has a returning loop because

$$\mathbb{P}(Z_1 = z \mid Z_0 = z) \geqslant \min(p,q) > 0, \qquad z \in \mathbb{S}.$$

Hence by *e.g.* Theorem 7.2 it admits a unique stationary distribution which coincides with its limiting distribution.

f) Under (7.3.9) we have

$$\begin{aligned}
\mathbb{P}(Z_1 = z) &= \mathbb{P}(Z_1 = z \mid Z_0 = z)\pi_z + \sum_{k=1}^N \mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k)\pi_{\bar{z}^k} \\
&= \pi_z \mathbb{P}(Z_1 = z \mid Z_0 = z) + \pi_z \sum_{k=1}^N \mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z) \\
&= \pi_z \left(\mathbb{P}(Z_1 = z \mid Z_0 = z) + \sum_{k=1}^N \mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z) \right) \\
&= \pi_z,
\end{aligned}$$

hence $(\pi_z)_{z\in\mathbb{S}}$ is a stationary distribution.

g) By (7.3.7) we have

$$\begin{aligned}
\frac{\mathbb{P}(Z_1 = \bar{z}^k \mid Z_0 = z)}{\mathbb{P}(Z_1 = z \mid Z_0 = \bar{z}^k)} &= \frac{1 + (p/q)^{\bar{z}^k(\bar{z}^k_{k-1} + \bar{z}^k_{k+1})/2}}{1 + (p/q)^{z_k(z_{k-1}+z_{k+1})/2}} \\
&= \frac{1 + (p/q)^{-z_k(z_{k-1}+z_{k+1})/2}}{1 + (p/q)^{z_k(z_{k-1}+z_{k+1})/2}} \\
&= \frac{q^{z_k(z_{k-1}+z_{k+1})/2}(1 + (q/p)^{z_k(z_{k-1}+z_{k+1})/2})}{q^{z_k(z_{k-1}+z_{k+1})/2} + p^{z_k(z_{k-1}+z_{k+1})/2}} \\
&= \frac{(q/p)^{z_k(z_{k-1}+z_{k+1})/2}(p^{z_k(z_{k-1}+z_{k+1})/2} + q^{z_k(z_{k-1}+z_{k+1})/2})}{q^{z_k(z_{k-1}+z_{k+1})/2} + p^{z_k(z_{k-1}+z_{k+1})/2}} \\
&= \left(\frac{q}{p}\right)^{z_k(z_{k-1}+z_{k+1})/2}, \quad k = 1, 2, \ldots, N.
\end{aligned}$$

h) For all $z \in \mathbb{S}$ we have

$$\begin{aligned}
\pi_{\bar{z}^k} &= C_\beta \exp\left(\beta \sum_{l=0}^{k-2} z_l z_{l+1} - \beta z_{k-1} z_k - \beta z_k z_{k+1} + \beta \sum_{l=k+1}^N z_l z_{l+1} \right) \\
&= C_\beta \exp\left(-2\beta z_k (z_{k-1} + z_{k+1}) + \beta \sum_{l=0}^N z_l z_{l+1} \right) \\
&= \pi_z \mathrm{e}^{-2\beta z_k (z_{k-1}+z_{k+1})}
\end{aligned}$$

$$= \left(\frac{q}{p}\right)^{z_k(z_{k-1}+z_{k+1})/2} \pi_z, \qquad k = 1, 2, \ldots, N,$$

and the *inverse temperature* β is given by

$$\beta = \frac{1}{4} \log \frac{p}{q},$$

i.e.

$$p = \frac{1}{1 + e^{-4\beta}}.$$

The constant C_β is chosen so that

$$\sum_{z \in \mathbb{S}} \pi_z = C_\beta \sum_{z \in \mathbb{S}} \exp\left(\beta \sum_{l=0}^{N} z_l z_{l+1}\right) = 1,$$

i.e.

$$C_\beta = \left(\sum_{z \in \mathbb{S}} \exp\left(\beta \sum_{l=0}^{N} z_l z_{l+1}\right)\right)^{-1}.$$

The stationary distribution $(\pi_z)_{z \in \mathbb{S}}$ is known as the *Boltzmann distribution.*

i) We have

$$\pi = \begin{bmatrix} \pi_{---} \\ \pi_{--+} \\ \pi_{-+-} \\ \pi_{-++} \\ \pi_{+--} \\ \pi_{+-+} \\ \pi_{++-} \\ \pi_{+++} \end{bmatrix} = \begin{bmatrix} C_\beta \\ C_\beta \\ C_\beta \mathrm{e}^{-4\beta} \\ C_\beta \\ C_\beta \\ C_\beta \\ C_\beta \\ C_\beta \mathrm{e}^{4\beta} \end{bmatrix} = C_\beta \begin{bmatrix} 1 \\ 1 \\ q/p \\ 1 \\ 1 \\ 1 \\ 1 \\ p/q \end{bmatrix} = \frac{1}{1+4pq} \begin{bmatrix} pq \\ pq \\ q^2 \\ pq \\ pq \\ pq \\ pq \\ p^2 \end{bmatrix} \begin{matrix} --- \\ --+ \\ -+- \\ -++ \\ +-- \\ +-+ \\ ++- \\ +++ \end{matrix}$$

and from the relation

$$\pi_{---} + \pi_{--+} + \pi_{-+-} + \pi_{-++} + \pi_{+--} + \pi_{+-+} + \pi_{++-} + \pi_{+++} = 1,$$

we find

$$C_\beta = \frac{1}{e^{4\beta} + \mathrm{e}^{-4\beta} + 6}$$

$$= \frac{1}{4\cosh^2(2\beta)+4}$$
$$= \frac{pq}{q^2+p^2+6pq}$$
$$= \frac{1}{6+p/q+q/p}$$
$$= \frac{pq}{1+4pq}.$$

We note that when $p > 1/2$ the configuration "$+++$" has the highest probability p^2, while "$-+-$" has the lowest probability q^2 in the long run, due to the presence of two "opinion leaders" $z_0 = +1$ and $z_4 = +1$ who will not change their minds.

We can also compute the probabilities of having more "+" than "-" in the long run, as

$$\pi_{-++} + \pi_{+-+} + \pi_{++-} + \pi_{+++} = \frac{p(1+2q)}{1+4pq},$$

while the probability of having more "-" than "+" is

$$\pi_{---} + \pi_{--+} + \pi_{-+-} + \pi_{+--} = \frac{q(1+2p)}{1+4pq}.$$

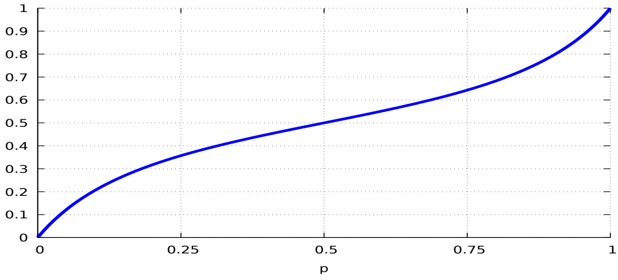


Fig. S.26: Probability of a majority of "+" in the long run as a function of $p \in [0,1]$.

Clearly, the end result is influenced by the boundary conditions $z_0 = z_4 = +1$.

For another example, taking $z_0 = -1$ and $z_4 = +1$, we have

$$\pi = \begin{bmatrix} \pi_{---} \\ \pi_{--+} \\ \pi_{-+-} \\ \pi_{-++} \\ \pi_{+--} \\ \pi_{+-+} \\ \pi_{++-} \\ \pi_{+++} \end{bmatrix} = \begin{bmatrix} C_\beta e^{2\beta} \\ C_\beta e^{2\beta} \\ C_\beta e^{-2\beta} \\ C_\beta e^{2\beta} \\ C_\beta e^{-2\beta} \\ C_\beta e^{-2\beta} \\ C_\beta e^{-2\beta} \\ C_\beta e^{2\beta} \end{bmatrix} = C_\beta \begin{bmatrix} \sqrt{p/q} \\ \sqrt{p/q} \\ \sqrt{q/p} \\ \sqrt{p/q} \\ \sqrt{q/p} \\ \sqrt{q/p} \\ \sqrt{q/p} \\ \sqrt{p/q} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} p \\ p \\ q \\ p \\ q \\ q \\ q \\ p \end{bmatrix} \begin{matrix} --- \\ --+ \\ -+- \\ -++ \\ +-- \\ +-+ \\ ++- \\ +++ \end{matrix}$$

where

$$C_\beta = \frac{1}{4\sqrt{p/q} + 4\sqrt{q/p}} = \frac{\sqrt{pq}}{4}.$$

The probabilities of having more "+" than "-" in the long run are

$$\pi_{-++} + \pi_{+-+} + \pi_{++-} + \pi_{+++} = \frac{1}{2}$$

while the probability of having more "-" than "+" is also

$$\pi_{---} + \pi_{--+} + \pi_{-+-} + \pi_{+--} = \frac{1}{2}.$$

The study undertaken in Bhattacharya and Mukherjee (2018) shows that $\beta = 0.853$ in the Ising model for a Facebook friendship-network with $N = 333$ nodes and 2519 edges, in which spin values refer to gender, *i.e.*

$$p = \frac{1}{1 + e^{-4\beta}} = 0.968077467.$$

A similar analysis on the 2012 US presidential election shows that $\beta = 0.96113$, *i.e.* $p = 0.979051557$.

Problem 7.37 (cf. Lezaud (1998))

a) By the Perron-Frobenius theorem applied to the nonnegative matrix P, the largest eigenvalue λ_0 of P has a single multiplicity and satisfies

$$1 = \min_{1 \leqslant i \leqslant d} \sum_{j=1}^{d} P_{i,j} \leqslant \lambda_0 \leqslant \operatorname{Max}_{1 \leqslant i \leqslant d} \sum_{j=1}^{d} P_{i,j} = 1.$$

Moreover, the eigenvector with eigenvalue $\lambda_0 = 1$ is clearly $\vec{e} = (1, \ldots, 1)$, as $P\vec{e} = \vec{e}$.

b) The projection operator Π onto $\vec{e}$ is the linear mapping given by

$$u \mapsto \Pi(u) = \frac{\langle u, \vec{e}\rangle}{\langle \vec{e}, \vec{e}\rangle}\vec{e} = \langle u, \vec{e}\rangle \vec{e} = \sum_{i=1}^{d} \langle u, \vec{e}\rangle \vec{e}_i,$$

where $\{\vec{e}_1, \ldots, \vec{e}_d\}$ is in the orthogonal basis

$$e_k := (0, \ldots, 0, \underset{\underset{k}{\uparrow}}{1}, 0, \ldots, 0), \qquad k = 1, 2, \ldots, d,$$

of $\mathbb{R}^d$. Its matrix in $\{\vec{e}_1, \ldots, \vec{e}_d\}$ is given by

$$\Pi = (\Pi_{i,j})_{1 \leqslant i,j \leqslant d} = (\langle \vec{e}_j, \vec{e}\rangle)_{1 \leqslant i,j \leqslant d} = (\pi_j)_{1 \leqslant i,j \leqslant d},$$

i.e.

$$\Pi := \begin{bmatrix} \pi \\ \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_d \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_d \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_d \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 & \cdots & \pi_d \end{bmatrix}.$$

We also note that Π is self-adjoint with respect to $\langle \cdot, \cdot \rangle$, as

$$\langle \Pi u, v\rangle = \sum_{i,j=1}^{d} \pi_i \pi_j u_i v_j = \langle u, \Pi v\rangle,$$

and its highest eigenvalue is 1.

c) The equality clearly holds for $n = 0$, due to the convention $\sum_{l=1}^{0} = 0$. Assuming that it holds at the rank $n \geqslant 0$, we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\alpha \sum_{l=1}^{n+1} f(X_l)\right) \,\Big|\, X_0 = k\right] = \mathbb{E}\left[e^{\alpha f(X_1)} \exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right) \,\Big|\, X_0 = k\right] \\
&= \sum_{r=1}^{d} \mathbb{E}\left[\mathbf{1}_{\{X_1 = r\}} e^{\alpha f(X_1)} \exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right) \,\Big|\, X_0 = k\right] \\
&= \frac{1}{\mathbb{P}(X_0 = k)} \sum_{r=1}^{d} e^{\alpha f(r)} \mathbb{E}\left[\mathbf{1}_{\{X_0 = k, X_1 = r\}} \exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right)\right] \\
&= \sum_{r=1}^{d} e^{\alpha f(r)} \frac{\mathbb{P}(X_0 = k, X_1 = r)}{\mathbb{P}(X_0 = k)} \mathbb{E}\left[\exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right) \,\Big|\, X_0 = k, \ X_1 = r\right]
\end{aligned}$$

$$= \sum_{r=1}^{d} e^{\alpha f(r)} \mathbb{P}(X_1 = r \mid X_0 = k) \mathbb{E}\left[\exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right) \Bigm| X_0 = k,\ X_1 = r\right]$$
$$= \sum_{r=1}^{d} e^{\alpha f(r)} P_{k,r} \mathbb{E}\left[\exp\left(\alpha \sum_{l=2}^{n+1} f(X_l)\right) \Bigm| X_1 = r\right]$$
$$= \sum_{r=1}^{d} e^{\alpha f(r)} P_{k,r} \mathbb{E}\left[\exp\left(\alpha \sum_{l=1}^{n} f(X_l)\right) \Bigm| X_= r\right]$$
$$= \sum_{r=1}^{d} P_{k,r} e^{\alpha f(r)} \sum_{l=1}^{d} \left[\left(P e^{\alpha D_f}\right)^n\right]_{r,l}$$
$$= \sum_{l=1}^{d} \left[\left(P e^{\alpha D_f}\right)^{n+1}\right]_{k,l}.$$

d) We have

$$\begin{aligned} e^{\alpha\gamma n} \mathbb{P}\left(\sum_{l=1}^{n} f(X_l) \geqslant n\gamma \Bigm| X_0 = k\right) &= e^{\alpha\gamma n} \mathbb{E}\left[\mathbf{1}_{\left\{\sum_{l=1}^{n} f(X_l) \geqslant n\gamma\right\}} \Bigm| X_0 = k\right] \\ &\leqslant \mathbb{E}\left[\exp\left(\alpha \sum_{l=1}^{n} f(X_l)\right) \Bigm| X_0 = k\right] \\ &= e^{-\alpha\gamma n} \sum_{l=1}^{d} \left[\left(P e^{\alpha D_f}\right)^n\right]_{k,l}, \quad n \geqslant 0. \end{aligned}$$

e) We have

$$\begin{aligned} \sum_{k,l=1}^{d} \pi_k \left[\left(P e^{\alpha D_f}\right)^n\right]_{k,l} &= \left\langle \vec{e}, \left(P e^{\alpha D_f}\right)^n \vec{e} \right\rangle \\ &= \left\langle \vec{e}, e^{-\alpha D_f/2} \left(e^{\alpha D_f/2} P e^{\alpha D_f/2}\right)^n e^{\alpha D_f/2} \vec{e} \right\rangle \\ &= \left\langle e^{-\alpha D_f/2} \vec{e}, \left(e^{\alpha D_f/2} P e^{\alpha D_f/2}\right)^n e^{\alpha D_f/2} \vec{e} \right\rangle \\ &\leqslant \| e^{-\alpha D_f/2} \vec{e} \| \cdot \left\| \left(e^{\alpha D_f/2} P e^{\alpha D_f/2}\right)^n e^{\alpha D_f/2} \vec{e} \right\| \\ &\leqslant \| e^{-\alpha D_f/2} \vec{e} \| \cdot \| e^{\alpha D_f/2} \vec{e} \| \cdot \left\| \left(e^{\alpha D_f/2} P e^{\alpha D_f/2}\right)^n \right\| \\ &\leqslant e^{\alpha} (\lambda_0(\alpha))^n. \end{aligned}$$

f) By Questions (d) and (e) we have

$$\mathbb{P}\left(\sum_{l=1}^{n} f(X_l) \geqslant n\gamma \Bigm| X_0 = k\right) \leqslant e^{-\alpha\gamma n} e^{\alpha} (\lambda_0(\alpha))^n = e^{\alpha - n(\alpha\gamma - \log \lambda_0(\alpha))},$$

$n \geqslant 0$.

g) The first equality follows from the fact that $\Pi P = P$. Next, letting $M = (M_{i,j})_{1\leqslant i,j\leqslant d}$, we have

$$\Pi D_f^n M D_f^m = \left(\sum_{l=1}^d \pi_l e^{nf(l)} M_{l,j} e^{mf(j)}\right)_{1\leqslant i,j\leqslant d},$$

hence

$$\mathrm{tr}(\Pi D_f^n M D_f^m) = \sum_{j=1}^d \sum_{l=1}^d \pi_l e^{nf(l)} M_{l,j} e^{mf(j)} = \langle f^n, M f^m \rangle.$$

h) We apply II-(2.31) in Kato (1995) by matching the expansion

$$P e^{\alpha D_f} = \sum_{n\geqslant 0} \alpha^n P \frac{(D_f)^n}{n!}$$

to II-(2.1) in Kato (1995) and by taking $m = 1$, see page 74 line -1 therein, since by Question (a) the multiplicity of the eigenvalue $\lambda_0(0) = 1$ of P is 1. We have

$$\begin{aligned} c_1 &= -\mathrm{tr}(P D_f S^{(0)}) \\ &= \mathrm{tr}(P D_f \Pi) \\ &= \mathrm{tr}(\Pi P D_f) \\ &= \mathrm{tr}(\Pi D_f) \\ &= \sum_{k=1}^d \pi_k f(k) \\ &= \mathbb{E}[f(X_1)] \\ &= 0, \end{aligned}$$

and

$$c_2 = -\frac{1}{2}\|f\|^2 + \frac{1}{2}\langle f, Sf\rangle \leqslant \frac{1}{2}\langle f, Sf\rangle \leqslant (1-\lambda_1)^{-1}.$$

where we used $S^{(0)} = -\Pi$ and $S^{(1)} = S$. Next, for $n \geqslant 2$ we have

$$c_n = \sum_{p=1}^n \frac{(-1)^p}{p} \sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \mathrm{tr}\left(P \frac{(D_f)^{\nu_1}}{\nu_1!} S^{(k_1)} \cdots P \frac{(D_f)^{\nu_p}}{\nu_p!} S^{(k_p)}\right)$$

$$= \sum_{p=1}^{n} \frac{(-1)^{p+1}}{p} \sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \mathrm{tr}\big(\Pi P(D_f)^{\nu_1} S^{(k'_1)} \cdots S^{(k'_{p-1})} P(D_f)^{\nu_p}\big)$$

$$= \sum_{p=1}^{n} \frac{(-1)^{p+1}}{p}$$

$$\sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \langle f^{\nu_1}, S^{k'_1} P(D_f)^{\nu_2} \cdots S^{k'_{p-2}} P(D_f)^{\nu_{p-1}} S^{k'_{p-1}} P f^{\nu_p}\rangle,$$

where we used $S^{(0)} = -\Pi$, $S^{(n)} = S^n$, Question (g), and the relation $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

i) We have

$$\sum_{\substack{k_1+\cdots+k_p=p-1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \mathbf{1} = \sum_{\substack{\nu_1+\cdots+\nu_p-p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1}} \mathbf{1} = \binom{2p-2}{p-1}.$$

j) Since $|\lambda_1| \leqslant 1$ by the Perron-Frobenius theorem, we have $0 \leqslant 1-\lambda_1 \leqslant 2$, hence

$$c_n = \sum_{p=1}^{n} \frac{(-1)^{p+1}}{p}$$

$$\sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \langle f^{\nu_1}, S^{k'_1} P(D_f)^{\nu_2} \cdots S^{k'_{p-2}} P(D_f)^{\nu_{p-1}} S^{k'_{p-1}} P f^{\nu_p}\rangle$$

$$\leqslant \sum_{p=1}^{n} \frac{1}{p}$$

$$\sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \|f^{\nu_1}\| \cdot \big\|S^{k'_1} P(D_f)^{\nu_2} \cdots S^{k'_{p-2}} P(D_f)^{\nu_{p-1}} S^{k'_{p-1}} P\big\| \cdot \|f^{\nu_p}\|$$

$$\leqslant \sum_{p=1}^{n} \frac{1}{p} \sum_{\substack{\nu_1+\cdots+\nu_p=n \\ k_1+\cdots+k_p=p-1 \\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1 \\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{\nu_1!\cdots\nu_p!} \big\|S^{k'_1} \cdots S^{k'_{p-1}}\big\|$$

$$
\begin{aligned}
&\leqslant \sum_{p=1}^n \frac{1}{p} \sum_{\substack{\nu_1+\cdots+\nu_p=n\\ k_1+\cdots+k_p=p-1\\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1\\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \frac{1}{2^{\nu_1-1}\cdots 2^{\nu_p-1}} \left\| S^{k_1'}\cdots S^{k_{p-1}'} \right\| \\
&\leqslant \sum_{p=1}^n \frac{(1-\lambda_1)^{-(p-1)}}{p2^{n-p}} \sum_{\substack{\nu_1+\cdots+\nu_p=n\\ k_1+\cdots+k_p=p-1\\ \nu_1\geqslant 1,\ldots,\nu_p\geqslant 1\\ k_1\geqslant 0,\ldots,k_p\geqslant 0}} \mathbf{1} \\
&\leqslant \sum_{p=1}^n \frac{((1-\lambda_1)/2)^{-(p-1)}}{p2^{n-1}} \binom{n-1}{p-1}\binom{2p-2}{p-1} \\
&\leqslant \sum_{p=1}^n \frac{((1-\lambda_1)/2)^{-(n-1)}}{p2^{n-1}} \binom{n-1}{p-1}\binom{2p-2}{p-1} \\
&= (1-\lambda_1)^{-(n-1)} \sum_{p=1}^n \frac{1}{p}\binom{n-1}{p-1}\binom{2p-2}{p-1} \\
&\leqslant (1-\lambda_1)^{-(n-1)} \left(1 + \sum_{p=2}^n \frac{1}{p}\binom{n-1}{p-1}\frac{2^{2p-2}}{\sqrt{\pi p}} \right) \\
&\leqslant (1-\lambda_1)^{-(n-1)} \sum_{p=0}^{n-1} \frac{1}{p+1}\binom{n-1}{p} 4^p, \qquad n \geqslant 2.
\end{aligned}
$$

Next, we note that for $x > 0$ we have

$$
\begin{aligned}
\sum_{p=0}^{n-1}\binom{n-1}{p}\frac{x^p}{p+1} &= \frac{1}{x}\int_0^x \sum_{p=0}^{n-1}\binom{n-1}{p} y^p dy \\
&= \frac{1}{x}\int_0^x (1+y)^{n-1} dy \\
&= \frac{(1+x)^n - 1}{nx} \\
&\leqslant \frac{(1+x)^n}{nx},
\end{aligned}
$$

hence, taking $x := 4$ we obtain

$$
c_n \leqslant (1-\lambda_1)^{-(n-1)} \frac{5^n}{4n} \leqslant (1-\lambda_1)^{-(n-1)} \frac{5^n}{25}, \qquad n \geqslant 7.
$$

and we check by hand calculation that the bound

$$
1 + \sum_{p=2}^n \frac{1}{p}\binom{n-1}{p-1}\frac{4^{p-1}}{\sqrt{\pi p}} \leqslant \frac{5^n}{25}
$$

is also valid for $n = 3, 4, 5, 6$, hence we have

$$c_n \leqslant (1-\lambda_1)^{-(n-1)} \frac{5^n}{25}, \qquad n \geqslant 2.$$

k) Noting that $c_1 = 0$, we have

$$\begin{aligned}
\lambda_0(\alpha) &= 1 + \sum_{n\geqslant 2} c_n \alpha^n \\
&\leqslant 1 + \sum_{n\geqslant 2} \frac{5^{n-2}\alpha^n}{(1-\lambda_1)^{n-1}} \\
&\leqslant 1 + \sum_{n\geqslant 2} \frac{5^{n-2}\alpha^n}{(1-\lambda_1)^{n-1}} \\
&\leqslant 1 + \frac{\alpha^2}{1-\lambda_1} \frac{1}{1-5\alpha/(1-\lambda_1)} \\
&= 1 + \frac{\alpha^2}{1-\lambda_1-5\alpha}, \qquad \alpha \in [0, (1-\lambda_1)/5),
\end{aligned}$$

hence

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{l=1}^n f(X_l) \geqslant \gamma\right) &\leqslant \exp\left(\alpha - n\left(\alpha\gamma - \log\left(1 + \frac{\alpha^2}{1-\lambda_1-5\alpha}\right)\right)\right) \\
&\leqslant \exp\left(\frac{1-\lambda_1}{5} - n\gamma\alpha + \frac{n\alpha^2}{1-\lambda_1-5\alpha}\right),
\end{aligned}$$

$\alpha \in [0, (1-\lambda_1)/5)$.

l) We minimize

$$\alpha \mapsto -\gamma\alpha + \frac{\alpha^2}{1-\lambda_1-5\alpha}$$

over $\alpha \in [0, (1-\lambda_1)/5)$ by noting that the vanishing of its derivative

$$5\left(\frac{\alpha}{1-\lambda_1-5\alpha}\right)^2 + 2\frac{\alpha}{1-\lambda_1-5\alpha} - \gamma = 0$$

occurs at

$$\frac{\alpha_*}{1-\lambda_1-5\alpha_*} = \frac{-1+\sqrt{1+5\gamma}}{5},$$

i.e.

$$\alpha_* = (1-\lambda_1)\frac{-1+\sqrt{1+5\gamma}}{5\sqrt{1+5\gamma}} = \frac{(1-\lambda_1)\gamma}{1+5\gamma+\sqrt{1+5\gamma}} < \frac{1-\lambda_1}{5},$$

hence

$$
\begin{aligned}
-\gamma\alpha_* + \frac{\alpha_*^2}{1-\lambda_1-5\alpha_*} &= \alpha_*\left(-\gamma + \frac{\alpha_*}{1-\lambda_1-5\alpha_*}\right) \\
&= \alpha_* \frac{-1-5\gamma+\sqrt{1+5\gamma}}{5} \\
&= (1-\lambda_1)\gamma \frac{-1-5\gamma+\sqrt{1+5\gamma}}{5(1+5\gamma+\sqrt{1+5\gamma})} \\
&= (1-\lambda_1)\gamma \frac{1+5\gamma-(1+5\gamma)^2}{5(1+5\gamma+\sqrt{1+5\gamma})^2} \\
&= -\frac{(1-\lambda_1)\gamma^2(1+5\gamma)}{(1+5\gamma+\sqrt{1+5\gamma})^2} \\
&= -\frac{(1-\lambda_1)\gamma^2}{(1+\sqrt{1+5\gamma})^2} \\
&\leqslant -\frac{(1-\lambda_1)\gamma^2}{(1+\sqrt{6})^2} \\
&\leqslant -\frac{(1-\lambda_1)\gamma^2}{7+2\sqrt{6}} \\
&< -(1-\lambda_1)\frac{\gamma^2}{12}.
\end{aligned}
$$

Problem 7.38 (Wolfer and Kontorovich (2021))

a) For all $i = 1, \ldots, d$ we have

$$
\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^d \left|\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j}\right|\right] &= \frac{1}{n}\sum_{j=1}^d \mathbb{E}\left[\left|\sum_{k=1}^n \mathbf{1}_{\{Z_i(k)=j\}} - nP_{i,j}\right|\right] \\
&\leqslant \frac{1}{n}\sum_{j=1}^d \sqrt{\mathbb{E}\left[\left|\sum_{k=1}^n \left(\mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j}\right)\right|^2\right]} \\
&= \frac{1}{n}\sum_{j=1}^d \sqrt{\mathrm{Var}\left[\sum_{k=1}^n \mathbf{1}_{\{Z_i(k)=j\}}\right]} \\
&= \frac{1}{n}\sum_{j=1}^d \sqrt{n(1-P_{i,j})P_{i,j}} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^d \sqrt{P_{i,j}} \\
&\leqslant \sqrt{\frac{d}{n}}\sqrt{\sum_{j=1}^d P_{i,j}}
\end{aligned}
$$

$$= \sqrt{\frac{d}{n}}, \qquad n \geqslant 1,$$

where we used the Cauchy-Schwarz inequality.

b) Using the inequality $||u| - |v|| \leqslant |u - v|$, $u, v \in \mathbb{R}$, we have

$$\begin{aligned}
&\left| \sum_{j=1}^{d} \left| \frac{1}{n} \mathbf{1}_{\{x=j\}} + \frac{1}{n} \sum_{k=1,\ k\neq i}^{n} \mathbf{1}_{\{z(k)=j\}} - P_{i,j} \right| \right. \\
&\qquad \left. - \sum_{j=1}^{d} \left| \frac{1}{n} \mathbf{1}_{\{y=j\}} + \frac{1}{n} \sum_{k=1,\ k\neq i}^{n} \mathbf{1}_{\{z(k)=j\}} - P_{i,j} \right| \right| \\
&\leqslant \frac{1}{n} \sum_{j=1}^{d} \left| \mathbf{1}_{\{x=j\}} + \sum_{k=1,\ k\neq i}^{n} \mathbf{1}_{\{z(k)=j\}} - P_{i,j} \right. \\
&\qquad \left. - \left(\sum_{j=1}^{d} \mathbf{1}_{\{y=j\}} + \sum_{k=1,\ k\neq i}^{n} \mathbf{1}_{\{z(k)=j\}} - P_{i,j} \right) \right| \\
&= \frac{1}{n} \sum_{j=1}^{d} \left| \mathbf{1}_{\{x=j\}} - \mathbf{1}_{\{y=j\}} \right| \\
&\leqslant \frac{2}{n} := c_i, \qquad i = 1, \ldots, n.
\end{aligned}$$

c) Using McDiarmid's inequality, for all $i = 1, \ldots, d$ we have

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| > \varepsilon \right) \\
&= \mathbb{P}\left(\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| - \mathbb{E}\left[\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| \right] \right. \\
&\qquad\qquad \left. > \varepsilon - \mathbb{E}\left[\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| \right] \right) \\
&\leqslant \mathbb{P}\left(\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| - \mathbb{E}\left[\sum_{j=1}^{d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k)=j\}} - P_{i,j} \right| \right] > \varepsilon - \sqrt{\frac{d}{n}} \right) \\
&\leqslant \exp\left(-\frac{2}{\sum_{i=1}^{d} c_i^2} \operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}} \right)^2 \right)
\end{aligned}$$

$$= \exp\left(-\frac{n}{2}\operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}}\right)^2\right).$$

d) When $\widetilde{N}_i(m) = n \geqslant 1$, we have

$$\begin{aligned}
\widetilde{P}_{i,j}(m) &:= \frac{1}{\widetilde{N}_i(m)} \sum_{k=1}^{m-1} \mathbf{1}_{\{\widetilde{X}_k = i, \widetilde{X}_{k+1} = j\}}, \\
&= \frac{1}{n} \sum_{k=1}^{m-1} \mathbf{1}_{\{\widetilde{X}_k = i, Z_{\widetilde{X}_k}(1+\widetilde{N}_{\widetilde{X}_k}(k)) = j\}}, \\
&= \frac{1}{n} \sum_{k=1}^{m-1} \mathbf{1}_{\{\widetilde{X}_k = i, Z_i(1+\widetilde{N}_i(k)) = j\}}, \\
&= \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Z_i(k) = j\}}, \qquad i, j = 1, \ldots, d.
\end{aligned}$$

e) This follows from the fact that $\widetilde{X}_{k+1}$ has the same distribution as Z_i given that $\widetilde{X}_k = i$.

f) Letting $n_i := \lceil m\pi_i/2 \rceil$, $i = 1, \ldots, d$, letting $c_1 := (1 - 1/\sqrt{2})^2$ we have

$$0 \leqslant \varepsilon - \sqrt{\frac{d}{n}} \leqslant \varepsilon\sqrt{c_1}, \qquad n \geqslant n_i \geqslant 2d/\varepsilon^2,$$

hence

$$\begin{aligned}
&\sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } N_i(m) = n\right) \\
&\quad = \sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widetilde{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } \widetilde{N}_i(m) = n\right) \\
&\quad = \sum_{n=n_i}^{3n_i} \exp\left(-\frac{n}{2}\operatorname{Max}\left(0, \varepsilon - \sqrt{\frac{d}{n}}\right)^2\right) \\
&\quad \leqslant \sum_{n=n_i}^{3n_i} \mathrm{e}^{-2nc_1\varepsilon^2} \\
&\quad \leqslant (2n_i + 1)\mathrm{e}^{-2n_ic_1\varepsilon^2} \\
&\quad \leqslant (2n_i + 1)\mathrm{e}^{-m\pi_ic_1\varepsilon^2},
\end{aligned}$$

provided that $n_i \geqslant 2d/\varepsilon^2$, or $m \geqslant 4d/(\varepsilon^2\pi_i)$.

g) We have

$$\begin{aligned}
&\sum_{i=1}^{d}\sum_{n=n_i}^{3n_i}\mathbb{P}\left(\sum_{j=1}^{d}\big|\widehat{P}_{i,j}(m)-P_{i,j}\big|>\varepsilon \text{ and } N_i(m)=n\right)\\
&\quad\leqslant \sum_{i=1}^{d}(2n_i+1)e^{-c_1 m\pi_i\varepsilon^2}\\
&\quad\leqslant \sum_{i=1}^{d}\frac{2n_i+1}{c_1 m\pi_i\varepsilon^2}e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad\leqslant \sum_{i=1}^{d}\frac{2\lceil m\pi_i/2\rceil+1}{c_1 m\pi_i\varepsilon^2}e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad\leqslant \sum_{i=1}^{d}\frac{m+3/\pi_i}{c_1 m\varepsilon^2}e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad\leqslant \frac{1}{c_1\varepsilon^2}\sum_{i=1}^{d}\left(1+\frac{3}{m\pi_*}\right)e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad= \frac{d}{c_1\varepsilon^2}\left(1+\frac{3}{m\pi_*}\right)e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad\leqslant \frac{d}{c_1\varepsilon^2}\left(1+\frac{3\varepsilon^2}{4d}\right)e^{-c_1 m\pi_i\varepsilon^2/2}\\
&\quad\leqslant \frac{2d}{c_1\varepsilon^2}e^{-c_1 m\pi_*\varepsilon^2/2},
\end{aligned}$$

provided that $m \geqslant 4d/(\varepsilon^2\pi_*)$ and $\varepsilon\in(0,1)$.

h) For all $\varepsilon>0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d}\sum_{j=1}^{d}\big|\widehat{P}_{i,j}(m)-P_{i,j}\big|>\varepsilon\right)\\
&= \mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d}\sum_{j=1}^{d}\big|\widehat{P}_{i,j}(m)-P_{i,j}\big|>\varepsilon \text{ and } \bigcap_{j=1}^{d}\{N_i(m)\in[n_i,3n_i]\}\right)\\
&\quad+\mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d}\sum_{j=1}^{d}\big|\widehat{P}_{i,j}(m)-P_{i,j}\big|>\varepsilon \text{ and } \bigcup_{j=1}^{d}\{N_i(m)\notin[n_i,3n_i]\}\right)\\
&\leqslant \mathbb{P}\left(\bigcup_{i=1,\ldots,d}\left\{\sum_{j=1}^{d}\big|\widehat{P}_{i,j}(m)-P_{i,j}\big|>\varepsilon \text{ and } N_i(m)\in[n_i,3n_i]\right\}\right)\\
&\quad+\mathbb{P}\left(\bigcup_{j=1}^{d}\{N_i(m)\notin[n_i,3n_i]\}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sum_{i=1}^d \mathbb{P}\left(\sum_{j=1}^d \big|\widehat{P}_{i,j}(m)-P_{i,j}\big| > \varepsilon \text{ and } N_i(m)\in[n_i,3n_i]\right)\\
&\quad +\mathbb{P}\left(\bigcup_{j=1}^d \{N_i(m)\notin[n_i,3n_i]\}\right)\\
&= \sum_{i=1}^d \sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^d \big|\widehat{P}_{i,j}(m)-P_{i,j}\big| > \varepsilon \text{ and } N_i(m)=n\right)\\
&\quad +\mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)\notin[n_i,3n_i]).
\end{aligned}
$$

i) Letting $f_i(x) := \mathbf{1}_{\{x=i\}} - \pi_i$, $i=1,\ldots,d$, we have

$$
N_i(m)-(m-1)\pi_i = \sum_{k=1}^{m-1} f_i(X_k)
$$

and

$$
\mathbb{E}[f_i(X_k)] = \mathbb{E}[N_i(m)-(m-1)\pi_i] = \mathbb{E}\left[\sum_{k=1}^{m-1} f_i(X_k)\right] = (m-1)\pi_i = 0,
$$

hence by the bound in Question (l) of Problem 7.37, we have

$$
\begin{aligned}
&\mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)\notin[n_i,3n_i])\\
&= \mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)>3n_i) + \mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)<n_i)\\
&\leqslant \mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)>3(m-1)\pi_i/2)\\
&\quad +\mathbb{P}(\exists i\in\{1,\ldots,d\} \ : \ N_i(m)<2+(m-1)\pi_i/2)\\
&= \mathbb{P}\left(\exists i\in\{1,\ldots,d\} \ : \ \frac{1}{m-1}\sum_{k=1}^{m-1} f_i(X_k) > \frac{\pi_i}{2}\right)\\
&\quad +\mathbb{P}\left(\exists i\in\{1,\ldots,d\} \ : \ \frac{1}{m-1}\sum_{k=1}^{m-1} f_i(X_k) < -\frac{\pi_i}{2}+\frac{2}{m-1}\right)\\
&\leqslant \mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d} \frac{1}{m-1}\sum_{k=1}^{m-1} f_i(X_k) > \frac{\pi_i}{2}\right)\\
&\quad +\mathbb{P}\left(\operatorname*{Max}_{i=1,\ldots,d} \frac{1}{m-1}\sum_{k=1}^{m-1} (-f_i(X_k)) > \frac{\pi_i}{2}-\frac{2}{m-1}\right)\\
&\leqslant e^{(1-\lambda_1)/5} e^{-(1-\lambda_1)m\pi_i^2/48} + e^{(1-\lambda_1)/5} e^{-(1-\lambda_1)m(\pi_i/2-2/(m-1))^2/12}\\
&\leqslant c_2 d e^{-c_3 m(1-\lambda_1)\pi_*^2}, \qquad m\geqslant 2,
\end{aligned}
$$

where $c_2 = 2e^{(1-\lambda_1)/5}$ and

$$c_3 = \mathrm{Max}\left(\frac{1}{48}, \frac{1}{12}\left(1 - \frac{4}{\pi_*(m-1)}\right)\right) \leqslant \frac{5}{12},$$

provided that $m \geqslant 1 + 4/\pi_*$.

j) We upper bound

$$\sum_{i=1}^{d} \sum_{n=n_i}^{3n_i} \mathbb{P}\left(\sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| > \varepsilon \text{ and } N_i(m) = n\right)$$
$$\leqslant \frac{2d}{c_1\varepsilon^2} e^{-c_1 m \pi_* \varepsilon^2/2}$$
$$< \frac{\delta}{2}$$

and

$$\mathbb{P}(\exists i \in \{1, \ldots, d\} \ : \ N_i(m) \notin [n_i, 3n_i]) \leqslant c_2 d e^{-c_3 m(1-\lambda_1)\pi_*^2}$$
$$< \frac{\delta}{2},$$

which yields

$$m > \frac{2}{c_1 \pi_* \varepsilon^2} \log \frac{4d}{\delta c_1 \varepsilon^2}$$

and

$$m > \frac{1}{c_3(1-\lambda_1)\pi_*^2} \log \frac{2c_2 d}{\delta},$$

hence, using the facts that $d \geqslant 2$ and $y + \log x < 2\log x$, $x > e^y$, we find that there is a constant $c > 0$ such that for all

$$m \geqslant c\,\mathrm{Max}\left(\frac{1}{\varepsilon^2 \pi_*} \mathrm{Max}\left(d, \log \frac{d}{\delta\varepsilon}\right), \frac{1}{(1-\lambda_1)\pi_*^2} \log \frac{d}{\delta}\right),$$

we have

$$\mathbb{P}\left(\mathop{\mathrm{Max}}_{i=1,\ldots,d} \sum_{j=1}^{d} \left|\widehat{P}_{i,j}(m) - P_{i,j}\right| \leqslant \varepsilon\right) \geqslant 1 - \delta.$$

For example, taking $\varepsilon = \delta = 5\%$ and $\pi_* = 1/d$ with $d = 26$ we find $m \gtrapprox 62300$.

Chapter 8 - Branching Processes

Exercise 8.1

a) We have

$$\begin{aligned}\mathbb{P}(X_2 > 0) &= 1 - \mathbb{P}(X_2 = 0)\\ &= 1 - \left(\frac{1}{5} + \frac{3}{5} \times \frac{1}{5} + \frac{1}{5} \times \frac{1}{5} \times \frac{1}{5}\right)\\ &= 1 - \frac{41}{125} = \frac{84}{125}.\end{aligned}$$

b) We have

$$\begin{aligned}\mathbb{P}(X_2 = 1) &= \mathbb{P}(X_2 = 1 \text{ and } X_1 = 0) + \mathbb{P}(X_2 = 1 \text{ and } X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 1 \text{ and } X_1 = 2)\\ &= \mathbb{P}(X_2 = 1 \mid X_1 = 0)\mathbb{P}(X_1 = 0)\\ &\quad + \mathbb{P}(X_2 = 1 \mid X_1 = 1)\mathbb{P}(X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 1 \mid X_1 = 2)\mathbb{P}(X_1 = 2)\\ &= 0 \times \frac{1}{5} + \frac{3}{5} \times \frac{3}{5} + 2 \times \frac{1}{5} \times \frac{3}{5} \times \frac{1}{5}\\ &= \frac{51}{125}.\end{aligned}$$

We could also compute

$$\begin{aligned}\mathbb{P}(X_2 = 2) &= \mathbb{P}(X_2 = 2 \text{ and } X_1 = 0) + \mathbb{P}(X_2 = 2 \text{ and } X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 2 \text{ and } X_1 = 2)\\ &= \mathbb{P}(X_2 = 2 \mid X_1 = 0)\mathbb{P}(X_1 = 0)\\ &\quad + \mathbb{P}(X_2 = 2 \mid X_1 = 1)\mathbb{P}(X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 2 \mid X_1 = 2)\mathbb{P}(X_1 = 2)\\ &= 0 \times \frac{1}{5} + \frac{1}{5} \times \frac{3}{5} + \times \frac{1}{5} \times \left(\frac{3}{5}\right)^2 + 2 \times \frac{1}{5} \times \frac{1}{5} \times \frac{1}{5}\\ &= \frac{26}{125},\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(X_2 = 3) &= \mathbb{P}(X_2 = 3 \text{ and } X_1 = 0) + \mathbb{P}(X_2 = 3 \text{ and } X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 3 \text{ and } X_1 = 2)\\ &= \mathbb{P}(X_2 = 3 \mid X_1 = 0)\mathbb{P}(X_1 = 0)\\ &\quad + \mathbb{P}(X_2 = 3 \mid X_1 = 1)\mathbb{P}(X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 3 \mid X_1 = 2)\mathbb{P}(X_1 = 2)\\ &= 2 \times \frac{1}{5} \times \frac{1}{5} \times \frac{3}{5}\\ &= \frac{6}{125},\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(X_2 = 4) &= \mathbb{P}(X_2 = 4 \text{ and } X_1 = 0) + \mathbb{P}(X_2 = 4 \text{ and } X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 4 \text{ and } X_1 = 2)\\ &= \mathbb{P}(X_2 = 4 \mid X_1 = 0)\mathbb{P}(X_1 = 0)\\ &\quad + \mathbb{P}(X_2 = 4 \mid X_1 = 1)\mathbb{P}(X_1 = 1)\\ &\quad + \mathbb{P}(X_2 = 4 \mid X_1 = 2)\mathbb{P}(X_1 = 2)\\ &= \frac{1}{5} \times \frac{1}{5} \times \frac{1}{5}\\ &= \frac{1}{125},\end{aligned}$$

which recovers

$$\mathbb{P}(X_2 \geqslant 1) = \frac{51 + 26 + 6 + 1}{125} = \frac{84}{125}.$$

c) We have

$$\begin{aligned}\mathbb{P}(X_1 = 2 \mid X_2 = 1) &= \frac{\mathbb{P}(X_1 = 2 \text{ and } X_2 = 1)}{\mathbb{P}(X_2 = 1)}\\ &= \mathbb{P}(X_2 = 1 \mid X_1 = 2)\frac{\mathbb{P}(X_1 = 2)}{\mathbb{P}(X_2 = 1)}\\ &= 2 \times \frac{1}{5} \times \frac{3}{5} \times \frac{1}{5} \times \frac{125}{51} = \frac{2}{17}.\end{aligned}$$

We can also compute

$$\begin{aligned}\mathbb{P}(X_1 = 1 \mid X_2 = 1) &= \mathbb{P}(X_2 = 1 \mid X_1 = 1)\frac{\mathbb{P}(X_1 = 1)}{\mathbb{P}(X_2 = 1)}\\ &= \frac{3}{5} \times \frac{3/5}{51/125}\\ &= \frac{15}{17},\end{aligned}$$

which allows us to check that

$$\mathbb{P}(X_1 = 2 \mid X_2 = 1) + \mathbb{P}(X_1 = 1 \mid X_2 = 1) = \frac{2}{17} + \frac{15}{17} = 1,$$

since $\mathbb{P}(X_1 = 0 \mid X_2 = 1) = 0$.

Exercise 8.2

a) We have

$$G_1(s) = \mathbb{E}[s^Y] = s^0\mathbb{P}(Y = 0) + s^1\mathbb{P}(Y = 1) = \frac{1}{2} + \frac{1}{2}s, \qquad s \in \mathbb{R}.$$

b) We prove this statement by induction. Clearly, it holds at the order 1. Next, assuming that (8.3.10) holds at the order $n \geqslant 1$ we get

$$G_{n+1}(s) = G_1(G_n(s)) = G_1\left(1 - \frac{1}{2^n} + \frac{s}{2^n}\right)$$
$$= \frac{1}{2} + \frac{1}{2}\left(1 - \frac{1}{2^n} + \frac{s}{2^n}\right) = 1 - \frac{1}{2^{n+1}} + \frac{s}{2^{n+1}}.$$

Additional comments:

(i) We may also directly note that

$$\mathbb{P}(X_n = 1 \mid X_0 = 1) = \mathbb{P}(Y_1 = 1, Y_2 = 1, \ldots, Y_n = 1) = (\mathbb{P}(Y = 1))^n = \frac{1}{2^n},$$

hence

$$\mathbb{P}(X_n = 0 \mid X_0 = 1) = 1 - \mathbb{P}(X_n = 1 \mid X_0 = 1) = 1 - \frac{1}{2^n},$$

and

$$G_n(s) = \mathbb{P}(X_n = 0 \mid X_0 = 1) + s\mathbb{P}(X_n = 1 \mid X_0 = 1) = 1 - \frac{1}{2^n} + \frac{s}{2^n}.$$

(ii) It is also possible to write

$$G_n(s) = \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} + \frac{s}{2^n} = \frac{s}{2^n} + \frac{1}{2}\sum_{k=0}^{n-1} \frac{1}{2^k}$$
$$= \frac{1}{2^n}s + \frac{1}{2} \times \frac{1 - (1/2)^n}{1 - 1/2} = 1 - \frac{1}{2^n} + \frac{1}{2^n}s,$$

although this is not recommended here.

(iii) It is **wrong** (why?) to write

$$G_n(s) = \mathbb{E}\left[s^{Y_1 + \cdots + Y_{X_n}}\right] = \cancel{\mathbb{E}\left[s^{Y_1}\right] \cdots \mathbb{E}\left[s^{Y_{X_n}}\right]}.$$

Note that the left-hand side is a well-defined deterministic number, while the right-hand side is not well-defined as a random product.

(iv) For $n \geqslant 2$ we do **not** have $G_n(s) = \cancel{(G_1(s))^n}$.

c) We have

$$\mathbb{P}(X_n = 0 \mid X_0 = 1) = G_n(0) = 1 - \frac{1}{2^n}.$$

Additional comments:

(i) We may also directly write

$$\mathbb{P}(X_n = 0 \mid X_0 = 1) = 1 - \mathbb{P}(X_n = 1 \mid X_0 = 1)$$

$$= 1 - \mathbb{P}(Y_1, Y_2 = 1, \ldots, Y_n = 1)$$
$$= 1 - (\mathbb{P}(Y_1 = 1))^n$$
$$= 1 - \frac{1}{2^n}.$$

On the other hand, we do **not** have

$$\mathbb{P}(X_n = 0 \mid X_0 = 1) = \not(\not\mathbb{P}\not(\not Y_1\not = \not 0\not)\not)^n,$$

since the events $\{X_n = 0\}$ and $\{Y_1 = 0, Y_2 = 0, \ldots, Y_n = 0\}$ are **not** equivalent, more precisely we only have

$$\{Y_1 = 0, Y_2 = 0, \ldots, Y_n = 0\} \subsetneqq \{X_n = 0\},$$

hence

$$\frac{1}{2^n} \leqslant \mathbb{P}(X_n = 0 \mid X_0 = 1) = 1 - \frac{1}{2^n}, \qquad n \geqslant 1.$$

(ii) The probability $\beta_n := \mathbb{P}(X_n = 0 \mid X_0 = 1)$ is **not** solution of $G_n(\beta_n) = \beta_n$. It is easy to check that the equality

$$G_n\left(1 - \frac{1}{2^n}\right) = \not 1 \not - \not\frac{1}{2^n}$$

does *not* hold for $n \geqslant 1$.

(iii) In fact, $\{X_n = 0\}$ means that extinction occurs at time n, *or* has already occurred *before* time n.

d) We have

$$\mathbb{E}[X_n \mid X_0 = 1] = G'_n(s)_{|s=1} = (\mathbb{E}[Y_1])^n = \frac{1}{2^n}.$$

Additional comment:

In this simple setting we could also write

$$\begin{aligned}\mathbb{E}[X_n \mid X_0 = 1] &= 0 \times \mathbb{P}(X_n = 0 \mid X_0 = 1) + 1 \times \mathbb{P}(X_n = 1 \mid X_0 = 1)\\ &= \mathbb{P}(X_n = 1 \mid X_0 = 1)\\ &= \frac{1}{2^n}.\end{aligned}$$

e) The extinction probability α is solution of $G_1(\alpha) = \alpha$, *i.e.*

$$\alpha = \frac{1}{2} + \frac{1}{2}\alpha,$$

with unique solution $\alpha = 1$.

Additional comment:

Since the sequence of events $(\{X_n = 0\})_{n\geqslant 1}$ is increasing, we also have

$$\alpha = \mathbb{P}\left(\bigcup_{n\geqslant 1}\{X_n = 0\}\right) = \lim_{n\to\infty}\mathbb{P}(X_n = 0) = \lim_{n\to\infty} 1 - \frac{1}{2^n} = 1.$$

Exercise 8.3

a) We have $G_1(s) = 0.2 + 0.5s + 0.3s^2$ and

$$\mathbb{E}[X_1] = \mathbb{E}[\xi] = G_1'(1) = 0.5 + 2 \times 0.3 = 1.1,$$

hence

$$\mathbb{E}[X_2] = (G_1'(1))^2 = (\mathbb{E}[\xi])^2 = (1.1)^2,$$

by Proposition 8.2. On the other hand, we have

$$\begin{aligned} G_2(s) &= G_1(G_1(s)) \\ &= G_1(0.2 + 0.5s + 0.3s^2) \\ &= 0.2 + 0.5(0.2 + 0.5s + 0.3s^2) + 0.3(0.2 + 0.5s + 0.3s^2)^2 \\ &= 0.312 + 0.31s + 0.261s^2 + 0.09s^3 + 0.027s^4, \end{aligned}$$

with

$$G_2'(s) = 0.31 + 0.522s + 0.27s^2 + 0.108s^3$$

and

$$G_2''(s) = 0.522 + 0.54s + 0.324s^2,$$

hence

$$G_2'(1) = (G_1'(1))^2 = (1.1)^2 = 1.21 \qquad \text{and} \qquad G_2''(1) = 1.386,$$

and

$$\mathbb{E}[X_2^2] = G_2''(1) + G_2'(1) = 1.386 + 1.21 = 2.596.$$

By (1.7.8) this yields

$$\text{Var}[X_2] = 2.596 - (1.21)^2.$$

b) We have

$$G_2(s) = 0.312 + 0.31s + 0.261s^2 + 0.09s^3 + 0.027s^4,$$

hence

$$\mathbb{P}(X_2 = 0) = 0.312, \quad \mathbb{P}(X_2 = 1) = 0.31, \quad \mathbb{P}(X_2 = 2) = 0.261,$$

and

$$\mathbb{P}(X_2 = 3) = 0.09, \quad \mathbb{P}(X_2 = 4) = 0.027.$$

c) We have

$$\begin{aligned}
&\mathbb{P}(X_4 = 0) = G_4(0) \\
&= G_2(G_2(0)) \\
&= 0.312 + 0.31G_2(0) + 0.261(G_2(0))^2 + 0.09(G_2(0))^3 + 0.027(G_2(0))^4 \\
&= 0.312 + 0.31 \times 0.312 + 0.261 \times 0.312^2 + 0.09 \times 0.312^3 + 0.027 \times 0.312^4 \\
&\simeq 0.44314.
\end{aligned}$$

d) We have

$$\mathbb{E}[X_{10}] = (\mathbb{E}[X_1])^{10} = (G_1'(1))^{10} = (1.1)^{10} = 2.59,$$

since the mean population size grows by 10% at each time step.

e) The extinction probability α solves the equation

$$\alpha = G_1(\alpha) = 0.2 + 0.5\alpha + 0.3\alpha^2,$$

i.e.

$$0.3\alpha^2 - 0.5\alpha + 0.2 = 0.3(\alpha - 1)(\alpha - 2/3) = 0,$$

hence $\alpha = 2/3$.

Exercise 8.4

a) We have

$$G_1(s) = \mathbb{P}(Y = 0) + s\mathbb{P}(Y = 1) + s^2\mathbb{P}(Y = 2) = as^2 + bs + c, \qquad s \in \mathbb{R}.$$

b) Letting X_n denote the number of individuals in the population at generation $n \geqslant 0$, we have

$$\mathbb{P}(X_2 = 0 \mid X_0 = 1) = G_1(G_1(0)) = G_1(c) = ac^2 + bc + c.$$

This probability can actually be recovered by pathwise analysis, by noting that in order to reach $\{X_2 = 0\}$ we should have either

i) $Y_1 = 0$ with probability c, or
ii) $Y_1 = 1$ with probability b and then $Y_1 = 0$ with probability c, or
iii) $Y_1 = 2$ with probability a and then $Y_1 = 0$ (two times) with probability c,

which yields

$$\mathbb{P}(X_2 = 0 \mid X_0 = 1) = c + bc + ac^2.$$

c) We have

$$\mathbb{P}(X_2 = 0 \mid X_0 = 2) = (\mathbb{P}(X_2 = 0 \mid X_0 = 1))^2 = (ac^2 + bc + c)^2,$$

as in (8.3.1).

d) The extinction probability α_1 given that $X_0 = 1$ is solution of $G_1(\alpha) = \alpha$, *i.e.*

$$a\alpha^2 + b\alpha + c = \alpha,$$

or

$$0 = a\alpha^2 - (a+c)\alpha + c = (\alpha - 1)(a\alpha - c)$$

from the condition $a + b + c = 1$. The extinction probability α_1 is known to be the smallest solution of $G_1(\alpha) = \alpha$, hence it is $\alpha_1 = c/a$ when $0 < c \leqslant a$. The extinction probability α_2 given that $X_0 = 2$ is $\alpha_2 = (\alpha_1)^2$.

e) When $0 \leqslant a \leqslant c$ we have $\alpha_1 = 1$.

Exercise 8.5

a) We have

$$G_Y(s) = \mathbb{P}(Y = 0) + s\mathbb{P}(Y = 1) + s^2\mathbb{P}(Y = 2) = q^2 + 2spq + s^2p^2 = (q + ps)^2,$$

b) The extinction probability satisfies the equation $\alpha = G_Y(\alpha)$, *i.e.*

$$\alpha = q^2 + 2\alpha pq + \alpha^2 p^2 = (q + \alpha p)^2,$$

with solutions $\alpha = 1$ and $\alpha = q^2/p^2$, hence the extinction probability is

$$\min\left(1, \frac{q^2}{p^2}\right) = \begin{cases} 1 & \text{if } p \leqslant q, \\ \\ \dfrac{q^2}{p^2} & \text{if } p \geqslant q. \end{cases}$$

c) Since $\{Z_1 = 0\} \subset \{Z_2 = 0\}$, this probability is

$$\begin{aligned} \mathbb{P}(Z_2 = 0 \text{ and } Z_1 \geqslant 1) &= \mathbb{P}\left(Z_2 = 0 \bigcap \{Z_1 = 0\}^c\right) \\ &= \mathbb{P}(Z_2 = 0) - \mathbb{P}(Z_1 = 0) \\ &= G_Y(G_Y(0)) - G_Y(0) \\ &= 2pq^3 + q^2p^4. \end{aligned}$$

Alternatively, it can be recovered pathwise as

$$2pq \times q^2 + p^2 \times q^2 \times q^2 = 2pq^3 + p^2q^4.$$

d) When $q \leqslant p$ the extinction probability is q^2/p^2 starting from $Z_0 = 1$, hence when Z_0 has a Poisson distribution with parameter $\lambda > 0$ this extinction probability is

$$\begin{aligned}
\mathbb{P}(T_0 < \infty) &= \sum_{n \geqslant 0} \mathbb{P}(T_0 < \infty \text{ and } Z_0 = n) \\
&= \mathbb{P}(Z_0 = n) \sum_{n \geqslant 0} \mathbb{P}(T_0 < \infty \mid Z_0 = n) \\
&= \sum_{n \geqslant 0} \left(\frac{q^2}{p^2}\right)^n \mathbb{P}(Z_0 = n) \\
&= \mathrm{e}^{-\lambda} \sum_{n \geqslant 0} \frac{1}{n!} \left(\frac{\lambda q^2}{p^2}\right)^n \\
&= \mathrm{e}^{\lambda(q^2/p^2 - 1)} \\
&= \mathrm{e}^{\lambda(q^2 - p^2)/p^2} \\
&= \mathrm{e}^{\lambda(q-p)/p^2},
\end{aligned}$$

whereas when $q \geqslant p$ the extinction probability becomes

$$\sum_{n \geqslant 0} \mathbb{P}(Z_0 = n) = \mathrm{e}^{-\lambda} \sum_{n \geqslant 0} \frac{\lambda^n}{n!} = 1.$$

Exercise 8.6

a) When only red cells are generated, their number at time $n-1$ is s^{n-1}, hence the probability that only red cells are generated up to time n is

$$\begin{aligned}
\frac{1}{4} \times \left(\frac{1}{4}\right)^2 \times \cdots \times \left(\frac{1}{4}\right)^{2^{n-1}} &= \prod_{k=0}^{n-1} \left(\frac{1}{4}\right)^{2^k} \\
&= \left(\frac{1}{4}\right)^{\sum_{k=0}^{n-1} 2^k} \\
&= \left(\frac{1}{4}\right)^{2^n - 1}, \qquad n \geqslant 0.
\end{aligned}$$

b) Since white cells cannot reproduce, the extinction of the culture is equivalent to the extinction of the red cells, and this question can be solved as in the framework of Exercise 8.3. The probability distribution of the number Y of red cells produced from one red cell is

$$\mathbb{P}(Y = 0) = \frac{1}{12}, \qquad \mathbb{P}(Y = 1) = \frac{2}{3}, \qquad \mathbb{P}(Y = 2) = \frac{1}{4},$$

which has the generating function

$$G_1(s) = \mathbb{P}(Y = 0) + s\mathbb{P}(Y = 1) + s^2\mathbb{P}(Y = 2)$$

$$= \frac{1}{12} + \frac{2s}{3} + \frac{s^2}{4} = \frac{1}{12}(1 + 8s + 3s^2),$$

hence the equation $G_1(\alpha) = \alpha$ reads

$$3\alpha^2 - 4\alpha + 1 = 3(\alpha - 1)(\alpha - 1/3) = 0,$$

which has $\alpha = 1/3$ for smallest solution. Consequently, the extinction probability of the culture is equal to $1/3$.

c) The probability that only red cells are generated from time 0 to time n is

$$\frac{1}{3} \times \left(\frac{1}{3}\right)^2 \times \cdots \times \left(\frac{1}{3}\right)^{2^{n-1}} = \prod_{k=0}^{n-1} \left(\frac{1}{3}\right)^{2^k} = \left(\frac{1}{3}\right)^{2^n - 1},$$

$n \geqslant 0$. The probability distribution

$$\mathbf{P}(Y = 0) = \frac{1}{6}, \qquad \mathbf{P}(Y = 1) = \frac{2}{2}, \qquad \mathbf{P}(Y = 2) = \frac{1}{3},$$

of the number Y of red cells has the generating function

$$G_1(s) = \mathbf{P}(Y = 0) + s\mathbf{P}(Y = 1) + s^2\mathbf{P}(Y = 2)$$
$$= \frac{1}{6} + \frac{s}{2} + \frac{s^2}{3} = \frac{1}{12}(2 + 6s + 4s^2),$$

hence the equation $G_1(\alpha) = \alpha$ reads $1 + 3\alpha + 2\alpha^2 = 6\alpha$, or

$$2\alpha^2 - 3\alpha + 1 = 2(\alpha - 1)(\alpha - 1/2) = 0,$$

which has $\alpha = 1/2$ for smallest solution. Consequently, the extinction probability of the culture is equal to $1/2$.

Exercise 8.7 Using the relation*

$$\mathbb{E}[T_0 \mid X_0 = k] \leqslant k\mathbb{E}[T_0 \mid X_0 = 1], \qquad k \in \mathbb{N},$$

we have

$$\mathbb{E}[T_0 \mid X_0 = 1] = \sum_{k \geqslant 0} \mathbb{P}(Y_1 = k)\big(1 + \mathbb{E}[T_0 \mid X_0 = k]\big)$$
$$= \sum_{k \geqslant 0} \mathbb{P}(Y_1 = k) + \sum_{k \geqslant 0} \mathbb{P}(Y_1 = k)\mathbb{E}[T_0 \mid X_0 = k]$$
$$= 1 + \sum_{k \geqslant 0} \mathbb{P}(Y_1 = k)\mathbb{E}[T_0 \mid X_0 = k]$$

* This inequality follows from the relation $\mathbb{E}[\mathrm{Max}(X_1, \ldots, X_n)] \leqslant \mathbb{E}[X_1 + \cdots + X_n]$.

$$\leqslant 1 + \mathbb{E}[T_0 \mid X_0 = 1] \sum_{k \geqslant 0} k \mathbb{P}(Y_1 = k)$$
$$= 1 + \mathbb{E}[T_0 \mid X_0 = 1] \mathbb{E}[Y_1],$$

hence

$$\mathbb{E}[T_0 \mid X_0 = 1] \leqslant \frac{1}{1 - \mathbb{E}[Y_1]} < \infty.$$

Exercise 8.8 This is a particular case of Example (iv) on page 295.

a) We have

$$G_1(s) = q^2 \sum_{n \geqslant 1} n(ps)^{n-1} = \frac{q^2}{(1-ps)^2}.$$

b) The equation $G_1(\alpha) = \alpha$ reads

$$\alpha = \frac{q^2}{(1-p\alpha)^2},$$

i.e.

$$p^2\alpha^3 - 2p\alpha^2 + \alpha - q^2 = 0,$$

which is known to admit $\alpha = 1$ for solution, *i.e.*

$$(\alpha - 1)(p^2\alpha^2 - (1-q^2)\alpha + q^2) = (\alpha - 1)(p^2\alpha^2 - p(1+q)\alpha + q^2) = 0,$$

whose smallest solution

$$\alpha = \frac{p(1+q) - \sqrt{p^2(1+q)^2 - 4p^2q^2}}{2} = p\frac{2 - p - \sqrt{4p - 3p^2}}{2}$$

is the extinction probability.

Exercise 8.9

a) We have $\mathbb{P}(X = k) = (1/2)^{k+1}$, $k \geqslant 0$.
b) The probability generating function of X is given by

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k \geqslant 0} s^k \mathbb{P}(X = k) = \frac{1}{2} \sum_{k \geqslant 0} (s/2)^k = \frac{1}{2-s},$$

$-1 < s \leqslant 1$.
c) The probability we are looking for is

$$\mathbb{P}(X_3 = 0 \mid X_0 = 0) = G_X(G_X(G_X(0))) = \frac{1}{2 - \frac{1}{2-1/2}} = \frac{3}{4}.$$

d) Since giving birth to a girl is equivalent to having at least one child, and this happens to every couple with probability $3/4$, the probability we are looking for is equal to

$$\frac{1}{4}+\frac{3}{4}\times\frac{1}{4}+\left(\frac{3}{4}\right)^2\times\frac{1}{4}=\frac{1}{4}\times\frac{1-(3/4)^3}{1-3/4}=1-(3/4)^3=\frac{37}{64}=0.578125.$$

It can also be recovered from

$$G_Z^{(3)}(s)=G_Z(G_Z(G_Z(s)))=\frac{37}{64}+\frac{27s}{64}$$

at $s=0$, where G_Z is the probability generating function

$$G_Z(s)=\frac{1}{4}+\frac{3s}{4}.$$

Exercise 8.10

a) We have

$$\begin{aligned}G_{Z_1}(s)&=1-\frac{q}{1-p}+qs\sum_{k\geqslant 1}(ps)^{k-1}\\&=1-\frac{q}{1-p}+\frac{qs}{1-ps}\\&=\frac{1-p-q}{1-p}+\frac{qs}{1-ps}.\end{aligned}$$

b) We have

$$G'_{Z_1}(s)=\frac{q}{1-ps}+\frac{pqs}{(1-ps)^2},$$

hence

$$\mathbb{E}[Z_1]=G'_{Z_1}(1)=\frac{q}{1-p}+\frac{pq}{(1-p)^2}=\frac{q(1-p)+pq}{(1-p)^2}=\frac{q}{(1-p)^2}.$$

c) The critical value of q is $q=(1-p)^2$.

d) When $q=(1-p)^2$ we have

$$G_{Z_1}(s)=p+\frac{(1-p)^2s}{1-ps},\qquad -1<s<1.$$

The relation is clearly satisfied at the rank $k=1$. Next, assuming that the relation

$$G_{Z_k}(s)=\frac{kp-(kp+p-1)s}{1-p+kp-kps},\qquad -1<s<1,$$

holds at the rank $k\geqslant 1$, we have

$$G_{Z_{k+1}}(s)=G_{Z_k}(G_{Z_1}(s))$$

$$
\begin{aligned}
&= \frac{kp-(kp+p-1)G_{Z_1}(s)}{1-p+kp-kpG_{Z_1}(s)}, \\
&= \frac{kp-(kp+p-1)\left(p+\dfrac{(1-p)^2s}{1-ps}\right)}{1-p+kp-kp\left(p+\dfrac{(1-p)^2s}{1-ps}\right)} \\
&= \frac{kp(1-ps)-(kp+p-1)\left(p(1-ps)+(1-p)^2s\right)}{(1-p+kp)(1-ps)-kp\left(p(1-ps)+(1-p)^2s\right)} \\
&= \frac{kp(1-ps)-(kp+p-1)\left(p+(1-2p)s\right)}{(1-p+kp)(1-ps)-kp\left(p+(1-2p)s\right)} \\
&= \frac{(1-p)(k+1)p-(1-p)((k+2)p-1)s}{(1-p)(kp+1)-(1-p)(k+1)ps} \\
&= \frac{(k+1)p-((k+1)p+p-1)s}{1-p+(k+1)p-(k+1)ps}, \qquad -1 < s < 1.
\end{aligned}
$$

Exercise 8.11

a) We have

$$
\begin{aligned}
G(s) &= \mathbb{E}[s^X] \\
&= \mathbb{E}\left[s^{1+X_1+\cdots+X_N}\right] \\
&= s\mathbb{E}\left[\prod_{l=1}^{N} s^{X_l}\right] \\
&= s\sum_{k\geqslant 0}\mathbb{E}\left[\prod_{l=1}^{N} s^{X_l} \,\Big|\, N=k\right]\mathbb{P}(N=k) \\
&= s\sum_{k\geqslant 0}\mathbb{E}\left[\prod_{l=1}^{k} s^{X_l} \,\Big|\, N=k\right]\mathbb{P}(N=k) \\
&= s\sum_{k\geqslant 0}\mathbb{E}\left[\prod_{l=1}^{k} s^{X_l}\right]\mathbb{P}(N=k) \\
&= s\sum_{k\geqslant 0}\left(\prod_{l=1}^{k}\mathbb{E}\left[s^{X_l}\right]\right)\mathbb{P}(N=k) \\
&= s\mathrm{e}^{-\mu}\sum_{k\geqslant 0}\left(\mathbb{E}\left[s^{X_1}\right]\right)^k\frac{\mu^n}{n!} \\
&= sG_\mu\left(\mathbb{E}\left[s^{X_1}\right]\right) \\
&= sG_\mu(G(s)), \qquad -1 \leqslant s \leqslant 1,
\end{aligned}
$$

where $(X_k)_{k\geqslant 1}$ denotes a sequence of independent copies of X, see also the recursion in Proposition 8.1 and Relation (13) in Haight and Breuer (1960). This relation can be solved using Lagrange series, see page 145 of Pólya and Szegö (1998), as

$$G(s) = \sum_{n=1}^{\infty} s^n \mathbb{P}(X = n) = \sum_{n=1}^{\infty} s^n \mathrm{e}^{-\mu n} \frac{(\mu n)^{n-1}}{n!},$$

where

$$\mathbb{P}(X = n) = \mathrm{e}^{-\mu n} \frac{(\mu n)^{n-1}}{n!}, \qquad n \geqslant 1,$$

is the Borel distribution, see also Finner et al. (2015).

b) We have

$$G'(s) = G_\mu(G(s)) + sG'(s)G'_\mu(G(s))$$

at $s = 1$, which gives

$$G'(1) = G_\mu(1) + G'_\mu(1)G'(1) = 1 + \mu G'(1),$$

and

$$\mathbb{E}[X] = \frac{1}{1-\mu},$$

which is finite if $\mu < 1$.

c) Similarly, knowing that $G''_\mu(1) = \mu^2$, the relation

$$G''(s) = 2G'(s)G'_\mu(G(s)) + sG''(s)G'_\mu(G(s)) + s(G'(s))^2 G''_\mu(G(s))$$

at $s = 1$ gives

$$\begin{aligned} G''(1) &= 2G'(1)G'_\mu(1) + G''(1)G'_\mu(1) + (G'(1))^2 G''_\mu(1) \\ &= \frac{2\mu - \mu^2}{(1-\mu)^2} + \mu G''(1), \end{aligned}$$

hence

$$G''(1) = \frac{2\mu - \mu^2}{(1-\mu)^3}$$

and

$$\begin{aligned} \mathrm{Var}[X] &= G''(1) + G'(1) - (G'(1))^2 \\ &= \frac{2\mu - \mu^2}{(1-\mu)^3} + \frac{1}{1-\mu} - \frac{1}{(1-\mu)^2} \\ &= \frac{\mu}{(1-\mu)^3}, \end{aligned}$$

see § 7.2.2 of Johnson et al. (2005).

Problem 8.12

a) We have

$$\mathbb{E}[Z_n] = \sum_{k=1}^{n} \mathbb{E}[X_k] = \sum_{k=1}^{n} \mu^k = \mu \sum_{k=0}^{n-1} \mu^k = \mu \frac{1-\mu^n}{1-\mu}, \qquad n \in \mathbb{N}.$$

b) We have

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{k\geqslant 1} X_k\right] = \sum_{k\geqslant 1} \mathbb{E}[X_k] = \mu \sum_{k\geqslant 0} \mu^k = \frac{\mu}{1-\mu}, \qquad n \in \mathbb{N},$$

provided that $\mu < 1$.

c) We have

$$\begin{aligned} H(s) &= \mathbb{E}\big[s^Z \mid X_0 = 1\big] \\ &= s^0 \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \mathbb{E}\big[s^Z \mid X_1 = k\big] \mathbb{P}(Y_1 = k) \\ &= \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \big(\mathbb{E}\big[s^Z \mid X_1 = 1\big]\big)^k \mathbb{P}(Y_1 = k) \\ &= \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \left(\mathbb{E}\left[s^{1+\sum_{i\geqslant 2} X_i} \,\Big|\, X_1 = 1\right]\right)^k \mathbb{P}(Y_1 = k) \\ &= \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \left(\mathbb{E}\left[s^{1+\sum_{i\geqslant 1} X_i} \,\Big|\, X_0 = 1\right]\right)^k \mathbb{P}(Y_1 = k) \\ &= \mathbb{P}(Y_1 = 0) + \sum_{k\geqslant 1} \big(\mathbb{E}\big[s^{1+Z} \mid X_0 = 1\big]\big)^k \mathbb{P}(Y_1 = k) \\ &= \sum_{k\geqslant 0} \big(s\mathbb{E}\big[s^Z \mid X_0 = 1\big]\big)^k \mathbb{P}(Y_1 = k) \\ &= G_1(sH(s)). \end{aligned}$$

d) We have

$$H(s) = G_1(sH(s)) = \frac{1-p}{1-psH(s)},$$

hence

$$psH^2(s) - H(s) + q = 0,$$

and

$$H(s) = \frac{1 \pm \sqrt{1-4pqs}}{2ps} = \frac{1 - \sqrt{1-4pqs}}{2ps},$$

where we have chosen the minus sign since the plus sign leads to $H(0) = +\infty$ whereas we should have $H(0) = \mathbb{P}(Z = 0) \leqslant 1$. In addition we have

$\mu = p/q < 1$ hence $p < 1/2 < q$ and the minus sign gives

$$H(1) = \frac{1-\sqrt{1-4pq}}{2p} = \frac{1-|q-p|}{2p} = 1.$$

e) We have

$$\lim_{s \searrow 0^+} H(s) = \lim_{s \searrow 0^+} \frac{1-(1-2pqs)}{2ps} = q = \mathbb{P}(Z=0) = \mathbb{P}(Y_1=0) = H(0).$$

Alternatively, L'Hospital's rule can be used to compute the limit of $H(s)$ expressed as a ratio.

f) We have

$$H'(s) = \frac{pq}{ps\sqrt{1-4pqs}} - \frac{1-\sqrt{1-4pqs}}{2ps^2},$$

and

$$\begin{aligned} H'(1) &= \frac{pq}{p\sqrt{1-4pq}} - \frac{1-\sqrt{1-4pq}}{2p} = \frac{pq}{p(q-p)} - \frac{1-(q-p)}{2p} \\ &= \frac{q}{q-p} - 1 = \frac{p}{q-p} = \frac{\mu}{1-\mu}, \end{aligned}$$

with $\mu = p/q$ for $p < 1/2$, which shows that

$$\mathbb{E}[Z] = \frac{\mu}{1-\mu}$$

and recovers the result of Question (b).

g) We have

$$\begin{aligned} \mathbb{E}\left[\sum_{k=1}^{Z} U_k\right] &= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{Z} U_k \;\Big|\; Z=n\right] \mathbb{P}(Z=n) \\ &= \sum_{n\geqslant 0} \mathbb{E}\left[\sum_{k=1}^{n} U_k\right] \mathbb{P}(Z=n) \\ &= \sum_{n\geqslant 0} n\mathbb{E}[U_1]\mathbb{P}(Z=n) \\ &= \mathbb{E}[U_1]\mathbb{E}[Z] = \mathbb{E}[U_1]\frac{\mu}{1-\mu}. \end{aligned}$$

h) We have

$$\mathbb{P}(U_k < x,\ k=1,2,\ldots,Z) = \sum_{n\geqslant 0} \mathbb{P}(U_k < x,\ k=1,2,\ldots,Z \mid Z=n)\mathbb{P}(Z=n)$$

$$
\begin{aligned}
&= \sum_{n\geqslant 0} \mathbb{P}(U_k < x,\ k = 1,2,\ldots,n)\mathbb{P}(Z = n) \\
&= \sum_{n\geqslant 0} \mathbb{P}(U_1 < x) \times \cdots \times \mathbb{P}(U_n < x)\mathbb{P}(Z = n) \\
&= \sum_{n\geqslant 0} (\mathbb{P}(U_1 < x))^n \mathbb{P}(Z = n) = \sum_{n\geqslant 0} (F(x))^n \mathbb{P}(Z = n) = H(F(x)),
\end{aligned}
$$

under the convention that the condition is satisfied by default when $Z = 0$.

Remark. We could also compute the same probability given that $Z \geqslant 1$, and this would give

$$
\begin{aligned}
&\mathbb{P}(U_k < x,\ k = 1,2,\ldots,Z \mid Z \geqslant 1) \\
&= \frac{1}{\mathbb{P}(Z \geqslant 1)} \sum_{n\geqslant 1} \mathbb{P}(U_k < x,\ k = 1,2,\ldots,Z \mid Z = n)\mathbb{P}(Z = n) \\
&= \frac{1}{\mathbb{P}(Z \geqslant 1)} \sum_{n\geqslant 1} \mathbb{P}(U_k < x,\ k = 1,2,\ldots,n)\mathbb{P}(Z = n) \\
&= \frac{1}{\mathbb{P}(Z \geqslant 1)} \sum_{n\geqslant 1} \mathbb{P}(U_1 < x)^n \mathbb{P}(Z = n) = \sum_{n\geqslant 1} F(x)^n \frac{\mathbb{P}(Z = n)}{\mathbb{P}(Z \geqslant 1)} \\
&= \frac{1}{\mathbb{P}(Z \geqslant 1)} (H(F(x)) - \mathbb{P}(Z = 0)) = \frac{1}{p}(H(F(x)) - q).
\end{aligned}
$$

i) We have

$$
\mathbb{E}\left[\sum_{k=1}^{Z} U_k\right] = \mathbb{E}[U_1]\frac{\mu}{1-\mu} = \frac{\mu}{1-\mu} = \frac{p}{q-p}.
$$

We find

$$
\begin{aligned}
\mathbb{P}(U_k < x,\ k = 1,2,\ldots,Z) &= H(F(x)) \\
&= H(1 - e^{-x}) \\
&= \frac{1 - \sqrt{1 - 4pq(1 - e^{-x})}}{2p(1 - e^{-x})}.
\end{aligned}
$$

Chapter 9 - Continuous-Time Markov Chains

Exercise 9.1 We model the number of operating machines as a birth and death process $(X_t)_{t\in\mathbb{R}_+}$ on the state space $\{0, 1, 2, 3, 4, 5\}$.

A new machine can only be added at the rate λ since the repairman can fix only one machine at a time.

In order to determine the failure rate starting from state $k \in \{0,1,2,3,4,5\}$, let us assume that the number of working machines at time t is $X_t = k$. It is known that the lifetime τ_i of machine $i \in \{1,2,\ldots,k\}$ is an exponentially distributed random variable with parameter $\mu > 0$. On the other hand, we know that the first machine to fail will do so at time $\min(\tau_1,\tau_2,\ldots,\tau_k)$, and we have

$$\begin{aligned}\mathbb{P}(\min(\tau_1,\tau_2,\ldots,\tau_k) > t) &= \mathbb{P}(\tau_1 > t, \tau_2 > t, \ldots, \tau_k > t)\\ &= \mathbb{P}(\tau_1 > t)\mathbb{P}(\tau_2 > t)\cdots\mathbb{P}(\tau_k > t) = (\mathrm{e}^{-\mu t})^k = \mathrm{e}^{-k\mu t},\end{aligned}$$

$t \geqslant 0$, hence the time until the first machine failure is exponentially distributed with parameter $k\mu$, *i.e.* the birth rate μ_k of $(X_t)_{t\in\mathbb{R}_+}$ is $\mu_k = k\mu$, $k = 1,2,3,4,5$.

Consequently, the infinitesimal generator Q of $(X_t)_{t\in\mathbb{R}_+}$ is given by

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 \\ \mu & -\mu-\lambda & \lambda & 0 & 0 & 0 \\ 0 & 2\mu & -2\mu-\lambda & \lambda & 0 & 0 \\ 0 & 0 & 3\mu & -3\mu-\lambda & \lambda & 0 \\ 0 & 0 & 0 & 4\mu & -4\mu-\lambda & \lambda \\ 0 & 0 & 0 & 0 & 5\mu & -5\mu \end{bmatrix},$$

with $\lambda = 0.5$ and $\mu = 0.2$. We look for a stationary distribution of the form

$$\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$$

by solving $\pi Q = 0$, *i.e.*

$$\begin{cases} 0 = -\lambda\pi_0 + \mu\pi_1 \\ 0 = \lambda\pi_0 - (\lambda+\mu)\pi_1 + 2\mu\pi_2 \\ 0 = \lambda\pi_1 - (\lambda+2\mu)\pi_2 + 3\mu\pi_3 \\ 0 = \lambda\pi_2 - (\lambda+3\mu)\pi_3 + 4\mu\pi_4 \\ 0 = \lambda\pi_3 - (\lambda+4\mu)\pi_4 + 5\mu\pi_5 \\ 0 = \lambda\pi_4 - 5\mu\pi_5, \end{cases} \quad \text{which gives} \quad \begin{cases} 0 = -\lambda\pi_0 + \mu\pi_1 \\ 0 = -\lambda\pi_1 + 2\mu\pi_2 \\ 0 = -\lambda\pi_2 + 3\mu\pi_3 \\ 0 = -\lambda\pi_3 + 4\mu\pi_4 \\ 0 = -\lambda\pi_4 + 5\mu\pi_5, \end{cases}$$

hence

$$\pi_1 = \frac{\lambda}{\mu}\pi_0, \quad \pi_2 = \frac{\lambda}{2\mu}\pi_1, \quad \pi_3 = \frac{\lambda}{3\mu}\pi_2, \quad \pi_4 = \frac{\lambda}{4\mu}\pi_3, \quad \pi_5 = \frac{\lambda}{5\mu}\pi_4,$$

i.e.

$$\pi_1 = \frac{\lambda}{\mu}\pi_0, \quad \pi_2 = \frac{\lambda^2}{2\mu^2}\pi_0, \quad \pi_3 = \frac{\lambda^3}{3!\mu^3}\pi_0, \quad \pi_4 = \frac{\lambda^4}{4!\mu^4}\pi_0, \quad \pi_5 = \frac{\lambda^5}{5!\mu^5}\pi_0,$$

which is a truncated Poisson distribution with

$$\pi_0 + \frac{\lambda}{\mu}\pi_0 + \frac{\lambda^2}{2\mu^2}\pi_0 + \frac{\lambda^3}{3!\mu^3}\pi_0 + \frac{\lambda^4}{4!\mu^4}\pi_0 + \frac{\lambda^5}{5!\mu^5}\pi_0 = 1,$$

hence

$$\begin{aligned}\pi_0 &= \frac{1}{1 + \dfrac{\lambda}{\mu} + \dfrac{\lambda^2}{2\mu^2} + \dfrac{\lambda^3}{3!\mu^3} + \dfrac{\lambda^4}{4!\mu^4} + \dfrac{\lambda^5}{5!\mu^5}} \\ &= \frac{\mu^5}{\mu^5 + \lambda\mu^4 + \lambda^2\mu^3/2 + \lambda^3\mu^2/3! + \lambda^4\mu/4! + \lambda^5/5!}.\end{aligned}$$

Finally, since π_5 is the probability that all 5 machines are operating, the fraction of time the repairman is idle in the long run is

$$\pi_5 = \frac{\lambda^5}{120\mu^5 + 120\lambda\mu^4 + 60\lambda^2\mu^3 + 20\lambda^3\mu^2 + 5\lambda^4\mu + \lambda^5}.$$

Note that of at most two machines can be under repair, the infinitesimal generator Q of $(X_t)_{t\in\mathbb{R}_+}$ will become

$$Q = \begin{bmatrix} -2\lambda & 2\lambda & 0 & 0 & 0 & 0 \\ \mu & -\mu-2\lambda & 2\lambda & 0 & 0 & 0 \\ 0 & 2\mu & -2\mu-2\lambda & 2\lambda & 0 & 0 \\ 0 & 0 & 3\mu & -3\mu-2\lambda & 2\lambda & 0 \\ 0 & 0 & 0 & 4\mu & -4\mu-\lambda & \lambda \\ 0 & 0 & 0 & 0 & 5\mu & -5\mu \end{bmatrix}.$$

Exercise 9.2

a) Since the time τ_k^R spent between two Poisson arrivals n^o k and $k+1$ is an exponentially distributed random variable with parameter λ_R, the probability we are looking for is given by

$$\mathbb{P}(\tau_k^R > t) = \mathbb{P}(N_t^R = 0) = \mathrm{e}^{-\lambda_R t},$$

where N_t^R denotes a Poisson process with intensity λ_R.

b) This probability is given by

$$\mathbb{P}(N_t^W \leqslant 3) = \mathbb{P}(N_t^W = 0) + \mathbb{P}(N_t^W = 1) + \mathbb{P}(N_t^W = 2) + \mathbb{P}(N_t^W = 3)$$
$$= \mathrm{e}^{-\lambda_W t} + \lambda_W t \mathrm{e}^{-\lambda_W t} + \mathrm{e}^{-\lambda_W t}\frac{\lambda_W^2 t^2}{2} + \mathrm{e}^{-\lambda_W t}\frac{\lambda_W^3 t^3}{6},$$

where N_t^W denotes a Poisson process with intensity λ_W.

c) This probability is given by the ratio $\mathbb{P}(\tau^R < \tau^W) = \lambda_R/(\lambda_W + \lambda_R)$ of arrival rates, as follows from the probability computation (1.5.9), where τ^R and τ^W are independent exponential random variables with parameters λ_R and λ_W, representing the time until the next "read", resp. "write" consultation.

   Note that the difference $N_t^R - N_t^W$ between the number N_t^R of "read" consultations and the number N_t^W of "write" consultations is a birth and death process with state-independent birth and death rates λ_R and λ_W.

d) This distribution is given by $\mathbb{P}(N_t^R = k \mid N_t^R + N_t^W = n)$ where N_t^R, N_t^W are independent Poisson random variables with parameters $\lambda_R t$ and $\lambda_W t$ respectively. We have

$$\mathbb{P}(N_t^R = k \mid N_t^R + N_t^W = n) = \frac{\mathbb{P}(N_t^R = k \text{ and } N_t^R + N_t^W = n)}{\mathbb{P}(N_t^R + N_t^W = n)}$$
$$= \frac{\mathbb{P}(N_t^R = k \text{ and } N_t^W = n-k)}{\mathbb{P}(N_t^R + N_t^W = n)} = \frac{\mathbb{P}(N_t^R = k)\mathbb{P}(N_t^W = n-k)}{\mathbb{P}(N_t^R + N_t^W = n)}$$
$$= \mathrm{e}^{-\lambda_R t}\frac{(\lambda_R t)^k}{k!}\mathrm{e}^{-\lambda_W t}\frac{(\lambda_W t)^{n-k}}{(n-k)!}\left(\mathrm{e}^{-\lambda_W t - \lambda_R t}\frac{(\lambda_W + \lambda_R)^n t^n}{n!}\right)^{-1}$$
$$= \binom{n}{k}\left(\frac{\lambda_R}{\lambda_R + \lambda_W}\right)^k\left(\frac{\lambda_W}{\lambda_R + \lambda_W}\right)^{n-k}, \quad k = 0, 1, \ldots, n,$$

   cf. (S.10) in the solution of Exercise 1.7.

Exercise 9.3

a) The number X_t of machines operating at time t is a birth and death process on $\{0, 1, 2\}$ with infinitesimal generator

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda + \mu) & \lambda \\ 0 & 2\mu & -2\mu \end{bmatrix}.$$

   The stationary distribution $\pi = (\pi_0, \pi_1, \pi_2)$ is solution of $\pi Q = 0$, *i.e.*

$$\begin{cases} 0 = -\lambda\pi_0 + \mu\pi_1 \\ 0 = \lambda\pi_0 - (\lambda + \mu)\pi_1 + 2\mu\pi_2 \\ 0 = \lambda\pi_1 - 2\mu\pi_2 \end{cases}$$

   under the condition $\pi_0 + \pi_1 + \pi_2 = 1$, which yields

$$(\pi_0, \pi_1, \pi_2) = \left(\frac{2\mu^2}{2\mu^2 + 2\lambda\mu + \lambda^2}, \frac{2\lambda\mu}{2\mu^2 + 2\lambda\mu + \lambda^2}, \frac{\lambda^2}{2\mu^2 + 2\lambda\mu + \lambda^2} \right),$$

i.e. the probability that no machine is operating is $\pi_0 = 2/5$ when $\lambda = \mu = 1$.

b) The number X_t of machines operating at time t is now a birth and death process on $\{0, 1\}$. The time spent in state ⓪ is exponentially distributed with average $1/\lambda$. When the chain is in state ①, one machine is working while the other one may still be under repair, and the mean time $\mathbb{E}[T_0 \mid X_0 = 1]$ spent in state ① before switching to state ⓪ has to be computed using first step analysis on the discrete-time embedded chain. We have

$$\begin{aligned}
&\mathbb{E}[T_0 \mid X_0 = 1] \\
&\quad = \mathbb{P}(\tau_\lambda < \tau_\mu) \times \left(\frac{1}{\mu} + \mathbb{E}[T_0 \mid X_0 = 1] \right) + \mathbb{P}(\tau_\mu < \tau_\lambda) \times \frac{1}{\mu} \\
&\quad = \frac{1}{\mu} + \mathbb{P}(\tau_\lambda < \tau_\mu) \times \mathbb{E}[T_0 \mid X_0 = 1] \\
&\quad = \frac{1}{\mu} + \frac{\lambda}{\lambda + \mu} \mathbb{E}[T_0 \mid X_0 = 1],
\end{aligned}$$

where, by (9.6.2) and (9.6.3) or (1.5.9), $\mathbb{P}(\tau_\lambda < \tau_\mu) = \lambda/(\lambda + \mu)$ is the probability that an exponential random variable τ_λ with parameter $\lambda > 0$ is smaller than another independent exponential random variable τ_μ with parameter $\mu > 0$. In other words, $\mathbb{P}(\tau_\lambda < \tau_\mu)$ is the probability that the repair of the idle machine finishes before the working machine fails. This first step analysis argument is similar to the one used in the solution of Exercise 9.18. This yields

$$\mathbb{E}[T_0 \mid X_0 = 1] = \frac{\lambda + \mu}{\mu^2},$$

hence the corresponding rate is $\mu^2/(\lambda + \mu)$ and the infinitesimal generator of the chain becomes

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \frac{1}{\mathbb{E}[T_0 \mid X_0 = 1]} & -\frac{1}{\mathbb{E}[T_0 \mid X_0 = 1]} \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda \\ \frac{\mu^2}{\lambda + \mu} & -\frac{\mu^2}{\lambda + \mu} \end{bmatrix}.$$

The stationary distribution $\pi = (\pi_0, \pi_1)$ is solution of $\pi Q = 0$, *i.e.*

$$\begin{cases} 0 = -\lambda \pi_0 + \pi_1 \dfrac{\mu^2}{\lambda + \mu} \\ 0 = \lambda \pi_0 - \pi_1 \dfrac{\mu^2}{\lambda + \mu} \end{cases}$$

under the condition $\pi_0 + \pi_1 = 1$, which yields

$$(\pi_0, \pi_1) = \left(\frac{\mu^2}{\mu^2 + \lambda\mu + \lambda^2}, \frac{\lambda\mu + \lambda^2}{\mu^2 + \lambda\mu + \lambda^2} \right),$$

i.e. the probability that no machine is operating when $\lambda = \mu = 1$ is $\pi_0 = 1/3$.

Alternatively, this question can be solved by considering a birth and death process $(X_t)_{t\in\mathbb{R}_+}$ on $\{0, 1, 2\}$ with infinitesimal generator

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{bmatrix},$$

for which

$$\begin{array}{lcl} X_t = 0 & \Longleftrightarrow & \text{no machine is working,} \\ X_t = 1 & \Longleftrightarrow & \text{one machine is working and the other is under repair,} \\ X_t = 2 & \Longleftrightarrow & \text{one machine is working and the other one is waiting.} \end{array}$$

In this case, $\pi Q = 0$ yields

$$(\pi_0, \pi_1, \pi_2) = \left(\frac{\mu^2}{\mu^2 + \lambda\mu + \lambda^2}, \frac{\lambda\mu}{\mu^2 + \lambda\mu + \lambda^2}, \frac{\lambda^2}{\mu^2 + \lambda\mu + \lambda^2} \right),$$

hence

$$\pi_0 = \frac{1}{3}, \qquad \pi_1 = \frac{1}{3}, \qquad \pi_2 = \frac{1}{3},$$

when $\lambda = \mu$.

Exercise 9.4

a) The process $(X_t)_{t\in\mathbb{R}_+}$ is a continuous-time Markov chain due to the Poisson arrival of customers in the queue. The state space of the chain $(X_t)_{t\in\mathbb{R}_+}$ is $\mathbb{S} = \{0, 1, 2, 3\}$ since there cannot be more than 3 people in the queue as the cable car departs immediately as soon as there are more 4 people in the queue. The infinitesimal generator of the chain is given by

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & -\lambda & \lambda \\ \lambda & 0 & 0 & -\lambda \end{bmatrix}.$$

b) As the chain is irreducible, the limiting distribution $\pi = [\pi_0, \pi_1, \pi_2, \pi_3]$ is obtained from the stationary distribution, by solving the equation $\pi Q = 0$, which yields $\pi = [\pi_0, \pi_1, \pi_2, \pi_3] = [1/4, 1/4, 1/4, 1/4,]$

c) This mean time is the sum

$$\mathbb{E}[\tau_0] + \mathbb{E}[\tau_1] + \mathbb{E}[\tau_2] + \mathbb{E}[\tau_3] = \frac{4}{\lambda}$$

of the means of four exponential random variables with parameter $\lambda > 0$, hence it equals $4/\lambda$.

Exercise 9.5 Muni Toke (2015)

a) The process $(L_t)_{t\in\mathbb{R}_+}$ is Markov from its construction using Poisson processes and the memoryless property of exponential random times. The matrix of the infinitesimal generator is

$$\begin{bmatrix} -5\lambda & 5\lambda & 0 & 0 & 0 & 0 & \cdots \\ \mu+\theta & -\mu-\theta-5\lambda & 5\lambda & 0 & 0 & 0 & \cdots \\ 0 & 2\mu+\theta & -2\mu-\theta-5\lambda & 5\lambda & 0 & 0 & \cdots \\ 0 & 0 & 3\mu+\theta & -3\mu-\theta-5\lambda & 5\lambda & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

b) Given that the mean number of issued order until time t is λt and the mean number of cancelled order until time t is μt, we deduce the relation $\mu t = 0.95\lambda t$, *i.e.* $\mu = 0.95\lambda$.

Exercise 9.6 The size of the crack is viewed as a continuous-time birth process taking values in $\{1, 2, 3, \ldots\}$ with state-dependent rate $\lambda_k = (1+k)^\rho$, $k \geqslant 1$. Let us denote by τ_k the time spent at state $k \in \mathbb{N}$ between two increases, which is an exponentially distributed random variable with parameter λ_k. The time it takes for the crack length to grow to infinity is $\sum_{k\geqslant 1} \tau_k$. It is known that $\sum_{k\geqslant 1} \tau_k < \infty$ almost surely if the expectation $\mathbb{E}\left[\sum_{k\geqslant 1} \tau_k\right]$ is finite, and in this situation the crack grows to infinity within a finite time. We have

$$\mathbb{E}\left[\sum_{k\geqslant 1} \tau_k\right] = \sum_{k\geqslant 1} \mathbb{E}[\tau_k] = \sum_{k\geqslant 1} \frac{1}{\lambda_k} = \sum_{k\geqslant 1} \frac{1}{(1+k)^\rho}.$$

By comparison with the integral of the function $x \longmapsto 1/(1+x)^\rho$ we get

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geqslant 1} \tau_k\right] &= \sum_{k\geqslant 1} \frac{1}{(1+k)^\rho} \\ &\leqslant \sum_{k\geqslant 1} \int_{k-1}^{k} \frac{1}{(1+x)^\rho} dx\end{aligned}$$

$$\begin{aligned}
&\leqslant \int_0^\infty \frac{1}{(1+x)^\rho} dx \\
&= \frac{1}{1-\rho} \left[(1+x)^{1-\rho} \right]_0^\infty \\
&= \frac{1}{\rho - 1} < \infty,
\end{aligned}$$

provided that $\rho > 1$. We conclude that the time for the crack to grow to infinite length is (almost surely) finite when $\rho > 1$. Similarly, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{k\geqslant 1} \tau_k \right] &= \sum_{k\geqslant 1} \frac{1}{(1+k)^\rho} \\
&\geqslant \sum_{k\geqslant 1} \int_k^{k+1} \frac{1}{(1+x)^\rho} dx \\
&\geqslant \int_1^\infty \frac{1}{(1+x)^\rho} dx \\
&= \infty,
\end{aligned}$$

hence the mean time for the crack to grow to infinite length is infinite when $\rho \leqslant 1$.

Remark. This problem could also be treated in discrete time, assuming that τ_k is the (random) crack length increase at each step. In this case, the relation

$$\mathbb{E}\left[\sum_{k\geqslant 1} \tau_k \right] < \infty$$

for $\rho > 1$ shows that growth of the crack length to infinity cannot occur in this case.

Exercise 9.7

a) This time is the expected value of the third jump time T_3, *i.e.*

$$\mathbb{E}[T_3] = \mathbb{E}[\tau_0] + \mathbb{E}[\tau_1] + \mathbb{E}[\tau_2] = \frac{3}{\lambda} = 30 \text{ minutes.}$$

b) This probability is

$$\begin{aligned}
\mathbb{P}(N_{60} < 3) &= \mathbb{P}(N_{60} = 0) + \mathbb{P}(N_{60} = 1) + \mathbb{P}(N_{60} = 2) \\
&= \mathrm{e}^{-60\lambda}(1 + 60\lambda + (60\lambda)^2/2) \\
&= 25\,\mathrm{e}^{-6} \simeq 0.062.
\end{aligned}$$

Exercise 9.8

a) The number of users checking the website at time t consists in the number $N_t - N_{t-\min(\theta,t)}$ of users who have connected between time $t-\min(\theta,t)$ and time t, hence its average is given by

$$\mathbb{E}[N_t - N_{t-\min(\theta,t)}] = \lambda(t-(t-\min(\theta,t))) = \lambda\min(\theta,t).$$

b) The event that no user is checking the website at time t can be written as $\{N_t - N_{\mathrm{Max}(\theta,t)} = 0\}$, and its probability is given by

$$\mathbb{P}(N_t - N_{\mathrm{Max}(\theta,t)} = 0) = \mathrm{e}^{-\lambda(t-(t-\min(\theta,t)))} = \mathrm{e}^{-\lambda\min(\theta,t)}.$$

Exercise 9.9

a) By the independence of increments of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$ we find

$$\begin{aligned}\mathbb{P}(N_{t_3}=5 \mid N_{t_1}=1) &= \frac{\mathbb{P}(N_{t_3}=5 \text{ and } N_{t_1}=1)}{\mathbb{P}(N_{t_1}=1)}\\ &= \frac{\mathbb{P}(N_{t_3}-N_{t_1}=4 \text{ and } N_{t_1}=1)}{\mathbb{P}(N_{t_1}=1)} = \frac{\mathbb{P}(N_{t_3}-N_{t_1}=4)\mathbb{P}(N_{t_1}=1)}{\mathbb{P}(N_{t_1}=1)}\\ &= \mathbb{P}(N_{t_3}-N_{t_1}=4) = \frac{(\lambda(t_3-t_1))^4}{4!}\mathrm{e}^{-\lambda(t_3-t_1)}.\end{aligned}$$

b) We expand N_{t_4} into the telescoping sum

$$N_{t_4} = (N_{t_4}-N_{t_3}) + (N_{t_3}-N_{t_2}) + (N_{t_2}-N_{t_1}) + (N_{t_1}-N_0)$$

of independent increments on disjoint time intervals, to obtain

$$\begin{aligned}&\mathbb{E}[N_{t_1}N_{t_4}(N_{t_3}-N_{t_2})]\\ &= \mathbb{E}[N_{t_1}(N_{t_4}-N_{t_3})(N_{t_3}-N_{t_2})] + \mathbb{E}[N_{t_1}(N_{t_3}-N_{t_2})(N_{t_3}-N_{t_2})]\\ &\quad +\mathbb{E}[N_{t_1}(N_{t_2}-N_{t_1})(N_{t_3}-N_{t_2})] + \mathbb{E}[N_{t_1}N_{t_1}(N_{t_3}-N_{t_2})]\\ &= \mathbb{E}[N_{t_1}]\mathbb{E}[N_{t_4}-N_{t_3}]\mathbb{E}[N_{t_3}-N_{t_2}] + \mathbb{E}[N_{t_1}]\mathbb{E}[(N_{t_3}-N_{t_2})^2]\\ &\quad +\mathbb{E}[N_{t_1}]\mathbb{E}[N_{t_2}-N_{t_1}]\mathbb{E}[N_{t_3}-N_{t_2}] + \mathbb{E}[N_{t_1}^2]\mathbb{E}[N_{t_3}-N_{t_2}]\\ &= \lambda^3 t_1(t_4-t_3)(t_3-t_2) + \lambda^2 t_1(t_3-t_2)(1+\lambda(t_3-t_2))\\ &\quad +\lambda^3 t_1(t_2-t_1)(t_3-t_2) + \lambda^2 t_1(1+\lambda t_1)(t_3-t_2).\end{aligned}$$

c) We have $\{T_2 > t\} = \{N_t \leqslant 1\}$, $t \geqslant 0$, hence

$$\mathbb{E}[N_{t_2} \mid T_2 > t_1] = \mathbb{E}[N_{t_2} \mid N_{t_1} \leqslant 1] = \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[N_{t_2}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}],$$

by (1.6.6). Now, using the independence of increments between $N_{t_2}-N_{t_1}$ and N_{t_1}, we have

$$\begin{aligned}\mathbb{E}[N_{t_2}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] &= \mathbb{E}[(N_{t_2}-N_{t_1})\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] + \mathbb{E}[N_{t_1}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \mathbb{E}[N_{t_2}-N_{t_1}]\mathbb{E}[\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] + \mathbb{E}[N_{t_1}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \mathbb{E}[N_{t_2}-N_{t_1}]\mathbb{P}(N_{t_1}\leqslant 1) + \mathbb{E}[N_{t_1}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}],\end{aligned}$$

hence

$$\begin{aligned}&\mathbb{E}[N_{t_2} \mid T_2 > t_1] = \mathbb{E}[N_{t_2} \mid N_{t_1}\leqslant 1] = \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[N_{t_2}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[(N_{t_2}-N_{t_1}+N_{t_1})\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[(N_{t_2}-N_{t_1})\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] + \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[N_{t_1}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \mathbb{E}[N_{t_2}-N_{t_1}] + \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\mathbb{E}[N_{t_1}\mathbb{1}_{\{N_{t_1}\leqslant 1\}}] \\ &= \mathbb{E}[N_{t_2}-N_{t_1}] + \frac{1}{\mathbb{P}(N_{t_1}\leqslant 1)}\left(0\times\mathbb{P}(N_{t_1}=0) + 1\times\mathbb{P}(N_{t_1}=1)\right) \\ &= \mathbb{E}[N_{t_2}-N_{t_1}] + \frac{\mathbb{P}(N_{t_1}=1)}{\mathbb{P}(N_{t_1}\leqslant 1)} = \lambda(t_2-t_1) + \frac{\lambda t_1 \mathrm{e}^{-\lambda t_1}}{\mathrm{e}^{-\lambda t_1}+\lambda t_1\mathrm{e}^{-\lambda t_1}}.\end{aligned}$$

Exercise 9.10 The generator of the process is given by

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 \\ 0 & \mu_2 & -\mu_2 \end{bmatrix} = \begin{bmatrix} -2\alpha & 2\alpha & 0 \\ \beta & -\alpha-\beta & \alpha \\ 0 & 2\beta & -2\beta \end{bmatrix}.$$

Writing the equation $\pi Q = 0$ shows that

$$\pi_1 = \frac{2\alpha}{\beta}\pi_0 \quad \text{and} \quad \pi_2 = \frac{\alpha}{2\beta}\pi_1,$$

and the condition

$$\pi_0 + \pi_1 + \pi_2 = 1$$

shows that

$$\pi_0 = \left(1+\frac{\alpha}{\beta}\right)^{-2}$$

and

$$\pi_0 = \left(\frac{\beta}{\alpha+\beta}\right)^2, \quad \pi_1 = \frac{2\alpha\beta}{(\alpha+\beta)^2} \quad \text{and} \quad \pi_2 = \left(\frac{\alpha}{\alpha+\beta}\right)^2,$$

which is a binomial distribution with parameter $(2,p) = (2, \alpha/(\alpha+\beta))$, *i.e.*

$$\pi_0 = p^2, \quad \pi_1 = \binom{2}{1}p(1-p) \quad \text{and} \quad \pi_2 = (1-p)^2.$$

Exercise 9.11 This is an extension of Exercise 9.10. The generator of the process is given by

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots & 0 & 0 & 0 \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mu_{N-1} & -\lambda_{N-1} - \mu_{N-1} & \lambda_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & \mu_N & -\mu_N \end{bmatrix}$$

$$= \begin{bmatrix} -\alpha N & \alpha N & 0 & \cdots & 0 & 0 & 0 \\ \beta & -\alpha(N-1) - \beta & \alpha(N-1) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta(N-1) & -\alpha - \beta(N-1) & \alpha \\ 0 & 0 & 0 & \cdots & 0 & \beta N & -\beta N \end{bmatrix}.$$

Writing the equation $\pi Q = 0$ shows that we have $-\alpha N \pi_0 + \beta \pi_1 = 0$,

$$\beta(k+1)\pi_{k+1} - (\alpha(N-k) + \beta k)\pi_k + \alpha(N-(k-1))\pi_{k-1} = 0, \quad k = 1, 2, \ldots, N-1,$$

and $\alpha \pi_{N-1} - \beta N \pi_N = 0$, from which we deduce the recurrence relation

$$\pi_{k+1} = \frac{\alpha}{\beta} \frac{N-k}{k+1} \pi_k, \qquad k = 0, 1, \ldots, N-1,$$

and by induction on $k = 1, 2, \ldots, N$ we find $\pi_1 = N\pi_0 \alpha/\beta$ and

$$\pi_2 = \frac{\alpha}{\beta} \frac{(N-1)}{2} \pi_1 = \left(\frac{\alpha}{\beta}\right)^2 \frac{N(N-1)}{1 \times 2} \pi_0, \quad \pi_3 = \left(\frac{\alpha}{\beta}\right)^3 \frac{N(N-1)(N-2)}{1 \times 2 \times 3} \pi_0,$$

hence

$$\begin{aligned} \pi_k &= \left(\frac{\alpha}{\beta}\right)^k \frac{N(N-1)\cdots(N-k+1)}{k!} \pi_0 \\ &= \left(\frac{\alpha}{\beta}\right)^k \frac{N!}{(N-k)!k!} \pi_0 = \left(\frac{\alpha}{\beta}\right)^k \binom{N}{k} \pi_0, \end{aligned}$$

$k = 0, 1, \ldots, N$. The condition

$$\pi_0 + \pi_1 + \cdots + \pi_N = 1$$

shows that

$$1 = \pi_0 \sum_{k=0}^{N} \left(\frac{\alpha}{\beta}\right)^k \frac{N!}{(N-k)!k!} = \left(1 + \frac{\alpha}{\beta}\right)^N \pi_0,$$

hence

$$\pi_0 = \left(1 + \frac{\alpha}{\beta}\right)^{-N}$$

and we have

$$\begin{aligned}\pi_k &= \left(1 + \frac{\alpha}{\beta}\right)^{-N} \left(\frac{\alpha}{\beta}\right)^k \frac{N!}{(N-k)!k!} \\ &= \left(\frac{\alpha}{\alpha+\beta}\right)^k \left(\frac{\beta}{\alpha+\beta}\right)^{N-k} \frac{N!}{(N-k)!k!}, \qquad k = 0, 1, \ldots, N,\end{aligned}$$

hence the stationary distribution π is a binomial distribution with parameter $(N, p) = (N, \alpha/(\alpha+\beta))$.

Exercise 9.12 The generator Q of this pure birth process is given by

$$Q = [\ \lambda_{i,j}\ ]_{i,j\in\mathbb{N}} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & -3 & 3 & 0 & 0 & \cdots \\ 0 & 0 & -2 & 2 & 0 & \cdots \\ 0 & 0 & 0 & -5 & 5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix},$$

hence the forward Kolmogorov equation $P'(t) = P(t)Q$ reads

$$\begin{bmatrix} P'_{0,0}(t) & P'_{0,1}(t) & P'_{0,2}(t) & P'_{0,3}(t) & \cdots \\ P'_{1,0}(t) & P'_{1,1}(t) & P'_{1,2}(t) & P'_{1,3}(t) & \cdots \\ P'_{2,0}(t) & P'_{2,1}(t) & P'_{2,2}(t) & P'_{2,3}(t) & \cdots \\ P'_{3,0}(t) & P'_{3,1}(t) & P'_{3,2}(t) & P'_{3,3}(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & P_{0,3}(t) & \cdots \\ P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & P_{1,3}(t) & \cdots \\ P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & P_{2,3}(t) & \cdots \\ P_{3,0}(t) & P_{3,1}(t) & P_{3,2}(t) & P_{3,3}(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \times \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & -3 & 3 & 0 & 0 & \cdots \\ 0 & 0 & -2 & 2 & 0 & \cdots \\ 0 & 0 & 0 & -5 & 5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix},$$

which yields

$$\begin{cases} P'_{0,0}(t) = -P_{0,0}(t), \\ \\ P'_{0,1}(t) = P_{0,0}(t) - 3P_{0,1}(t), \\ \\ P'_{0,2}(t) = 3P_{0,1}(t) - 2P_{0,2}(t), \\ \\ P'_{0,3}(t) = 2P_{0,2}(t) - 5P_{0,3}(t). \end{cases}$$

The first equation is solved by (A.9) as

$$P_{0,0}(t) = P_{0,0}(0)\mathrm{e}^{-t} = \mathrm{e}^{-t}, \qquad t \geqslant 0,$$

and this solution can be easily recovered from

$$P_{0,0}(t) = \mathbb{P}(X_t = 0 \mid X_0 = 0) = \mathbb{P}(\tau_0 > t) = \mathrm{e}^{-t}, \qquad t \geqslant 0.$$

The second equation becomes

$$P'_{0,1}(t) = \mathrm{e}^{-t} - 3P_{0,1}(t), \tag{S.56}$$

its associated homogeneous equation is $P'_{0,1}(t) = -3P_{0,1}(t)$, which has $t \longmapsto C_1\mathrm{e}^{-3t}$ for homogeneous solution. By substituting the particular solution $t \longmapsto C_2\mathrm{e}^{-t}$ in (S.56) we get

$$-C_2\mathrm{e}^{-t} = \mathrm{e}^{-t} - 3C_2\mathrm{e}^{-t},$$

i.e. $C_2 = 1/2$, hence the general solution

$$P_{0,1}(t) = \frac{1}{2}\mathrm{e}^{-t} + C_1\mathrm{e}^{-3t},$$

and from the initial condition $P_{0,1}(0) = 0$ we get $C_1 = -1/2$, *i.e.*

$$P_{0,1}(t) = \frac{1}{2}(\mathrm{e}^{-t} - \mathrm{e}^{-3t}), \qquad t \geqslant 0.$$

The remaining equations can be solved similarly by searching for a suitable particular solution. For

$$P'_{0,2}(t) = \frac{3}{2}(\mathrm{e}^{-t} - \mathrm{e}^{-3t}) - 2P_{0,2}(t),$$

we find, searching for a particular solution of the form $t \longmapsto a\mathrm{e}^{-t} + b\mathrm{e}^{-3t}$,

$$P_{0,2}(t) = \frac{3}{2}\mathrm{e}^{-3t}(1 - \mathrm{e}^{t})^2, \qquad t \geqslant 0, \tag{S.57}$$

see *here*, and for

$$P'_{0,3}(t) = 3\mathrm{e}^{-3t}(1 - \mathrm{e}^{t})^2 - 5P_{0,3}(t),$$

we find

$$P_{0,3}(t) = \frac{1}{4}\mathrm{e}^{-5t}(\mathrm{e}^{t} - 1)^3(1 + 3e^{t}), \qquad t \geqslant 0, \tag{S.58}$$

see *here*. Note that using the *backward* Kolmogorov equation $P'(t) = QP(t)$ can lead to more complicated calculations.

Probabilistic approach

The above results may be recovered by a probabilistic approach, with the change of variable $z = x + y$, as

$$\begin{aligned}
P_{0,1}(t) &= \mathbb{P}(X_t = 1 \mid X_0 = 0) \\
&= \mathbb{P}(\tau_0 < t \text{ and } \tau_0 + \tau_1 > t) \\
&= \mathbb{P}(\tau_0 < t < \tau_0 + \tau_1) \\
&= \lambda_0 \lambda_1 \int_{\{(x,y)\ :\ 0<x<t,\ x+y>t\}} \mathrm{e}^{-\lambda_0 x} \mathrm{e}^{-\lambda_1 y} dx dy \\
&= \lambda_0 \lambda_1 \int_0^t \int_t^\infty \mathrm{e}^{-\lambda_0 x} \mathrm{e}^{-\lambda_1 (z-x)} dx dz \\
&= \lambda_0 \lambda_1 \int_0^t \int_t^\infty \mathrm{e}^{-(\lambda_0 - \lambda_1) x} \mathrm{e}^{-\lambda_1 z} dx dz \\
&= \mathrm{e}^{-\lambda_1 t} \lambda_0 \int_0^t \mathrm{e}^{-(\lambda_0 - \lambda_1) x} dx \\
&= \frac{\lambda_0}{\lambda_0 - \lambda_1} (\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\lambda_0 t}) \\
&= \frac{1}{2} (\mathrm{e}^{-t} - \mathrm{e}^{-3t}), \qquad t \geqslant 0. \qquad \text{(S.59)}
\end{aligned}$$

As for $P_{0,2}(t)$ it also reads

$$P_{0,2}(t) = \mathbb{P}(X_t = 2 \mid X_0 = 0) = \mathbb{P}(\tau_0 + \tau_1 < t \text{ and } \tau_0 + \tau_1 + \tau_2 > t),$$

and (S.57) can be recovered via a triple integral. Similarly, (S.58) can be approached by a quadruple integral, etc.

Exercise 9.13 Noting that the two events

$$\{T_1 > t,\ T_2 > t + s\} = \{X_t = 0,\ 0 \leqslant X_{t+s} \leqslant 1\}$$

coincides for all $s, t \geqslant 0$, we find that

$$\begin{aligned}
&\mathbb{P}(T_1 > t \text{ and } T_2 > t + s) = \mathbb{P}(X_t = 0 \text{ and } X_{t+s} \in \{0,1\} \mid X_0 = 0) \\
&\quad = \mathbb{P}(X_{t+s} \in \{0,1\} \mid X_t = 0) \mathbb{P}(X_t = 0 \mid X_0 = 0) \\
&\quad = \mathbb{P}(X_s \in \{0,1\} \mid X_0 = 0) \mathbb{P}(X_t = 0 \mid X_0 = 0) \\
&\quad = (\mathbb{P}(X_s = 0 \mid X_0 = 0) + \mathbb{P}(X_s = 1 \mid X_0 = 0)) \mathbb{P}(X_t = 0 \mid X_0 = 0) \\
&\quad = P_{0,0}(t)(P_{0,0}(s) + P_{0,1}(s)).
\end{aligned}$$

Next, we note that we have

$$P_{0,0}(t) = \mathrm{e}^{-\lambda_0 t}, \qquad t \geqslant 0,$$

and, from Relation (S.59) above,

$$P_{0,1}(t) = \frac{\lambda_0}{\lambda_1 - \lambda_0}\left(\mathrm{e}^{-\lambda_0 t} - \mathrm{e}^{-\lambda_1 t}\right), \qquad t \geqslant 0,$$

hence

$$\begin{aligned}\mathbb{P}(T_1 > t \text{ and } T_2 > t+s) &= \mathrm{e}^{-\lambda_0(t+s)} + \frac{\lambda_0}{\lambda_1 - \lambda_0}\left(\mathrm{e}^{-\lambda_0(t+s)} - \mathrm{e}^{-\lambda_0 t - \lambda_1 s}\right) \\ &= \frac{\lambda_1}{\lambda_1 - \lambda_0}\mathrm{e}^{-\lambda_0(t+s)} - \frac{\lambda_0}{\lambda_1 - \lambda_0}\mathrm{e}^{-\lambda_0 t - \lambda_1 s}.\end{aligned}$$

Then, since

$$\mathbb{P}(T_1 > x \text{ and } T_2 > y) = \int_x^\infty \int_y^\infty f_{(T_1,T_2)}(u,v)dudv,$$

by (1.5.4) we get

$$\begin{aligned}f_{(T_1,T_2)}(x,y) &= \frac{\partial^2}{\partial y \partial x}\mathbb{P}(T_1 > x \text{ and } T_2 > y) \\ &= \frac{\partial^2}{\partial y \partial x}\left(\frac{\lambda_1}{\lambda_1 - \lambda_0}\mathrm{e}^{-\lambda_0 y} - \frac{\lambda_0}{\lambda_1 - \lambda_0}\mathrm{e}^{-\lambda_0 x - \lambda_1(y-x)}\right) \\ &= -\lambda_0 \frac{\partial}{\partial y}\mathrm{e}^{-\lambda_0 x - \lambda_1(y-x)} = \lambda_0\lambda_1 \mathrm{e}^{-\lambda_0 x - \lambda_1(y-x)},\end{aligned}$$

provided that $y \geqslant x \geqslant 0$. When $x > y \geqslant 0$ we have

$$f_{(T_1,T_2)}(x,y) = 0.$$

The density of (τ_0, τ_1) is given under the change of variable $T_1 = \tau_0$, $T_2 = \tau_0 + \tau_1$, by

$$f_{(\tau_0,\tau_1)}(s,t) = f_{(T_1,T_2)}(s, s+t) = \lambda_0\lambda_1 \mathrm{e}^{-\lambda_0 s - \lambda_1 t}, \quad s,t \geqslant 0,$$

which shows that τ_0, τ_1 are two independent exponentially distributed random variables with parameters λ_0 and λ_1, respectively.

Exercise 9.14 Let $(N_t)_{t\in\mathbb{R}_+}$ denote a Poisson process with intensity $\lambda > 0$.

a) This probability is equal to

$$\mathbb{P}(N_T = 0) = \mathbb{P}(\tau_0 > T) = \mathrm{e}^{-\lambda T}.$$

b) Let t denote the expected time we are looking for. When the woman attempts to cross the street, she can do so immediately with probability $\mathbb{P}(N_T = 0) = \mathbb{P}(\tau_0 > T)$, in which case the waiting time is 0. Otherwise, with probability $1 - \mathbb{P}(N_T = 0)$, she has to wait on average (using Lemma 1.10 and (1.6.12))

$$\mathbb{E}[\tau_0 \mid \tau_0 < T] = \frac{1}{\mathbb{P}(\tau_0 < T)} \mathbb{E}[\tau_0 \mathbb{1}_{\{\tau_0 < T\}}]$$
$$= \frac{\lambda}{1 - \mathrm{e}^{-\lambda T}} \int_0^T x \mathrm{e}^{-\lambda x} dx$$
$$= \frac{1 - (1 + \lambda T)\mathrm{e}^{-\lambda T}}{\lambda(1 - \mathrm{e}^{-\lambda T})}$$

for the first car to pass, after which the process is reinitialized and the average waiting time is again t. Hence by first step analysis in continuous time we find the equation

$$t = 0 \times \mathbb{P}(N_T = 0) + (\mathbb{E}[\tau_0 \mid \tau_0 < T] + t) \times \mathbb{P}(\tau_0 \leqslant T)$$
$$= \frac{1 - (1 + \lambda T)\mathrm{e}^{-\lambda T}}{\lambda} + t(1 - \mathrm{e}^{-\lambda T}) \tag{S.60}$$

with unknown t, and solution

$$t = \frac{e^{\lambda T} - 1 - \lambda T}{\lambda}.$$

Alternatively, we could also rewrite the above equation (S.60) as

$$t = \mathbb{E}[0 \times \mathbb{1}_{\{\tau_0 \geqslant T\}} + (\tau_0 + t)\mathbb{1}_{\{\tau_0 < T\}}] = \frac{1 - (1 + \lambda T)\mathrm{e}^{-\lambda T}}{\lambda} + t(1 - \mathrm{e}^{-\lambda T}).$$

c) Denoting by t the mean time until she finishes crossing the street we have, by first step analysis in continuous time,

$$t = \mathbb{E}\left[T \mathbb{1}_{\{T < \tau_0\}} + (\tau_0 + t)\mathbb{1}_{\{T \geqslant \tau_0\}}\right]$$
$$= \lambda T \int_T^\infty \mathrm{e}^{-\lambda s} ds + \lambda \int_0^T (s + t)\mathrm{e}^{-\lambda s} ds$$
$$= T\mathbb{P}(\tau_0 > T) + \lambda \int_0^T s\mathrm{e}^{-\lambda s} ds + t\mathbb{P}(\tau_0 < T)$$
$$= T\mathrm{e}^{-\lambda T} + \frac{1 - \mathrm{e}^{-\lambda T}(1 + \lambda T)}{\lambda} + t(1 - \mathrm{e}^{-\lambda T}),$$

which yields

$$t = \frac{e^{\lambda T} - 1}{\lambda}.$$

Alternatively, we can write

$$t = T\mathbb{P}(\tau_0 > T) + \big(\mathbb{E}\left[\tau_0 \mid \tau_0 > T\right] + t\big)\mathbb{P}(\tau_0 > T)$$
$$= T\mathbb{P}(\tau_0 > T) + \mathbb{E}\left[\tau_0 \mathbb{1}_{\{\tau_0 > T\}}\right] + t\mathbb{P}(\tau_0 > T)$$
$$= T\mathbb{P}(\tau_0 > T) + \lambda \int_0^T s\mathrm{e}^{-\lambda s} ds + t\mathbb{P}(\tau_0 < T),$$

which yields the same result. The same conclusion follows by adding the extra crossing time T to the answer of Question (b).

d) In this case, T becomes an independent exponentially distributed random variable with parameter $\mu > 0$, hence we can write

$$\begin{aligned} t &= \mathbb{E}\left[\frac{e^{\lambda T}-1}{\lambda}\right] \\ &= \mu\int_0^\infty \frac{e^{\lambda u}-1}{\lambda}\mathrm{e}^{-\mu u}du \\ &= \frac{\mu}{\lambda}\int_0^\infty \mathrm{e}^{(\lambda-\mu)u}du - \frac{\mu}{\lambda}\int_0^\infty \mathrm{e}^{-\mu u}du \\ &= \frac{\mu}{\lambda(\mu-\lambda)} - \frac{1}{\lambda} \\ &= \frac{1}{\mu-\lambda} \end{aligned}$$

if $\mu > \lambda$, with $t = +\infty$ if $\mu \leqslant \lambda$.

Exercise 9.15

a) This probability is the probability that an exponential random variable with parameter μ is lower than T, *i.e.* $\mathbb{P}(\tau_0 > T) = \mathrm{e}^{-\mu T}$.

b) Denoting by $t = \mathbb{E}[W]$ the mean time until the machine breaks down we have, by first step analysis in continuous time,

$$\begin{aligned} t &= \mathbb{E}\left[T\mathbb{1}_{\{T<\tau_0\}} + (\tau_0+t)\mathbb{1}_{\{T\geqslant\tau_0\}}\right] \\ &= \mu T\int_T^\infty \mathrm{e}^{-\mu s}ds + \mu\int_0^T (s+t)\mathrm{e}^{-\mu s}ds \\ &= T\mathbb{P}(\tau_0 > T) + \mu\int_0^T s\mathrm{e}^{-\mu s}ds + t\mathbb{P}(\tau_0 < T) \\ &= T\mathrm{e}^{-\mu T} + \frac{1-\mathrm{e}^{-\mu T}(1+\mu T)}{\mu} + t(1-\mathrm{e}^{-\mu T}), \end{aligned}$$

which yields

$$t = \mathbb{E}[W] = \frac{e^{\mu T}-1}{\mu}.$$

Alternatively, we can write

$$\begin{aligned} t &= T\mathbb{P}(\tau_0 > T) + \big(\mathbb{E}\left[\tau_0 \mid \tau_0 > T\right] + t\big)\mathbb{P}(\tau_0 > T) \\ &= T\mathbb{P}(\tau_0 > T) + \mathbb{E}\left[\tau_0\mathbb{1}_{\{\tau_0>T\}}\right] + t\mathbb{P}(\tau_0 > T) \\ &= T\mathbb{P}(\tau_0 > T) + \mu\int_0^T s\mathrm{e}^{-\mu s}ds + t\mathbb{P}(\tau_0 < T), \end{aligned}$$

which yields the same result.

c) This proportion is

$$\frac{\mathbb{E}[W]}{\mathbb{E}[W]+1/\lambda} = \frac{\mathrm{e}^{\mu T}-1}{e^{\mu T}-1+\mu/\lambda}.$$

The state of the machine can be modeled using a two-state Markov chain with infinitesimal generator

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \dfrac{1}{\mathbb{E}[W]} & -\dfrac{1}{\mathbb{E}[W]} \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda \\ \dfrac{\mu}{\mathrm{e}^{\mu T}-1} & \dfrac{\mu}{1-\mathrm{e}^{\mu T}} \end{bmatrix},$$

and the above proportion can be rewritten using the stationary distribution

$$\begin{aligned}
[\pi_0,\pi_1] &= \left[\frac{1/\lambda}{1/\lambda+(\mathrm{e}^{\mu T}-1)/\mu}, \frac{(\mathrm{e}^{\mu T}-1)/\mu}{1/\lambda+(\mathrm{e}^{\mu T}-1)/\mu}\right] \\
&= \left[\frac{\mu}{\mu+\lambda(\mathrm{e}^{\mu T}-1)}, \frac{\lambda(\mathrm{e}^{\mu T}-1)}{\mu+\lambda(\mathrm{e}^{\mu T}-1)}\right] \\
&= \left[\frac{1/\lambda}{\mathbb{E}[W]+1/\lambda}, \frac{\mathbb{E}[W]}{\mathbb{E}[W]+1/\lambda}\right].
\end{aligned}$$

Exercise 9.16

a) The generator Q of $(X_t)_{t\in\mathbb{R}_+}$ is given by

$$Q = \begin{bmatrix} -0.5 & 0.5 & 0 \\ 0.2 & -0.45 & 0.25 \\ 0 & 0.4 & -0.4 \end{bmatrix}.$$

b) Solving for $\pi Q = 0$ we have

$$\begin{aligned}
\pi Q &= [\pi_0,\pi_1,\pi_2] \times \begin{bmatrix} -0.5 & 0.5 & 0 \\ 0.2 & -0.45 & 0.25 \\ 0 & 0.4 & -0.4 \end{bmatrix} \\
&= \begin{bmatrix} -0.5\times\pi_0 + 0.2\times\pi_1 \\ 0.5\times\pi_0 - 0.45\times\pi_1 + 0.4\times\pi_2 \\ 0.25\times\pi_1 - 0.4\times\pi_2 \end{bmatrix}^T = [0,0,0],
\end{aligned}$$

i.e. $\pi_0 = 0.4\times\pi_1 = 0.64\times\pi_2$ under the condition $\pi_0+\pi_1+\pi_2 = 1$, which gives $\pi_0 = 16/81$, $\pi_1 = 40/81$, $\pi_2 = 25/81$.

c) In the long run the average is

$$0\times\pi_0 + 1\times\pi_1 + 2\times\pi_2 = \frac{40}{81} + \frac{50}{81} = \frac{90}{81}.$$

d) We find

$$100\times\frac{90}{81} = \frac{1000}{9}.$$

e) We have

$$Q = \begin{bmatrix} -0.5 & 0.5 & 0 \\ 0.2 & -0.7 & 0.5 \\ 0 & 0.4 & -0.4 \end{bmatrix},$$

and solving $\pi Q = 0$ shows that

$$[\pi_0, \pi_1, \pi_2] = \left[\frac{0.32}{2.12}, \frac{0.8}{2.12}, \frac{1}{2.12}\right] = [0.15094, 0.37736, 0.47170].$$

Exercise 9.17 Both chains $(X_1(t))_{t\in\mathbb{R}}$ and $(X_2(t))_{t\in\mathbb{R}}$ have the same infinitesimal generator

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.$$

The infinitesimal generator of $Z(t) := X_1(t) + X_2(t)$ is given by

$$\begin{bmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -\lambda-\mu & \lambda \\ 0 & 2\mu & -2\mu \end{bmatrix},$$

as the birth rate λ is doubled when both chains are in state ⓪, and the death rate μ is also doubled when both chains are in state ①.

Note that the generator of the sum $(Z(t))_{t\in\mathbb{R}_+}$ of $(X_1(t))_{t\in\mathbb{R}_+}$ and $(X_2(t))_{t\in\mathbb{R}_+}$ is not the sum of the matrix generators of $(X_1(t))_{t\in\mathbb{R}_+}$ and $(X_2(t))_{t\in\mathbb{R}_+}$. Recall that by Proposition 9.6, the semi-groups of $X_1(t)$ and $X_2(t)$ are given by

$$\begin{aligned}
&\begin{bmatrix} \mathbb{P}(X_1(t)=0 \mid X_1(0)=0) & \mathbb{P}(X_1(t)=1 \mid X_1(0)=0) \\ \\ \mathbb{P}(X_1(t)=0 \mid X_1(0)=1) & \mathbb{P}(X_1(t)=1 \mid X_1(0)=1) \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{P}(X_2(t)=0 \mid X_2(0)=0) & \mathbb{P}(X_2(t)=1 \mid X_2(0)=0) \\ \\ \mathbb{P}(X_2(t)=0 \mid X_2(0)=1) & \mathbb{P}(X_2(t)=1 \mid X_2(0)=1) \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\mu}{\lambda+\mu} + \dfrac{\lambda}{\lambda+\mu}\mathrm{e}^{-t(\lambda+\mu)} & \dfrac{\lambda}{\lambda+\mu} - \dfrac{\lambda}{\lambda+\mu}\mathrm{e}^{-t(\lambda+\mu)} \\ \\ \dfrac{\mu}{\lambda+\mu} - \dfrac{\mu}{\lambda+\mu}\mathrm{e}^{-t(\lambda+\mu)} & \dfrac{\lambda}{\lambda+\mu} + \dfrac{\mu}{\lambda+\mu}\mathrm{e}^{-t(\lambda+\mu)} \end{bmatrix}.
\end{aligned}$$

As for the transition semi-group of $Z(t)$, we have

$$\begin{aligned}
P_{0,0}(t) &= \mathbb{P}(Z(t)=0 \mid Z(0)=0) \\
&= \mathbb{P}(X_1(t)=0 \text{ and } X_2(t)=0 \mid X_1(0)=0 \text{ and } X_2(0)=0) \\
&= \mathbb{P}(X_1(t)=0 \mid X_1(0)=0)\mathbb{P}(X_2(t)=0 \mid X_2(0)=0)
\end{aligned}$$

$$
\begin{aligned}
&= (\mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0))^2 \\
&= \left(\frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\mathrm{e}^{-t(\lambda+\mu)}\right)^2.
\end{aligned}
$$

For $P_{0,1}(t)$ we have

$$
\begin{aligned}
P_{0,1}(t) &= \mathbb{P}(Z(t) = 1 \mid Z(0) = 0) \\
&= \mathbb{P}(X_1(t) = 0 \text{ and } X_2(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 0) \\
&\quad + \mathbb{P}(X_1(t) = 1 \text{ and } X_2(t) = 0 \mid X_1(0) = 0 \text{ and } X_2(0) = 0) \\
&= \mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 0 \mid X_2(0) = 0) \\
&\quad + \mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 0) \\
&= 2\mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 0 \mid X_2(0) = 0).
\end{aligned}
$$

Starting from $Z(0) = 1$ and ending at $Z(t) = 1$ we have two possibilities $(0,1)$ or $(1,0)$ for the terminal condition. As for the initial condition $Z(0) = 1$ the two possibilities $(0,1)$ and $(1,0)$ count for one only since they both give $Z(0) = 1$. We have

$$
\begin{aligned}
&\mathbb{P}(Z(t) = 1 \mid Z(0) = 1) \\
&= \mathbb{P}(Z(t) = 1 \mid \{X_1(0) = 0 \text{ and } X_2(0) = 1\} \cup \{X_1(0) = 1 \text{ and } X_2(0) = 0\}) \\
&= \frac{\mathbb{P}(\{Z(t) = 1\} \cap (\{X_1(0) = 0 \text{ and } X_2(0) = 1\} \cup \{X_1(0) = 1 \text{ and } X_2(0) = 0\}))}{\mathbb{P}(\{X_1(0) = 0 \text{ and } X_2(0) = 1\} \cup \{X_1(0) = 1 \text{ and } X_2(0) = 0\})} \\
&= \frac{\mathbb{P}(\{Z(t) = 1\} \cap \{X_1(0) = 0 \text{ and } X_2(0) = 1\})}{\mathbb{P}(Z(0) = 1)} \\
&\quad + \frac{\mathbb{P}(\{Z(t) = 1\} \cap \{X_1(0) = 1 \text{ and } X_2(0) = 0\})}{\mathbb{P}(Z(0) = 1)} \\
&= \mathbb{P}(Z(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\frac{\mathbb{P}(X_1(0) = 0 \text{ and } X_2(0) = 1)}{\mathbb{P}(Z(0) = 1)} \\
&\quad + \mathbb{P}(Z(t) = 1 \mid X_1(0) = 1 \text{ and } X_2(0) = 0)\frac{\mathbb{P}(X_1(0) = 1 \text{ and } X_2(0) = 0)}{\mathbb{P}(Z(0) = 1)} \\
&= \mathbb{P}(Z(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\frac{\mathbb{P}(X_1(0) = 0 \text{ and } X_2(0) = 1)}{\mathbb{P}(Z(0) = 1)} \\
&\quad + \mathbb{P}(Z(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\frac{\mathbb{P}(X_1(0) = 1 \text{ and } X_2(0) = 0)}{\mathbb{P}(Z(0) = 1)} \\
&= \mathbb{P}(Z(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1).
\end{aligned}
$$

The above equality

$$
\mathbb{P}(Z(t) = 1 \mid X_1(0) = 1 \text{ and } X_2(0) = 0) = \mathbb{P}(Z(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)
$$

is justified by the fact that the transition probabilities of $(Z(t))_{t\in\mathbb{R}_+}$ depend only on $Z(0) = X_1(0) + X_2(0) = 0 + 1 = 1 + 0 = 1$.

Thus, in order to compute $P_{1,1}(t)$ we can choose to assign the value 0 to $X_1(0)$ and the value 1 to $X_2(0)$ without influencing the final result, as the other choice would lead to the same probability value. Hence for $P_{1,1}(t)$ we have

$$\begin{aligned}P_{1,1}(t) &= \mathbb{P}(Z(t) = 1 \mid Z(0) = 1)\\ &= \mathbb{P}(X_1(t) = 0 \text{ and } X_2(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\\ &\quad + \mathbb{P}(X_1(t) = 1 \text{ and } X_2(t) = 0 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\\ &= \mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 1)\\ &\quad + \mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 0 \mid X_2(0) = 1).\end{aligned}$$

Note that in the above calculation we chose to represent the event $\{Z(0) = 1\}$ by $\{X_1(0) = 0 \text{ and } X_2(0) = 1\}$, however making the other choice $\{X_1(0) = 1 \text{ and } X_2(0) = 0\}$ would lead to the same result because $X_1(t)$ and $X_2(t)$ have same infinitesimal generators.

Concerning $P_{1,0}(t)$ we have

$$\begin{aligned}P_{1,0}(t) &= \mathbb{P}(Z(t) = 0 \mid Z(0) = 1)\\ &= \mathbb{P}(X_1(t) = 0 \text{ and } X_2(t) = 0 \mid X_1(0) = 0 \text{ and } X_2(0) = 1).\end{aligned}$$

On the other hand, we have

$$\begin{aligned}P_{1,2}(t) &= \mathbb{P}(Z(t) = 2 \mid Z(0) = 1)\\ &= \mathbb{P}(X_1(t) = 1 \text{ and } X_2(t) = 1 \mid X_1(0) = 0 \text{ and } X_2(0) = 1)\\ &= \mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 1).\end{aligned}$$

We check that

$$\begin{aligned}P_{1,0}(t) + P_{1,1}(t) + P_{1,2}(t) &= \mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 0 \mid X_2(0) = 1)\\ &\quad + \mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 1)\\ &\quad + \mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 0 \mid X_2(0) = 1)\\ &\quad + \mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0)\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 1)\\ &= (\mathbb{P}(X_1(t) = 1 \mid X_1(0) = 0) + \mathbb{P}(X_1(t) = 0 \mid X_1(0) = 0))\\ &\quad \times(\mathbb{P}(X_2(t) = 1 \mid X_2(0) = 1) + \mathbb{P}(X_2(t) = 0 \mid X_2(0) = 1))\\ &= 1.\end{aligned}$$

Note that for $Z(t)$ to be Markov, the processes $X_1(t)$ and $X_2(t)$ should have same infinitesimal generators. For example, if $X_1(t)$ and $X_2(t)$ have different transition rates, then starting from $Z(t) = 1$ we need the information whether $X_1(t) = 1$ or $X_2(t) = 1$ in order to determine what will be the next transition rate. However, the knowledge of $Z(t) = 1$ is not sufficient for this.

Altogether, there are $3 \times 3 = 9$ transition probabilities to compute since the chain $Z(t)$ has 3 states $\{0, 1, 2\}$, and the remaining computations are left to the reader.

Exercise 9.18 Starting from state ⓪, the process $X_t = \xi_{N_t}$ stays at state ⓪ during an exponentially distributed Poisson interjump time with parameter λ, after which N_t increases by one unit. In this case, $\xi_{N_t} = 0$ becomes $\xi_{N_t+1} = 1$ with probability 1, from the transition matrix (9.7.4), hence the birth rate of X_t from state ⓪ to state ① is λ.

Then, starting from state ①, the process X_t stays at ① during an exponentially distributed time with parameter λ. The difference is that when N_t increases by one unit, $\xi_{N_t} = 1$ may move to $\xi_{N_t+1} = 0$ with probability $1 - \alpha$, or remain at $\xi_{N_t+1} = 1$ with probability α.

In fact, due to the Markov property, X_t will remain at 1 during an exponentially distributed time whose expectation may be higher than $1/\lambda$ when $\alpha > 0$. We will compute the expectation of this random time.

a) We have

$$\begin{aligned}\mathbb{E}\big[T_0^r \mid X_0 = 1\big] &= \alpha\left(\frac{1}{\lambda} + \mathbb{E}\big[T_0^r \mid X_0 = 1\big]\right) + (1-\alpha) \times \left(\frac{1}{\lambda} + 0\right) \\ &= \frac{1}{\lambda} + \alpha \mathbb{E}\big[T_0^r \mid X_0 = 1\big],\end{aligned}$$

hence

$$\mathbb{E}\big[T_0^r \mid X_0 = 1\big] = \frac{1}{\lambda(1-\alpha)} \tag{S.61}$$

and

$$\mathbb{E}\big[T_0^r \mid X_0 = 0\big] = \frac{1}{\lambda} + 1 \times \mathbb{E}\big[T_0^r \mid X_0 = 1\big] = \frac{2-\alpha}{\lambda(1-\alpha)}.$$

Note that (S.61) can also be recovered from (5.3.3) by letting $b = 1 - \alpha$ and multiplying by the average Poisson interjump time $1/\lambda$.

b) We have

$$\mathbb{E}\big[T_1^r \mid X_0 = 1\big] = \frac{1}{\lambda} + (1-\alpha)\mathbb{E}\big[T_1^r \mid X_0 = 0\big] = \frac{2-\alpha}{\lambda},$$

since $\mathbb{E}\big[T_1^r \mid X_0 = 0\big] = 1/\lambda$.

c) This continuous-time first step analysis argument is similar to the one used in the solution of Exercise 9.3. Since

$$\mathbb{E}\big[T_0^r \mid X_0 = 1\big] = \frac{1}{\lambda(1-\alpha)},$$

it takes an exponential random time with parameter $\lambda(1-\alpha)$ for the process $(X_t)_{t\in\mathbb{R}_+}$ to switch from state ① to state ⓪. Hence the death rate is

$$\frac{1}{\mathbb{E}\big[T_0^r \mid X_0 = 1\big]} = \lambda(1-\alpha),$$

and the infinitesimal generator Q of X_t is

$$\left[\begin{array}{cc} -\lambda & \lambda \\ \dfrac{1}{\mathbb{E}\big[T_0^r \mid X_0 = 1\big]} & -\dfrac{1}{\mathbb{E}\big[T_0^r \mid X_0 = 1\big]} \end{array}\right] = \left[\begin{array}{cc} -\lambda & \lambda \\ (1-\alpha)\lambda & -(1-\alpha)\lambda \end{array}\right].$$

Problem 9.19

a) We need to show the following properties.

(i) The process $(N_t^1 + N_t^2)_{t\in\mathbb{R}_+}$ is a counting process.

Clearly, the jump heights are positive integers and they can only be equal to one since the probability that N_t^1 and N_t^2 jumps simultaneously is 0.

(ii) The process $(N_t^1 + N_t^2)_{t\in\mathbb{R}_+}$ has independent increments.

Letting $0 < t_1 < t_2 < \cdots < t_n$, the family

$$\begin{aligned}&\Big(N_{t_n}^1 + N_{t_n}^2 - (N_{t_{n-1}}^1 + N_{t_{n-1}}^2), \ldots, N_{t_2}^1 + N_{t_2}^2 - (N_{t_1}^1 + N_{t_1}^2)\Big) \\ &\quad = \Big(N_{t_n}^1 - N_{t_{n-1}}^1 + N_{t_n}^2 - N_{t_{n-1}}^2, \ldots, N_{t_2}^1 - N_{t_1}^1 + N_{t_2}^2 - N_{t_1}^2\Big)\end{aligned}$$

is a family of independent random variables. In order to see this we note that $N_{t_n}^1 - N_{t_{n-1}}^1$ is independent of

$$N_{t_{n-1}}^1 - N_{t_{n-2}}^1, \ldots, N_{t_2}^1 - N_{t_1}^1,$$

and of

$$N_{t_n}^2 - N_{t_{n-1}}^2, \ldots, N_{t_2}^2 - N_{t_1}^2,$$

hence it is also independent of

$$N_{t_{n-1}}^1 - N_{t_{n-2}}^1 + N_{t_{n-1}}^2 - N_{t_{n-2}}^2, \ldots, N_{t_2}^1 - N_{t_1}^1 + N_{t_2}^2 - N_{t_1}^2.$$

Similarly it follows that $N_{t_n}^2 - N_{t_{n-1}}^2$ is independent of

$$N_{t_{n-1}}^1 - N_{t_{n-2}}^1 + N_{t_{n-1}}^2 - N_{t_{n-2}}^2, \ldots, N_{t_2}^1 - N_{t_1}^1 + N_{t_2}^2 - N_{t_1}^2,$$

hence $N^1_{t_n}+N^2_{t_n}-(N^1_{t_{n-1}}+N^2_{t_{n-1}})$ is independent of

$$N^1_{t_{n-1}}-N^1_{t_{n-2}}+N^2_{t_{n-1}}-N^2_{t_{n-2}},\ldots,N^1_{t_2}-N^1_{t_1}+N^2_{t_2}-N^2_{t_1}.$$

This shows the required mutual independence by induction on $n\geqslant 1$.

(iii) The process $(N^1_t+N^2_t)_{t\in\mathbb{R}_+}$ has stationary increments.

We note that the distributions of the random variables $N^1_{t+h}-N^1_{s+h}$ and $N^2_{t+h}-N^2_{s+h}$ do not depend on $h\in\mathbb{R}_+$, hence by the law of total probability we check that

$$\begin{aligned}&\mathbb{P}(N^1_{t+h}+N^2_{t+h}-(N^1_{s+h}+N^2_{s+h})=n)\\&\quad=\sum_{k=0}^{n}\mathbb{P}(N^1_{t+h}-N^1_{s+h}=k)\mathbb{P}(N^2_{t+h}-N^2_{s+h}=n-k)\end{aligned}$$

is independent of $h\in\mathbb{R}_+$.

The intensity of $N^1_t+N^2_t$ is $\lambda_1+\lambda_2$.

b) (i) The proof of independence of increments is similar to that of Question (a).

(ii) Concerning the stationarity of increments we have

$$\begin{aligned}\mathbb{P}(M_{t+h}-M_t=n)&=\mathbb{P}(N^1_{t+h}-N^2_{t+h}-(N^1_{s+h}-N^2_{s+h})=n)\\&=\mathbb{P}(N^1_{t+h}-N^1_{s+h}-(N^2_{t+h}-N^2_{s+h})=n)\\&=\sum_{k\geqslant 0}\mathbb{P}(N^1_{t+h}-N^1_{s+h}=n+k)\mathbb{P}(N^2_{t+h}-N^2_{s+h}=k)\end{aligned}$$

which is independent of $h\in\mathbb{R}_+$ since the distributions of $N^1_{t+h}-N^1_{s+h}$ and $N^2_{t+h}-N^2_{s+h}$ are independent of $h\in\mathbb{R}_+$.

c) For $n\in\mathbb{N}$ we have

$$\begin{aligned}\mathbb{P}(M_t=n)&=\mathbb{P}(N^1_t-N^2_t=n)\\&=\sum_{k\geqslant \mathrm{Max}(0,-n)}\mathbb{P}(N^1_t=n+k)\mathbb{P}(N^2_t=k)\\&=\mathrm{e}^{-(\lambda_1+\lambda_2)t}\sum_{k\geqslant \mathrm{Max}(0,-n)}\frac{\lambda_1^{n+k}\lambda_2^k t^{n+2k}}{k!(n+k)!}\\&=\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2}\mathrm{e}^{-(\lambda_1+\lambda_2)t}\sum_{k\geqslant \mathrm{Max}(0,-n)}\frac{(t\sqrt{\lambda_1\lambda_2})^{n+2k}}{(n+k)!k!}\\&=\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2}\mathrm{e}^{-(\lambda_1+\lambda_2)t}I_{|n|}(2t\sqrt{\lambda_1\lambda_2}),\end{aligned}$$

where

$$I_n(x) = \sum_{k\geqslant 0} \frac{(x/2)^{n+2k}}{k!(n+k)!}, \qquad x > 0,$$

is the modified Bessel function with parameter $n \geqslant 0$. When $n \leqslant 0$, by exchanging λ_1 and λ_2 we get

$$\begin{aligned}\mathbb{P}(M_t = n) &= \mathbb{P}(-M_t = -n) \\ &= \left(\frac{\lambda_2}{\lambda_1}\right)^{-n/2} \mathrm{e}^{-(\lambda_1+\lambda_2)t} I_{-n}(2t\sqrt{\lambda_1\lambda_2}) \\ &= \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \mathrm{e}^{-(\lambda_1+\lambda_2)t} I_{-n}(2t\sqrt{\lambda_1\lambda_2}),\end{aligned}$$

hence in the general case we have

$$\mathbb{P}(M_t = n) = \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \mathrm{e}^{-(\lambda_1+\lambda_2)t} I_{|n|}(2t\sqrt{\lambda_1\lambda_2}), \qquad n \in \mathbb{Z},$$

which is known as the *Skellam distribution*.

This also shows that the semigroup $P(t)$ of the birth and death process with state-independent birth and death rates λ_1 and λ_2 satisfies

$$\begin{aligned}P_{i,j}(t) &= \mathbb{P}(M_t = j \mid M_0 = i) \\ &= \mathbb{P}(M_t = j - i) \\ &= \left(\frac{\lambda_1}{\lambda_2}\right)^{(j-i)/2} \mathrm{e}^{-(\lambda_1+\lambda_2)t} I_{|j-i|}(2t\sqrt{\lambda_1\lambda_2}),\end{aligned}$$

$i, j \in \mathbb{Z}$, $t \geqslant 0$.

When $\lambda_1 = \lambda_2 = \lambda$ we find

$$\mathbb{P}(M_t = n) = \mathrm{e}^{-2\lambda t} I_{|n|}(2\lambda t).$$

d) From the bound*

$$I_{|n|}(y) < C_n y^{|n|} \mathrm{e}^{y}, \qquad y > 1,$$

we get

$$\mathbb{P}(M_t = n) \leqslant C_n \mathrm{e}^{-(\lambda_1+\lambda_2)t} \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \left(2t\sqrt{\lambda_1\lambda_2}\right)^{|n|} \mathrm{e}^{2t\sqrt{\lambda_1\lambda_2}}$$

* See *e.g.* Theorem 2.1 of Laforgia (1991) for a proof of this inequality.

$$= C_n \mathrm{e}^{-t(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2} \left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} \left(2t\sqrt{\lambda_1\lambda_2}\right)^{|n|},$$

which tends to 0 as t goes to infinity when $\lambda_1 \neq \lambda_2$.

Hence we have*

$$\lim_{t\to\infty} \mathbb{P}(|M_t| < c) = \sum_{-c<k<c} \lim_{t\to\infty} \mathbb{P}(|M_t| = k) = 0, \qquad c > 0. \qquad \text{(S.62)}$$

Remark. There exists a shorter proof by a probabilistic argument (not required here), which is displayed below, cf. Bosq and Nguyen (1996).

Noting that since $\mathbb{E}[M_t] = t(\lambda_1 - \lambda_2)$, for t large enough we have, depending on the sign of $\lambda_1 - \lambda_2$,

$$c - \mathbb{E}[M_t] \leqslant -\frac{1}{2}\mathbb{E}[M_t] \quad \text{or} \quad -c - \mathbb{E}[M_t] \geqslant \frac{1}{2}\mathbb{E}[M_t],$$

we get

$$-c \leqslant M_t \leqslant c \quad \Longrightarrow \quad M_t - \mathbb{E}[M_t] \leqslant -\frac{1}{2}\mathbb{E}[M_t] \quad \text{or} \quad M_t - \mathbb{E}[M_t] \geqslant \frac{1}{2}\mathbb{E}[M_t],$$

hence by Chebyshev's inequality, we have

$$\begin{aligned}\mathbb{P}(|M_t| \leqslant c) &\leqslant \mathbb{P}\left(|M_t - \mathbb{E}[M_t]| \geqslant \frac{1}{2}\mathbb{E}[M_t]\right)\\ &\leqslant 4\frac{\mathrm{Var}[M_t]}{(\mathbb{E}[M_t])^2} = \frac{4}{t}\frac{\lambda_1+\lambda_2}{(\lambda_1-\lambda_2)^2},\end{aligned}$$

which tends to 0 as t tends to infinity, provided that $\lambda_1 \neq \lambda_2$. A probabilistic proof is also available in case $\lambda_1 = \lambda_2$, using the central limit theorem.

e) When $M_t \geqslant 0$, M_t represents the number of waiting customers. When $M_t \leqslant 0$, $-M_t$ represents the number of waiting drivers.

Relation (S.62) shows that for any fixed $c > 0$, the probability of having either more than c waiting customers or more than c waiting drivers is high in the long run.

Problem 9.20

a) We have

* Treating the case $\lambda_1 = \lambda_2$ is more complicated and not required.

$$Q = \begin{bmatrix} -N\lambda & N\lambda & 0 & \cdots & \cdots & 0 & 0 & 0 \\ \mu & -\mu-(N-1)\lambda & (N-1)\lambda & \cdots & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & (N-1)\mu & -(N-1)\mu-\lambda & \lambda \\ 0 & 0 & 0 & \cdots & \cdots & 0 & N\mu & -N\mu \end{bmatrix}.$$

b) The system of equations follows by writing the matrix multiplication $P'(t) = P(t)Q$ term by term.

c) We apply the result of Question (b) to

$$\frac{\partial G_k}{\partial t}(s,t) = \sum_{n=0}^{N} s^n P'_{k,n}(t),$$

and use the expression

$$\frac{\partial G_k}{\partial s}(s,t) = \sum_{n=1}^{N} n s^{n-1} P'_{k,n}(t).$$

d) We have

$$\begin{aligned}
&\lambda N(s-1)G_k(s,t) + (\mu + (\lambda-\mu)s - \lambda s^2)\frac{\partial G_k}{\partial s}(s,t) - \frac{\partial G_k}{\partial t}(s,t) \\
&= -(s-1)(\lambda+\mu)(N-k)\left((s-1)\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^k \lambda\left(-(s-1)\lambda \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{N-k-1} \mathrm{e}^{-(\lambda+\mu)t} \\
&+(s-1)(\lambda+\mu)k\mu\left((s-1)\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{k-1}\left(-(s-1)\lambda \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{N-k} \mathrm{e}^{-(\lambda+\mu)t} \\
&+(s-1)\left((s-1)\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^k N\lambda\left(-(s-1)\lambda \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{N-k} \\
&+\left(\lambda s^2 - (\lambda-\mu)s - \mu\right)(N-k)\left(\lambda \mathrm{e}^{-(\lambda+\mu)t} - \lambda\right)\left((s-1)\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^k\left(-(s-1)\lambda \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{N-k-1} \\
&-\left(\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda\right)k\left((s-1)\mu \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{k-1}\left(-(s-1)\lambda \mathrm{e}^{-(\lambda+\mu)t} + \lambda s + \mu\right)^{N-k}\left(\lambda s^2 - (\lambda-\mu)s - \mu\right) \\
&= 0.
\end{aligned}$$

e) This expression follows from the relation

$$\mathbb{E}[X_t \mid X_0 = k] = \frac{\partial G_k}{\partial s}(s,t)_{|s=1}$$

and the result of Question (d).

f) We have

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{E}[X_t \mid X_0 = k] &= k\frac{\lambda(\lambda+\mu)^{k-1}}{(\lambda+\mu)^N}(\mu+\lambda)^{N-k} \\
&\quad +(N-k)\frac{(\mu+\lambda)^k\lambda(\mu+\lambda)^{N-k-1}}{(\lambda+\mu)^N} = \frac{N\lambda}{\lambda+\mu}.
\end{aligned}$$

Problem 9.21 First, we note that in case

$$f(x) = \mathbb{1}_{[a,b]}(x), \qquad 0 \leqslant a \leqslant b \leqslant t,$$

we have, by direct counting,

$$\sum_{k=1}^{N_t} f(T_k) = \sum_{k=1}^{N_t} \mathbb{1}_{[a,b]}(T_k) = N_b - N_a.$$

Hence

$$\mathbb{E}\left[\sum_{k=1}^{N_t} f(T_k)\right] = \mathbb{E}[N_b - N_a] = \lambda(b-a) = \lambda \int_0^t \mathbb{1}_{[a,b]}(s)ds = \lambda \int_0^t f(s)ds,$$

hence (9.7.6) is proved for $f(x) = \mathbb{1}_{[a,b]}(x)$.

Next, we check by linearity that the (9.7.6) extends to a linear combination of indicator functions of the form

$$f(x) = \sum_{k=1}^{n} \alpha_k \mathbb{1}_{[a_{k-1},a_k]}(x), \qquad 0 \leqslant a_0 < a_1 < \cdots < a_n < t.$$

The difficult part is to do the extension from the linear combinations of indicator functions to "any" integrable function $f : [0,t] \longrightarrow \mathbb{R}$. This requires the knowledge of measure theory. For a different proof using the exponentially distributed jump times of the Poisson process, see Proposition 2.3.4 in Privault (2009).

Let g_n, $n \geqslant 1$, be defined as

$$\begin{aligned} g_n(t_1,t_2,\ldots,t_n) = &\sum_{k=1}^{n} \mathbb{1}_{\{t_{k-1}<t<t_k\}}(f(t_1)+\cdots+f(t_{k-1})) \\ &+\mathbb{1}_{\{t_n<t\}}(f(t_1)+\cdots+f(t_n)), \end{aligned}$$

with $t_0 := 0$, so that

$$\sum_{k=1}^{\min(n,N_t)} f(T_k) = g_n(T_1,T_2,\ldots,T_n).$$

Then

$$\mathbb{E}\left[\sum_{k=1}^{\min(n,N_t)} f(T_k)\right] = \mathbb{E}[g(T_1,T_2,\ldots,T_n)]$$

$$= \lambda^n \sum_{k=1}^{n} \int_0^\infty \mathrm{e}^{-\lambda t_n} \int_0^{t_n} \cdots \int_0^{t_2} \mathbb{1}_{\{t_{k-1}<t<t_k\}}(f(t_1)+\cdots+f(t_{k-1}))dt_1\cdots dt_n$$

$$
\begin{aligned}
&+\lambda^n \int_0^t \mathrm{e}^{-\lambda t_n} \int_0^{t_n} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_n)) dt_1 \cdots dt_n \\
&= \lambda^n \sum_{k=1}^n \int_t^\infty \mathrm{e}^{-\lambda t_n} \frac{(t_n-t)^{n-k}}{(n-k-1)!} dt_n \\
&\quad \times \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_{k-1})) dt_1 \cdots dt_{k-1} \\
&\quad +\lambda^n \int_0^t \mathrm{e}^{-\lambda t_n} \int_0^{t_n} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_n)) dt_1 \cdots dt_n \\
&= \mathrm{e}^{-\lambda t} \sum_{k=1}^n \lambda^k \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-k}}{(n-k-1)!} dt \\
&\quad \times \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_{k-1})) dt_1 \cdots dt_{k-1} \\
&\quad +\lambda^n \int_0^t \mathrm{e}^{-\lambda t_n} \int_0^{t_n} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_n)) dt_1 \cdots dt_n \\
&= \mathrm{e}^{-\lambda t} \sum_{k=1}^n \lambda^k \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_{k-1})) dt_1 \cdots dt_{k-1} \\
&\quad +\mathrm{e}^{-\lambda t} \sum_{k \geqslant n} \lambda^k \int_0^t \frac{(t-t_n)^{k-n}}{(k-n)!} \int_0^{t_n} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_n)) dt_1 \cdots dt_n \\
&= \mathrm{e}^{-\lambda t} \sum_{k=1}^n \lambda^k \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_{k-1})) dt_1 \cdots dt_{k-1} \\
&\quad +\mathrm{e}^{-\lambda t} \sum_{k \geqslant n} \lambda^k \int_0^t \int_{t_n}^t \int_{t_n}^{t_k} \cdots \int_{t_n}^{t_{n+2}} dt_{n+1} \cdots dt_k \\
&\quad \times \int_0^{t_{n+1}} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_n)) dt_1 \cdots dt_n \\
&= \mathrm{e}^{-\lambda t} \sum_{k \geqslant 0} \lambda^k \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} (f(t_1)+\cdots+f(t_{\min(k,n)})) dt_1 \cdots dt_k \\
&= \mathrm{e}^{-\lambda t} \sum_{k \geqslant 0} \frac{\lambda^k}{k!} \int_0^t \cdots \int_0^t (f(t_1)+\cdots+f(t_{\min(k,n)})) dt_1 \cdots dt_k.
\end{aligned}
$$

Hence as n goes to infinity,

$$
\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{N_t} f(T_k)\right] &= \lim_{n\to\infty} \mathbb{E}\left[\sum_{k=1}^{\min(n,N_t)} f(T_k)\right] \\
&= \lim_{n\to\infty} \mathrm{e}^{-\lambda t} \sum_{k \geqslant 0} \frac{\lambda^k}{k!} \int_0^t \cdots \int_0^t (f(t_1)+\cdots+f(t_{\min(k,n)})) dt_1 \cdots dt_k \\
&= \mathrm{e}^{-\lambda t} \sum_{k \geqslant 0} \frac{\lambda^k}{k!} \int_0^t \cdots \int_0^t (f(t_1)+\cdots+f(t_k)) dt_1 \cdots dt_k
\end{aligned}
$$

$$= \lambda \mathrm{e}^{-\lambda t} \int_0^t f(s) ds \sum_{k \geqslant 1} \frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda \int_0^t f(s) ds.$$

Chapter 10 - Discrete-Time Martingales

Exercise 10.1 For all $k = 0, 1, \ldots, B$ and $n \geqslant 1$ we have

$$\begin{aligned} \mathbb{P}(\tau_{0,B} = \infty \mid S_0 = k) &\leqslant \mathbb{P}(\tau_{0,B} > nB \mid S_0 = k) \\ &\leqslant \mathbb{P}\left(\bigcap_{k=0}^{n-1} \{X_{kB+1} = 1, \ldots, X_{kB+B} = 1\}^c \right) \\ &= (1 - p^k)^n, \end{aligned}$$

from which we obtain $\mathbb{P}(\tau_{0,B} = \infty \mid S_0 = k) = 0$ after letting n tend to infinity when $p \in [0, 1)$, hence $\mathbb{P}(\tau_{0,B} < \infty \mid S_0 = k) = 1$. In case $p = 1$, we clearly have $\mathbb{P}(\tau_{0,B} < \infty \mid S_0 = k) = 1$.

Exercise 10.2

a) We have

$$\begin{aligned} \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[\sum_{k=1}^{n+1} 2^{k-1} X_k \;\middle|\; \mathcal{F}_n \right] \\ &= \mathbb{E}\left[\sum_{k=1}^{n} 2^{k-1} X_k \;\middle|\; \mathcal{F}_n \right] + \mathbb{E}[2^n X_{n+1} \mid \mathcal{F}_n] \\ &= \sum_{k=1}^{n} 2^{k-1} X_k + 2^n \mathbb{E}[X_{n+1}] \\ &= M_n, \qquad n \geqslant 0. \end{aligned}$$

b) This random time is a hitting time, so it is a stopping time.
c) The strategy of the gambler is to double the stakes each time he loses (double down strategy), and to quit the game as soon as his gains reach \$1.
d) The two possible values of $M_{\tau \wedge n}$ are 1 and

$$-\sum_{k=1}^{n} 2^{k-1} = -\frac{1 - 2^n}{1 - 2} = 1 - 2^n, \qquad n \geqslant 1.$$

We have

$$\mathbb{P}(M_{n\wedge\tau} = 1 - 2^n) = 2^{-n} \quad \text{and} \quad \mathbb{P}(M_{n\wedge\tau} = 1) = 1 - 2^{-n}, \qquad n \geqslant 1.$$

e) We have

$$\begin{aligned}\mathbb{E}[M_{n\wedge\tau}] &= (1-2^n)\mathbb{P}(M_{n\wedge\tau} = 1-2^n) + \mathbb{P}(M_{n\wedge\tau} = 1) \\ &= (1-2^n)2^{-n} + (1-2^{-n}) \\ &= 0, \qquad n \geqslant 1.\end{aligned}$$

f) The Stopping Time Theorem 10.9 directly states that

$$\mathbb{E}[M_{n\wedge\tau}] = \mathbb{E}[M_0] = 0.$$

Exercise 10.3

a) The random time τ is a stopping time because for every $n \geqslant 0$, the validity of the event $\{\tau > n\}$ can be decided by studying the path of the process $(S_k)_{k\geqslant 0}$ until time n.
b) The range of possible values of S_τ is $\{-1, 0, 1, 2, 3, 4, \ldots\}$.

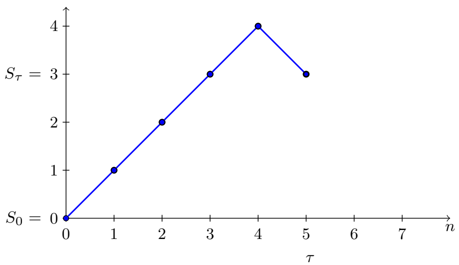


Fig. S.28: Sample path of the random walk $(S_n)_{n\geqslant 0}$.

c) According to the stopping time theorem, the stopped process $(S_{n\wedge\tau})_{n\geqslant 0}$ is a martingale, and therefore we have

$$\mathbb{E}[S_\tau] = \mathbb{E}\Big[\lim_{n\to\infty} S_{n\wedge\tau}\Big] = \lim_{n\to\infty} \mathbb{E}[S_{n\wedge\tau}] = \lim_{n\to\infty} \mathbb{E}[S_0] = 0.$$

The exchange between limit and expected value can be justified from the dominated convergence theorem, see the next question.
d) We have $\mathbb{P}(S_\tau = -1) = 1/2$ and $\mathbb{P}(S_\tau = k) = 1/2^{k+2}$, $k \geqslant 0$. In particular, we have $S_{n\wedge\tau} \leqslant 1 + S_\tau$, and S_τ is integrable according to the next question.
e) We have

$$\begin{aligned}\mathbb{E}[S_\tau] &= \sum_{k\in\mathbb{Z}} k\mathbb{P}(S_\tau = k)\\ &= -\frac{1}{2} + \sum_{k\geqslant 0} k\mathbb{P}(S_\tau = k)\\ &= -\frac{1}{2} + \sum_{k\geqslant 0} \frac{k}{2^{k+2}}\\ &= -\frac{1}{2} + \frac{1}{8}\sum_{k\geqslant 0} \frac{k}{2^{k-1}}\\ &= -\frac{1}{2} + \frac{1}{8}\frac{1}{(1-1/2)^2}\\ &= 0.\end{aligned}$$

Exercise 10.4 We note that for all $n \geqslant 1$ we have

$$\{\tau = n\} = \{\tau \geqslant n\} \setminus \{\tau > n\} = \{\tau > n-1\} \cap \{\tau > n\}^c \in \mathcal{F}_n,$$

since $\{\tau > n\} \in \mathcal{F}_n$ and $\{\tau > n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$. In case $n = 0$, we have

$$\{\tau = 0\} = \{\tau > 0\}^c \in \mathcal{F}_0.$$

Conversely, assuming that (10.5.1) holds, we have

$$\{\tau > n\} = \{\tau \leqslant n\}^c = \left(\bigcup_{k=0}^n \{\tau = k\}\right)^c = \bigcap_{k=0}^n \{\tau = k\}^c \in \mathcal{F}_n,$$

since

$$\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n, \qquad k = 0, 1, \ldots, n.$$

Exercise 10.5 We consider three examples.

i) Consider the stopped martingale $(X_n)_{n\geqslant 0} := (M_{\tau\wedge n})_{n\geqslant 0}$ of Exercise 10.2, for which by Question (e) therein we have $\mathbb{E}[X_n] = \mathbb{E}[M_{\tau\wedge n}] = 0$, $n \geqslant 0$, hence

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[M_{\tau\wedge n}] = 0,$$

while

$$\lim_{n\to\infty} X_n = \lim_{n\to\infty} M_{\tau\wedge n} = 1,$$

a.s.. Indeed, by (1.2.5), the *non-decreasing* sequence of events $(\{X_n = 1\})_{n\geqslant 1}$ satisfies

$$\mathbb{P}\big(\lim_{n\to\infty} X_n = 1\big) = \mathbb{P}\left(\bigcup_{n\geqslant 1}\{X_n = 1\}\right)$$
$$= \lim_{n\to\infty} \mathbb{P}(X_n = 1)$$
$$= \lim_{n\to\infty}\left(1 - \frac{1}{2^n}\right)$$
$$= 1.$$

Note that here the sequence $(M_n)_{n\geqslant 0}$ is unbounded, as

$$\inf_{n\geqslant 0} M_n = -\infty \quad and \quad \operatorname*{Sup}_{n\geqslant 0} |M_n| = +\infty.$$

ii) Consider the random sequence $X_n := n\mathbb{1}_{\{U<1/n\}}$, $n \geqslant 1$, where $U \simeq U(0,1]$ is a uniformly distributed random variable on $(0,1]$. More formally, $(X_n)_{n\geqslant 1}$ can be constructed by *e.g.* taking $\Omega = (0,1)$ with the probability measure $\mathbb{P}$ defined by $\mathbb{P}([a,b]) = b-a$, $0 < a \leqslant b < 1$, and defining the random variable $X_n(\omega) = n\mathbb{1}_{\{0<\omega<1/n\}}$, $\omega \in (0,1)$, in which case we have $\operatorname{Sup}_{n\geqslant 1}|X_n| = +\infty$, $\mathbb{E}[X_n] = n \times (1/n - 0) = 1$, $n \geqslant 1$, and $\lim_{n\to\infty} X_n(\omega) = 0$ for all $\omega \in (0,1)$.

iii) Consider the random sequence $X_n := n^2\mathbb{1}_{\{U_n<1/n^2\}}$, $n \geqslant 1$, where $(U_n)_{n\geqslant 1} \simeq U(0,1]$ is a sequence of uniformly distributed random variables on $(0,1]$. On the one hand, we have

$$\mathbb{E}[X_n] = n^2\mathbb{E}[\mathbb{1}_{\{U_n<1/n^2\}}] = n^2\mathbb{P}(\mathbb{1}_{\{U_n<1/n^2\}}) = 1, \qquad n \geqslant 1.$$

On the other hand, we have

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n > 0) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

hence by the Borel-Cantelli Lemma, the probability that $X_n > 0$ infinitely many times is zero, which yields $\lim_{n\to\infty} X_n = 0$ almost surely, *i.e.* with probability one.

Exercise 10.6

a) From the tower property of conditional expectations (1.6.8), we have:

$$\mathbb{E}[M_{n+1}] = \mathbb{E}[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]] \geqslant \mathbb{E}[M_n], \qquad n \geqslant 0.$$

b) If $(Z_n)_{n\geqslant 0}$ is a stochastic process with independent increments having nonnegative expectations, we have

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[Z_n \mid \mathcal{F}_n] + \mathbb{E}\left[Z_{n+1} - Z_n \mid \mathcal{F}_n\right]$$
$$= \mathbb{E}[Z_n \mid \mathcal{F}_n] + \mathbb{E}[Z_{n+1} - Z_n]$$

$$\geqslant \mathbb{E}[Z_n \mid \mathcal{F}_n] = Z_n, \qquad n \geqslant 0.$$

c) We define $(A_n)_{n\geqslant 0}$ by induction with $A_0 := 0$ and

$$A_{n+1} := A_n + \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n], \qquad n \geqslant 0,$$

and let

$$N_n := M_n - A_n, \qquad n \geqslant 0. \tag{S.63}$$

(i) For all $n \geqslant 0$, we have

$$\begin{aligned}
&\mathbb{E}[N_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[M_{n+1} - A_{n+1} \mid \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1} - A_n - \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] \mid \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1} - A_n \mid \mathcal{F}_n] - \mathbb{E}[\mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] \mid \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1} - A_n \mid \mathcal{F}_n] - \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] \\
&= -\mathbb{E}[A_n \mid \mathcal{F}_n] + \mathbb{E}[M_n \mid \mathcal{F}_n] \\
&= M_n - A_n \\
&= N_n,
\end{aligned}$$

hence $(N_n)_{n\geqslant 0}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geqslant 0}$.

(ii) For all $n \geqslant 0$, we have

$$\begin{aligned}
A_{n+1} - A_n &= \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] - \mathbb{E}[M_n \mid \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] - M_n \geqslant 0,
\end{aligned}$$

since $(M_n)_{n\geqslant 0}$ is a submartingale.

(iii) By induction we have $A_{n+1} = A_n + \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n]$, $n \geqslant 0$, which is $\mathcal{F}_n$-measurable if A_n is $\mathcal{F}_{n-1}$-measurable, $n \geqslant 1$.

(iv) This property is obtained by construction in (S.63).

d) For all bounded stopping times σ and τ such that $\sigma \leqslant \tau$ *a.s.*, we have $\mathbb{E}[A_\sigma] \leqslant \mathbb{E}[A_\tau]$, and $\mathbb{E}[N_\sigma] = \mathbb{E}[N_\tau]$ by (10.3.3), hence

$$\begin{aligned}
\mathbb{E}[M_\sigma] &= \mathbb{E}[N_\sigma] + \mathbb{E}[A_\sigma] \\
&= \mathbb{E}[N_\tau] + \mathbb{E}[A_\sigma] \\
&\leqslant \mathbb{E}[N_\tau] + \mathbb{E}[A_\tau] \\
&= \mathbb{E}[M_\tau],
\end{aligned}$$

by (10.3.3), since $(N_n)_{n\geqslant 0}$ is a martingale and $(A_n)_{n\geqslant 0}$ is non-decreasing.

Exercise 10.7

a) We will show more generally that

$$\phi(p_1x_1 + p_2x_2 + \cdots + p_nx_n) \leqslant p_1\phi(x_1) + p_2\phi(x_2) + \cdots + p_n\phi(x_n), \tag{S.64}$$

$x_1, \ldots, x_n \in \mathbb{R}$, for any sequence of coefficients $p_1, p_2, \ldots, p_n \geqslant 0$ such that $p_1 + p_2 + \cdots + p_n = 1$. The inequality (S.64) clearly holds for $n = 1$, and for $n = 2$ it coincides with the convexity property of ϕ, *i.e.*

$$\phi(p_1x_1 + p_2x_2) \leqslant p_1\phi(x_1) + p_2\phi(x_2), \qquad x_1, x_2 \in \mathbb{R}.$$

Assuming that (S.64) holds for some $n \geqslant 1$ and taking $p_1, p_2, \ldots, p_{n+1} \geqslant 0$ such that $p_1 + p_2 + \cdots + p_{n+1} = 1$ and $0 < p_{n+1} < 1$ and applying (10.5.2) at the second order, we have

$$\begin{aligned}
&\phi(p_1x_1 + p_2x_2 + \cdots + p_{n+1}x_{n+1}) \\
&= \phi\left((1-p_{n+1})\frac{p_1x_1 + p_2x_2 + \cdots + p_nx_n}{1-p_{n+1}} + p_{n+1}x_{n+1}\right) \\
&\leqslant (1-p_{n+1})\phi\left(\frac{p_1x_1 + p_2x_2 + \cdots + p_nx_n}{1-p_{n+1}}\right) + p_{n+1}\phi(x_{n+1}) \\
&\leqslant (1-p_{n+1})\left(\frac{p_1\phi(x_1) + p_2\phi(x_2) + \cdots + p_n\phi(x_n)}{1-p_{n+1}}\right) + p_{n+1}\phi(x_{n+1}) \\
&= p_1\phi(x_1) + p_2\phi(x_2) + \cdots + p_{n+1}\phi(x_{n+1}),
\end{aligned}$$

and we conclude by induction.

b) Taking $p_1 = p_1 = \cdots = p_N = 1/N$, we have

$$\begin{aligned}
&\mathbb{E}^*\left[\phi\left(\frac{S_1 + \cdots + S_N}{N}\right)\right] \leqslant \mathbb{E}^*\left[\frac{\phi(S_1) + \cdots + \phi(S_N)}{N}\right] && \text{since } \phi \text{ is convex,} \\
&= \frac{\mathbb{E}^*[\phi(S_1)] + \cdots + \mathbb{E}^*[\phi(S_N)]}{N} \\
&= \frac{\mathbb{E}^*[\phi(\mathbb{E}^*[S_N \mid \mathcal{F}_1])] + \cdots + \mathbb{E}^*[\phi(\mathbb{E}^*[S_N \mid \mathcal{F}_N])]}{N} && \text{because } (S_n)_{n\geqslant 0} \text{ is a martingale,} \\
&\leqslant \frac{\mathbb{E}^*[\mathbb{E}^*[\phi(S_N) \mid \mathcal{F}_1]] + \cdots + \mathbb{E}^*[\mathbb{E}^*[\phi(S_N) \mid \mathcal{F}_N]]}{N} && \text{by Jensen's inequality,} \\
&= \frac{\mathbb{E}^*[\phi(S_N)] + \cdots + \mathbb{E}^*[\phi(S_N)]}{N} && \text{by the tower property,} \\
&= \mathbb{E}^*[\phi(S_N)].
\end{aligned}$$

c) This is an application of the above bound to the convex payoff function $x \mapsto (x-K)^+$, see Figure S.29 and the R code below for an illustration.

```
nSim=99999;p=0.4;q=1-p;n=7;a=q/p;r=1;european=0;asian=0;K=1.5
dev.new(width=16,height=7)
for (j in 1:nSim){S<-a^cumsum(2*rbinom(n,1,p)-1);color="blue"
A<-sum(c(1,S))/(n+1);if (S[n]>=K) {european=european+S[n]-K}
if (A>=K) {asian=asian+A-K};if (S[n]>A) {color="darkred"} else {color="darkgreen"}
plot(seq(0,n),c(1,S), xlab = "Time", xlim=c(0,n), type='o',ylim = c(0,a^(n-2)), lwd = 3, ylab = "",
       col = color,main=paste("Asian Price=",format(round(asian,2)),"/",
       j,"=",format(round(asian/j,2)),"European Price=",format(round(european,2)),
       "/",j,"=",format(round(european/j,2))), xaxs='i',xaxt='n',yaxt='n', yaxs='i', yaxp =
       c(0,10,10))
text(3,6,paste("A-Payoff=",format(round(max(A-K,0),2))," E-Payoff=",
       format(round(max(S[n]-K,0),2))), col=color,cex=2)
axis(1, at=seq(0,n), labels=seq(0,n), las=1)
axis(2, at=c(0,K,A,1,2,3,4,5,6,7,8,9,10), labels=c(0,"K","Average",1,2,3,4,5,6,7,8,9,10), las=2)
lines(seq(0,n),rep(K,n+1),col = "red",lty = 1, lwd = 4);
lines(seq(0,n),rep(A,n+1),col = "darkgreen",lty = 2, lwd = 4); Sys.sleep(0.1)
if (S[n]>K || A>K) {readline(prompt = "Pause. Press <Enter> to continue...")}}
```

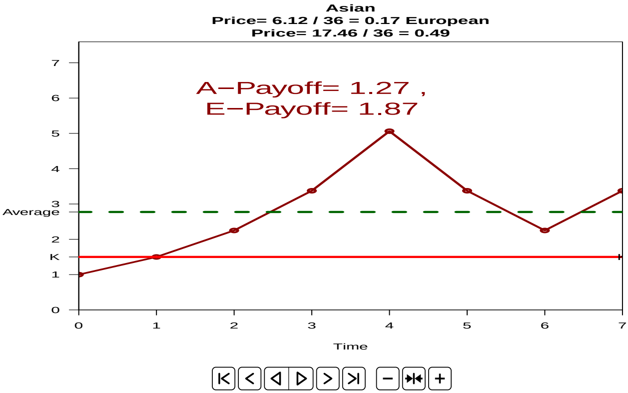


Fig. S.29: Asian option price *vs.* European option price.*

Exercise 10.8

a) We have

$$\mathbb{E}[M_n] \leqslant \mathbb{E}[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]] = \mathbb{E}[M_{n+1}], \qquad n \geqslant 0.$$

b) We write

$$S_n - \alpha n = \sum_{k=1}^{n} (X_k - p) + n(p - \alpha)$$

* The animation works in Acrobat Reader.

$$= \underbrace{(X_1 + X_2 + \cdots + X_n - np)}_{\text{martingale}} + n(p - \alpha)$$
$$= S_n - np + (p - \alpha)n$$

as the sum of a martingale (a stochastic process with centered independent increments) and $(p - \alpha)n$. As in the Doob-Meyer decomposition of Exercise 10.6-(c), we conclude that $(S_n)_{n \geqslant 0}$ is a *sub*martingale if and only if $p \geqslant \alpha$. Indeed, we have

$$\begin{aligned}
\mathbb{E}[S_n - \alpha n \mid \mathcal{F}_k] &= \mathbb{E}[X_1 + X_2 + \cdots + X_n - np \mid \mathcal{F}_k] + (p - \alpha)n \\
&= X_1 + X_2 + \cdots + X_k - kp + (p - \alpha)n \\
&= X_1 + X_2 + \cdots + X_k - k\alpha + (n - k)(p - \alpha) \\
&\geqslant S_k - k\alpha, \qquad k = 0, 1, \ldots, n,
\end{aligned}$$

if and only if $p \geqslant \alpha$.

Exercise 10.9

a) We have

$$\phi(M_k) = \phi\left(\mathbb{E}[M_n \mid \mathcal{F}_k]\right) \leqslant \mathbb{E}[\phi(M_n) \mid \mathcal{F}_k], \qquad k = 0, 1, \ldots, n$$

b) We have

$$\phi(M_k) \leqslant \phi\left(\mathbb{E}[M_n \mid \mathcal{F}_k]\right) \leqslant \mathbb{E}[\phi(M_n) \mid \mathcal{F}_k], \qquad k = 0, 1, \ldots, n.$$

Problem 10.10

a) We have

$$\{\tau_x > n\} = \bigcap_{k=0}^{n} \{M_k < x\}.$$

On the other hand, for all $k = 0, 1, \ldots, n$ we have $\{M_k < x\} \in \mathcal{F}_k \subset \mathcal{F}_n$, hence $\{\tau_x > n\} \in \mathcal{F}_n$ by the stability property of σ-algebras under intersection, cf. (1.1.1).

b) We have

$$\begin{aligned}
x\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x\Big) &= x\mathbb{P}(\tau_x \leqslant n) \\
&= x\mathbb{E}[\mathbb{1}_{\{\tau_x \leqslant n\}}] \\
&\leqslant \mathbb{E}\left[M_{\tau_x}\mathbb{1}_{\{\tau_x \leqslant n\}}\right] \\
&= \mathbb{E}\left[M_{\tau_x \wedge n}\mathbb{1}_{\{\tau_x \leqslant n\}}\right] \qquad \text{(S.65)} \\
&\leqslant \mathbb{E}[M_{\tau_x \wedge n}] \qquad \text{(S.66)} \\
&= \mathbb{E}[M_{\tau_x \wedge 0}]
\end{aligned}$$

$$= \mathbb{E}[M_0]$$
$$= \mathbb{E}[M_n], \qquad (S.67)$$

where we used the condition $M_{\tau_x \wedge n} \geqslant 0$ from (S.65) to (S.66), hence

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[M_n]}{x}, \qquad x > 0. \qquad (S.68)$$

Remark. The nonnegativity of $(M_n)_{n\geqslant 0}$ is used to reach (S.66), and the Doob Stopping Time Theorem 10.9 is used to derive (S.67).

c) When $(M_n)_{n\geqslant 0}$ is a *sub*martingale, by the Doob Stopping Time Theorem 10.9 for *sub*martingales we have

$$\mathbb{E}[M_{\tau_x \wedge n}] \leqslant \mathbb{E}[M_n],$$

see Exercise 10.6-(d), showing that (S.67) and (S.68) above still hold in this case.

d) Since $x \mapsto x^2$ is a convex function, $((M_n)^2)_{n\geqslant 0}$ is a *sub*martingale by Question (a), hence by Question (c) we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) = \mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} (M_k)^2 \geqslant x^2 \Big) \leqslant \frac{\mathbb{E}[(M_n)^2]}{x^2}, \qquad x > 0.$$

e) Similarly to Question (d), $x \mapsto x^p$ is a convex function for all $p \geqslant 1$ hence $((M_n)^p)_{n\geqslant 0}$ is a *sub*martingale by Question (a), and by Question (c) we find

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) = \mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} (M_k)^p \geqslant x^p \Big) \leqslant \frac{\mathbb{E}[(M_n)^p]}{x^p}, \qquad x > 0.$$

f) We note that $(S_n)_{n\geqslant 0}$ is a martingale because it has centered and independent increments, with

$$\mathbb{E}[(S_n)^2] = \operatorname{Var}[S_n] = n \operatorname{Var}[Y_1] = n\sigma^2,$$

hence by Question (d) we have

$$\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} S_k \geqslant x \Big) \leqslant \frac{\mathbb{E}[(S_n)^2]}{x^2} = \frac{n\sigma^2}{x^2}, \qquad x > 0.$$

g) When $(M_n)_{n\geqslant 0}$ is a (not necessarily nonnegative) *sub*martingale we can modify the answer to Question (b) using the Doob Stopping Time Theorem 10.9 for *sub*martingales, see Exercise 10.6-(d), as follows:

$$x\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x \Big) = x\mathbb{E}[\mathbb{1}_{\{\tau_x \leqslant n\}}]$$
$$\leqslant \mathbb{E}\left[M_{\tau_x \wedge n} \mathbb{1}_{\{\tau_x \leqslant n\}} \right]$$

$$\leqslant \mathbb{E}[(M_{\tau_x \wedge n})^+]$$
$$\leqslant \mathbb{E}[(M_n)^+],$$

since $((M_k)^+)_{k\in\mathbb{N}}$ is a *sub*martingale because $x \mapsto x^+$ is a non-decreasing convex function, *cf.* Exercise 10.9-(b), hence

$$\mathbb{P}\Big(\underset{k=0,1,\ldots,n}{\mathrm{Max}} M_k \geqslant x\Big) \leqslant \frac{\mathbb{E}[(M_n)^+]}{x}, \qquad x > 0.$$

h) We have

$$\begin{aligned} x\mathbb{P}\Big(\underset{k=0,1,\ldots,n}{\mathrm{Max}} M_k \geqslant x\Big) &= x\mathbb{P}(\tau_x \leqslant n) \\ &= x\mathbb{E}[\mathbb{1}_{\{\tau_x \leqslant n\}}] \\ &\leqslant \mathbb{E}\left[M_{\tau_x \wedge n}\mathbb{1}_{\{\tau_x \leqslant n\}}\right] \\ &\leqslant \mathbb{E}[M_{\tau_x \wedge n}] \\ &\leqslant \mathbb{E}[M_0]. \end{aligned}$$

i) We have

$$\begin{aligned} x\mathbb{P}\Big(\underset{k=0,1,\ldots,n}{\mathrm{Max}} \phi(M_k) \geqslant x\Big) &= x\mathbb{E}\left[\mathbb{1}_{\{\mathrm{Max}_{k=0,1,\ldots,n}\phi(M_k)\geqslant x\}}\right] \\ &= x\mathbb{E}\left[\mathbb{1}_{\{\tau_x^\phi \leqslant n\}}\right] \\ &= \mathbb{E}\left[\phi\left(M_{\tau_x^\phi \wedge n}\right)\mathbb{1}_{\{\tau_x^\phi \leqslant n\}}\right] \\ &= \mathbb{E}\big[\phi\big(M_{\tau_x^\phi \wedge n}\big)\big] \\ &\leqslant \mathbb{E}[\phi(M_n)], \end{aligned}$$

where the last inequality follows from Exercise 10.6-(d), since both $\tau_x^\phi \wedge n$ and n are stopping times.

j) Consider for example any nonnegative martingale such as $M_n = (p/q)^{S_n}$ where $S_n = X_1 + \cdots + X_n$ and $(X_k)_{k\geqslant 1}$ is a sequence of independent identically distributed Bernoulli random variables with $p = \mathbb{P}(X_k = 1)$ and $q = 1-p = \mathbb{P}(X_k = -1)$, $k \geqslant 1$. Then $Z_n := \mathrm{e}^{-n}M_n$ will be a *super*martingale, since

$$\begin{aligned} \mathbb{E}[Z_n \mid \mathcal{F}_k] &= \mathrm{e}^{-n}\mathbb{E}[M_n \mid \mathcal{F}_k] \\ &= \mathrm{e}^{-n}M_k \\ &\leqslant \mathrm{e}^{-k}M_k \\ &= Z_k, \qquad k = 0,1,\ldots,n. \end{aligned}$$

Exercise 10.11 We have

$$
\begin{aligned}
\mathbb{E}[(M_n)^r] &= \mathbb{E}\left[\left(\frac{q}{p}\right)^{rS_n}\right] \\
&= \mathbb{E}\left[\prod_{k=1}^n \left(\frac{q}{p}\right)^{r(S_k-S_{k-1})}\right] \\
&= \prod_{k=1}^n \mathbb{E}\left[\left(\frac{q}{p}\right)^{r(S_k-S_{k-1})}\right] \\
&= \prod_{k=1}^n \left(p\left(\frac{q}{p}\right)^r + q\left(\frac{q}{p}\right)^{-r}\right) \\
&= \left(p\left(\frac{q}{p}\right)^r + q\left(\frac{p}{q}\right)^r\right)^n \\
&= \left(\frac{pq^{2r}+qp^{2r}}{(pq)^r}\right)^n \\
&= \big(\mathbb{E}[(M_1)^r]\big)^n.
\end{aligned}
$$

We note that by Jensen's inequality we have

$$
\mathbb{E}[(M_1)^r] \geqslant \big(\mathbb{E}[M_1]\big)^r = 1,
$$

and $\mathbb{E}[(M_n)^r]$ is non-decreasing in $n \geqslant 0$ as $((M_n)^r)_{n\geqslant 0}$ is a *sub*martingale. By Problem 10.10-(e), for all $n \geqslant 0$ and $r \geqslant 1$ we have

$$
\begin{aligned}
\mathbb{P}\Big(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x\Big) &\leqslant \frac{\mathbb{E}[(M_n)^r]}{x^r} \\
&= \frac{(p(q/p)^r + q(p/q)^r)^n}{x^r}, \qquad x > 0. \qquad \text{(S.69)}
\end{aligned}
$$

By (3.3.3), we also check that

$$
\begin{aligned}
\mathbb{E}[(M_{2n})^r] &= \mathbb{E}\left[\left(\frac{q}{p}\right)^{rS_{2n}}\right] \\
&= \sum_{k=-n}^n \left(\frac{q}{p}\right)^{2kr} \mathbb{P}(S_{2n} = 2k) \\
&= \sum_{k=-n}^n \binom{2n}{n+k}\left(\frac{q}{p}\right)^{2kr} p^{n+k}q^{n-k} \\
&= \sum_{k=0}^{2n} \binom{2n}{k}\left(\frac{q}{p}\right)^{2(k-n)r} p^k q^{2n-k}
\end{aligned}
$$

$$= \left(\frac{q}{p}\right)^{-2nr} q^{2n} \sum_{k=0}^{2n} \binom{2n}{k} \left(\left(\frac{q}{p}\right)^{2r-1}\right)^k$$
$$= \left(p\left(\frac{q}{p}\right)^r + q\left(\frac{p}{q}\right)^r\right)^{2n}.$$

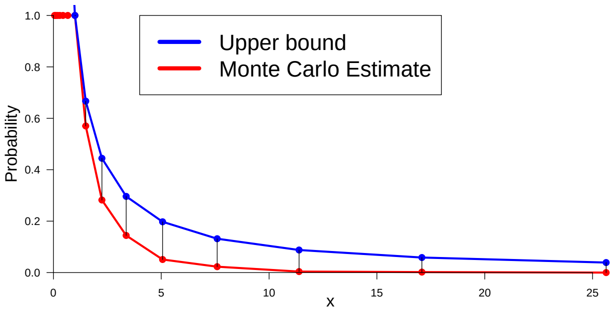


Fig. S.30: Supremum deviation probability with $n = 7$ and $p = 0.4$.

Note also that when $r = 1$ we find

$$\mathbb{P}\left(\operatorname*{Max}_{k=0,1,\ldots,n} M_k \geqslant x\right) \leqslant \frac{\mathbb{E}[M_1]}{x} = \frac{1}{x}, \qquad x > 0.$$

The following R code together with Figures S.30-S.31 provide a numerical confirmation of the upper bound (S.69) when $r = 1$.

```
nSim=99999;p=0.4;q=1-p;n=7;a=q/p;r=1;prob=rep(0,2*n+2)
par(mar=c(2,3,2,2));par(mgp=c(1,1,0)); for (i in (-n):(n+1)){for (j in 1:nSim){
M<-a^cumsum(2*rbinom(n,1,p)-1);color="blue"
if (max(c(1,M))>=a^i) {prob[n+1+i]=prob[n+1+i]+1;color="darkred"}
if ((j%%10000)==0){
plot(seq(0,n),c(1,M), xlab = "Time", xlim=c(0,n), type='o',ylim = c(0,a^(n-1)), lwd
        = 3, ylab = "", col = color, main='',xaxs='i',yaxs='i',yaxp = c(0, 11, 11), las
        =1)
text(3,8,paste("x=",format(round(a^i,2)),",
        ",prob[n+1+i],"/",j,"=",format(prob[n+1+i]/j,digits=4)),col=color,cex=2)
axis(2, at=c(a^i), labels=c("x"), las=1)
lines(seq(0,n),rep(a^i,n+1),col = "black",lty = 2, lwd = 2); text(-0.1,a^i,
        paste("x"));Sys.sleep(0.9)}}
prob[n+1+i]=prob[n+1+i]/nSim};x=a^seq(-n,n+1);
        par(mar=c(2,3,2,2));par(mgp=c(1,1,0))
plot(x,prob,type="o",pch=19,lwd=3,col="red",xlab="x",ylim=c(0,1),
        ylab="Probability",main="",axes=FALSE,cex.axis=1.5, cex.lab=1.5)
lines(x,(p*a^r+q*(p/q)^r)^n/x^r,type="o",pch=19,lwd=3,col="blue",main="")
axis(1, pos=0,las =1);axis(2, pos=0,las =1);
legend(4, 1, legend=c("Upper bound", "Monte Carlo estimate"), col=c("blue","red"),
        lty=1, lwd=6,cex=2)
for (i in 0:(n+1)){segments(x0=a^i, y0=prob[n+1+i],x1=a^i,
        (p*a^r+q*a^(-r))^n/a^(i*r),col="black")}
```

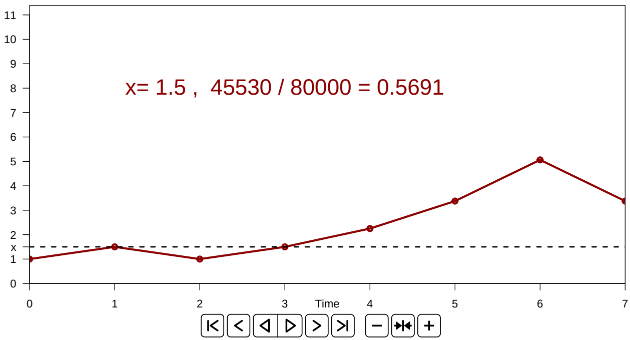


Fig. S.31: Martingale supremum as a function of time.*

Chapter 11 - Spatial Poisson Processes

Exercise 11.1 We note that N_t can be constructed as $N_t(\omega) = \omega([0, t])$, and by (11.1.1) we find that given $\{N_T = k\}$, the distribution of $(T_1, T_2, \ldots, T_k)$

* The animation works in Acrobat Reader.

is uniform on

$$\{(t_1, t_2, \ldots, t_k) \ : \ 0 \leqslant t_1 \leqslant \cdots \leqslant t_k \leqslant T\}.$$

This yields

$$\mathbb{E}[T_1 \mid N_T = 2] = \frac{2}{T^2} \int_0^T \int_0^y x dx dy = \frac{1}{T^2} \int_0^T y^2 dy = \frac{T}{3},$$

and

$$\mathbb{E}[T_2 \mid N_T = 2] = \frac{2}{T^2} \int_0^T y \int_0^y dx dy = \frac{2}{T^2} \int_0^T y^2 dy = \frac{2T}{3},$$

and we check that

$$\mathbb{E}[T_1 + T_2 \mid N_T = 2] = T = \frac{2}{T} \int_0^T x dx,$$

which is consistent with the fact that the (unordered) locations of the two jump times are uniformly distributed on $[0, T]$ given that $N_T = 2$, see also (11.1.1) page 388.

On the other hand, by the memoryless property of the standard Poisson process we have

$$\mathbb{E}[T_3 \mid N_T = 2] = T + \mathbb{E}[T_1 \mid N_T = 0] = T + \frac{1}{\lambda}.$$

Exercise 11.2 The probability that there are 10 events within a circle of radius 3 meters is

$$\mathrm{e}^{-9\pi\lambda} \frac{(9\pi\lambda)^{10}}{10!} = \mathrm{e}^{-9\pi/2} \frac{(9\pi/2)^{10}}{10!} \simeq 0.0637.$$

Exercise 11.3 The probability that more than two living organisms are in this measured volume is

$$\begin{aligned}\mathbb{P}(N \geqslant 3) = 1 - \mathbb{P}(N \leqslant 2) &= 1 - \mathrm{e}^{-10\theta} \left(1 + 10\theta + \frac{(10\theta)^2}{2}\right) \\ &= 1 - \mathrm{e}^{-6} \left(1 + 6 + \frac{6^2}{2}\right) = 1 - 25\,\mathrm{e}^{-6} \simeq 0.938.\end{aligned}$$

Exercise 11.4 Let X_A, resp. X_B, the number of defects found by the first, resp. second, inspection. We know that X_A and X_B are independent Poisson random variables with intensities 0.5, hence the probability that both inspections yield defects is

$$\begin{aligned}\mathbb{P}(X_A \geqslant 1 \text{ and } X_B \geqslant 1) &= \mathbb{P}(X_A \geqslant 1)\mathbb{P}(X_B \geqslant 1) \\ &= (1 - \mathbb{P}(X_A = 0))(1 - \mathbb{P}(X_B = 0))\end{aligned}$$

$$= (1 - \mathrm{e}^{-0.5})^2 \simeq 0.2212.$$

Exercise 11.5 The number X_N of points in the interval $[0, \lambda]$ has a binomial distribution with parameter $(N, \lambda/N)$, *i.e.*

$$\mathbb{P}(X_N = k) = \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{N-k}, \qquad k = 0, 1, \ldots, N,$$

and we find

$$\lim_{N\to\infty} \mathbb{P}(X_N = k) = \frac{\lambda^k}{k!} \lim_{N\to\infty} \left(\left(1 - \frac{\lambda}{N}\right)^{N-k} \prod_{i=0}^{k-1} \frac{N-i}{N} \right) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!},$$

which is the Poisson distribution with parameter $\lambda > 0$.

Exercise 11.6

a) Based on the area $\pi r^2 = 9\pi$, this probability is given by

$$\mathrm{e}^{-9\pi/2} \frac{(9\pi/2)^{10}}{10!}.$$

b) This probability is

$$\mathrm{e}^{-9\pi/2} \frac{(9\pi/2)^5}{5!} \times \mathrm{e}^{-9\pi/2} \frac{(9\pi/2)^3}{3!}.$$

c) This probability is

$$\mathrm{e}^{-9\pi} \frac{(9\pi)^8}{8!}.$$

d) Since the location of points are uniformly distributed by (11.1.1), the probability that a point in the disk $D((0,0),1)$ is located in the subdisk $D((1/2,0),1/2)$ is given by the ratio $\pi/4/\pi = 1/4$ of their surfaces. Hence, given that 5 items are found in $D((0,0),1)$, the number of points located within $D((1/2,0),1/2)$ has a binomial distribution with parameter $(5, 1/4)$, cf. (S.10) in the solution of Exercise 1.7 and Exercise 9.2-(d), and we find the probability

$$\binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 = \frac{45}{512} \simeq 0.08789.$$

Exercise 11.7 (Wang et al. (2012)) By the moment identity Relation (11.4.1) we have

$$\mathbb{E}\left[\left| \frac{S_n - \lambda n}{\sqrt{n}} \right|^p \right] = n^{-p/2} \sum_{k=0}^{p} (n\lambda)^k S_2(p,k) = n^{-p/2} \sum_{k=0}^{p/2} (n\lambda)^k S_2(p,k)$$

$$\begin{aligned} &= n^{-p/2} \sum_{k=0}^{p/2} (n\lambda)^k p^k = n^{-p/2} \frac{(np\lambda)^{1+p/2} - 1}{np\lambda - 1} \\ &\leqslant \frac{(p\lambda)^{1+p/2}}{p\lambda - 1/n} < C_p, \end{aligned}$$

where $C_p > 0$ is a finite constant.

Exercise 11.8

a) We have

$$\begin{aligned} M'(s) &= \int_X f(x)(\mathrm{e}^{sf(x)} - 1)\sigma(dx)\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\exp\left(s\int_0^\infty f(y)(dN_y - dy)\right)\right] \\ &= s\int_X |f(x)|^2 \frac{\mathrm{e}^{sf(x)} - 1}{sf(x)}\sigma(dx)\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\exp\left(s\int_0^\infty f(y)(dN_y - dy)\right)\right] \\ &\leqslant \frac{\mathrm{e}^{sK} - 1}{K}\int_X |f(x)|^2\sigma(dx)\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\exp\left(s\int_0^\infty f(y)(dN_y - dy)\right)\right] \\ &= \alpha^2\frac{\mathrm{e}^{sK} - 1}{K}\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\exp\left(s\int_0^\infty f(y)(dN_y - dy)\right)\right] = \alpha^2\frac{\mathrm{e}^{sK} - 1}{K}M(s), \end{aligned}$$

which shows that

$$\frac{M'(s)}{M(s)} \leqslant h(s) := \alpha^2\frac{\mathrm{e}^{sK} - 1}{K}, \qquad s \geqslant 0.$$

b) We have

$$\begin{aligned} \log M(t) &= \log M(0) + \int_0^t d\log M(s) \\ &\leqslant \int_0^t \frac{M'(s)}{M(s)}ds \\ &\leqslant \int_0^t h(s)ds, \end{aligned}$$

hence

$$M(t) \leqslant \exp\left(\int_0^t h(s)ds\right) = \exp\left(\alpha^2\int_0^t \frac{\mathrm{e}^{sK} - 1}{K}ds\right), \qquad t \geqslant 0.$$

c) By Markov's inequality, we have

$$\begin{aligned} \mathbb{P}_\sigma^X\left(\int_0^\infty f(y)(dN_y - dy) \geqslant x\right) &= \mathbb{E}_{\mathbb{P}_\sigma^X}\left[\mathbb{1}_{\{\int_0^\infty f(y)(dN_y - dy) \geqslant x\}}\right] \\ &\leqslant \mathrm{e}^{-tx}\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\mathbb{1}_{\{\int_0^\infty f(y)(dN_y - dy) \geqslant x\}}\exp\left(t\int_0^\infty f(y)dN_y\right)\right] \\ &\leqslant \mathrm{e}^{-tx}\mathbb{E}_{\mathbb{P}_\sigma^X}\left[\exp\left(t\int_0^\infty f(y)dN_y\right)\right] \end{aligned}$$

$$
\begin{aligned}
&\leqslant \exp\left(-tx + \int_0^t h(s)ds\right)\\
&\leqslant \exp\left(-tx + \alpha^2 \int_0^t \frac{\mathrm{e}^{sK}-1}{K} ds\right)\\
&= \exp\left(-tx + \frac{\alpha^2}{K^2}(\mathrm{e}^{tK} - tK - 1)\right),
\end{aligned}
$$

which also yields

$$
\mathbb{P}\left(\int_0^\infty f(y)(dN_y - dy) \geqslant x\right) \leqslant \exp\left(-tx + \alpha^2 \int_0^t \frac{e^{sK}-1}{K} ds\right).
$$

d) Minimizing the above term in t with $t = K^{-1}\log(1 + Kx/\alpha^2)$ shows that

$$
\begin{aligned}
\mathbb{P}_\sigma^X\left(\int_0^\infty f(y)(dN_y - dy) \geqslant x\right) &\leqslant \exp\left(\frac{x}{K} - \left(\frac{x}{K} + \frac{\alpha^2}{K^2}\right)\log\left(1 + \frac{xK}{\alpha^2}\right)\right)\\
&\leqslant \exp\left(-\frac{x}{2K}\log\left(1 + \frac{xK}{\alpha^2}\right)\right)\\
&= \left(1 + \frac{xK}{\alpha^2}\right)^{-x/2K}.
\end{aligned}
$$

Exercise 11.9

a) We have

$$
\begin{aligned}
\mathcal{G}_\Phi(f) &= \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n=0}^\infty \frac{1}{n!} \int_{\mathbb{X}^n} \prod_{i=1}^n f(x_i)\sigma(dx_1)\cdots\sigma(dx_n)\\
&= \mathrm{e}^{-\sigma(\mathbb{X})} \sum_{n=0}^\infty \frac{1}{n!} \left(\int_{\mathbb{X}} f(x_1)\sigma(dx_1)\right)^n\\
&= \exp\left(\int_{\mathbb{X}} f(x)\sigma(\mathrm{d}x) - \sigma(\mathbb{X})\right)\\
&= \exp\left(\int_{\mathbb{X}} (f(x)-1)\sigma(dx)\right), \qquad f \in L^1(\mathbb{X}, \mu).
\end{aligned}
$$

b) We have

$$
\begin{aligned}
\mathbb{P}(\Phi \cap A = \emptyset) &= \mathbb{E}\left[\mathbf{1}_{\{\Phi\cap A=\emptyset\}}\right]\\
&= \mathbb{E}\left[\prod_{x\in\Phi} \mathbf{1}_{A^c}(x)\right]\\
&= \mathcal{G}_\Phi(\mathbf{1}_{A^c})\\
&= \exp\left(\int_{\mathbb{X}} (\mathbf{1}_{A^c}(x) - 1)\sigma(dx)\right)
\end{aligned}
$$

$$= \exp\left(-\int_{\mathbb{X}} \mathbf{1}_A(x)\sigma(dx)\right)$$
$$= \mathrm{e}^{-\sigma(A)}.$$

c) Letting

$$\mathcal{C} := \{(x,r) \in \mathbb{R}^d \times \mathbb{R}_+ \ : \ \|x\| \leqslant r\},$$

we have

$$\begin{aligned}
\mathbb{P}(0 \in \mathcal{B}) &= 1 - \mathbb{P}(0 \notin \mathcal{B}) \\
&= 1 - \mathbb{P}(\Phi \cap \mathcal{C} = \emptyset) \\
&= 1 - \mathcal{G}_\Phi(\mathbf{1}_{\mathcal{C}^c}) \\
&= 1 - \exp\left(-\int_{\mathbb{R}^d \times \mathbb{R}_+} \mathbf{1}_{\mathcal{C}}(x)\sigma(dx)\right) \\
&= 1 - \exp\left(-\lambda \int_{\mathbb{R}^d \times \mathbb{R}_+} \mathbf{1}_{\{y \in B(0,r)\}} dy \rho(r) dr\right) \\
&= 1 - \exp\left(-\lambda \int_{\mathbb{R}^d} \rho(r) \int_0^\infty \mathbf{1}_{\{y \in B(0,r)\}} dy dr\right) \\
&= 1 - \exp\left(-\lambda v_d \int_{\mathbb{R}^d} \rho(r) r^d dr\right).
\end{aligned}$$

d) By translation invariance, this probability is given by

$$\mathbb{P}(0 \in \mathcal{B}) = 1 - \exp\left(-\lambda v_d \int_{\mathbb{R}^d} \rho(r) r^d dr\right).$$

Exercise 11.10

a) This probability is given by

$$\exp\left(-\int_{[0,1]^d} \sigma(dy) \int_0^{1/2} \mathrm{e}^{-r} dr\right) = \mathrm{e}^{-\sigma([0,1]^d)(1-\mathrm{e}^{-1/2})}.$$

b) The mean is given by

$$\int_{[0,1]^d} \sigma(dy) \int_0^{1/2} \mathrm{e}^{-r} dr = \sigma([0,1]^d)(1 - \mathrm{e}^{-1/2}).$$

Chapter 12 - Reliability and Renewal Processes

Exercise 12.1

a) We have

$$F_\beta(t) = \mathbb{P}(\tau < t) = \int_0^t f_\beta(x) dx = \beta \int_0^t x^{\beta-1} \mathrm{e}^{-x^\beta} dx = -\left[\mathrm{e}^{-x^\beta}\right]_0^t = 1 - \mathrm{e}^{-t^\beta},$$

$t \geqslant 0$.

b) We have

$$R(t) = \mathbb{P}(\tau > t) = 1 - F_\beta(t) = \mathrm{e}^{-t^\beta}, \qquad t \geqslant 0.$$

c) We have

$$\lambda(t) = -\frac{d}{dt}\log R(t) = \beta t^{\beta-1} \qquad t \geqslant 0.$$

d) By (12.3.1) we have

$$\mathbb{E}[\tau] = \int_0^\infty R(t)dt = \int_0^\infty \mathrm{e}^{-t^\beta}dt.$$

In particular this yields $\mathbb{E}[\tau] = \sqrt{\pi}/2$ when $\beta = 2$.

References

[1] Agapie, A. and Höns, R. (2007). Analysis of a voter model. *Math. Rep. (Bucur.)*, 9(59)(2):135–145. (Cited on page 270.)

[2] Aldous, D. and Diaconis, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348. (Cited on pages 254 and 261.)

[3] Althoen, S., King, L., and Schilling, K. (1993). How long is a game of Snakes and Ladders? *The Mathematical Gazette*, 77(478):71–76. (Cited on page 178.)

[4] Anantharam, V., Varaiya, P., and Walrand, J. (1987). Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays - Part II: Markovian rewards. *IEEE Transactions on Automatic Control*, 32(11):977–982. (Cited on page 612.)

[5] Anděl, J. and Hudecová, Š. (2012). Variance of the game duration in the gambler's ruin problem. *Statist. Probab. Lett.*, 82(9):1750–1754. (Cited on page 83.)

[6] Antal, T. and Redner, S. (2005). The excited random walk in one dimension. *J. Phys. A*, 38(12):2555–2577. (Cited on pages 108, 109, 115, 118, 225, 466, 468, and 484.)

[7] Asmussen, S. (2003). *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition. Stochastic Modelling and Applied Probability. (Cited on page 219.)

[8] Benjamini, I. and Wilson, D. (2003). Excited random walk. *Electron. Comm. Probab.*, 8:86–92. (Cited on page 115.)

[9] Bhattacharya, B. and Mukherjee, S. (2018). Inference in Ising models. *Bernoulli*, 24(1):493–525. (Cited on page 641.)

[10] Borodin, A. (2017). *Stochastic processes.* Probability and its Applications. Birkhäuser/Springer, Cham. Original Russian edition published by LAN Publishing, St. Petersburg, 2013. (Cited on page 466.)

[11] Bosq, D. and Nguyen, H. (1996). *A Course in Stochastic Processes: Stochastic Models and Statistical Inference.* Mathematical and Statistical Methods. Kluwer. (Cited on pages v, vi, 237, 307, and 694.)

[12] Bouneffouf, D., Rish, I., and Aggarwal, C. (2020). A survey on practical applications of multi-armed and contextual bandits. In *IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. (Cited on page 259.)

[13] Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122. (Cited on page 614.)

[14] Champion, W., Mills, T., and Smith, S. (2007). Lost in space. *Math. Scientist*, 32:88–96. (Cited on page 476.)

[15] Chen, M.-F. (2004). *From Markov chains to non-equilibrium particle systems.* World Scientific Publishing Co., second edition. (Cited on page 196.)

[16] Chewi, S. (2017). Wald's identity. https://inst.eecs.berkeley.edu/~ee126/fa17/wald.pdf. Accessed: 2022-08-25. (Cited on pages 258 and 611.)

[17] Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall Inc., Englewood Cliffs, N.J. (Cited on pages v and vi.)

[18] Çınlar, E. (2011). *Probability and stochastics*, volume 261 of *Graduate Texts in Mathematics*. Springer, New York. (Cited on page 19.)

[19] Decreusefond, L., Ferraz, E., and Martins, P. (2009). Upper bound of loss probability for the dimensioning of OFDMA systems with multi class randomly located users. In *WiOpt, workshop SPASWIN 2009. Seoul, South Korea*. (Cited on page 399.)

[20] Devore, J. L. (2003). *Probability and Statistics for Engineering and the Sciences*. Duxbury Press, sixth edition edition. (Cited on page 9.)

[21] Doob, J. (1953). *Stochastic processes*. John Wiley & Sons Inc., New York. (Cited on page 385.)

[22] Doob, J. (1984). *Classical potential theory and its probabilistic counterpart*. Springer-Verlag, Berlin. (Cited on page 385.)

[23] Durrett, R. (1999). *Essentials of Stochastic Processes*. Springer. (Cited on page v.)

[24] Finner, H., Kern, P., and Scheer, M. (2015). On some compound distributions with Borel summands. *Insurance Math. Econom.*, 62:234–244. (Cited on page 666.)

[25] Foucart, S. (2010). Linear algebra and matrix analysis, Lecture 6. https://www.math.drexel.edu/~foucart/TeachingFiles/F12/M504Lect6.pdf. Accessed: 2022-08-25. (Cited on pages 257, 276, 277, and 607.)

[26] Freedman, D. (1983). *Markov chains*. Springer-Verlag, New York-Berlin. (Cited on pages 258 and 611.)

[27] Goldberg, S. (1986). *Introduction to difference equations*. Dover Publications, Inc., New York, second edition. With illustrative examples from economics, psychology, and sociology. (Cited on page 158.)

[28] Grimmett, G. R. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press, Oxford. (Cited on page v.)

[29] Gusev, V. (2014). Synchronizing automata with random inputs. In *Developments in Language Theory*, volume 8633 of *Lecture Notes in Computer Science*, pages 68–75. Springer, New York. (Cited on pages 198, 251, and 554.)

[30] Haight, F. and Breuer, M. (1960). The Borel-Tanner distribution. *Biometrika*, 47:143–150. (Cited on page 666.)

[31] Houdré, C. and Privault, N. (2002). Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli*, 8(6):697–720. (Cited on page 399.)

[32] Iribarren, J. and Moro, E. (2011). Branching dynamics of viral information spreading. *Phys. Rev. E*, 84:046116, 13. (Cited on page 283.)

[33] Jacod, J. and Protter, P. (2000). *Probability essentials*. Springer-Verlag, Berlin. (Cited on pages 9 and 425.)

[34] Johnson, N., Kemp, A., and Kotz, S. (2005). *Univariate discrete distributions*. Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, NJ, third edition. (Cited on page 666.)

[35] Jonasson, J. (2009). The mathematics of card shuffling. http://www.math.chalmers.se/~jonasson/convrates.pdf. Accessed: 2019-10-10. (Cited on pages 254 and 261.)

[36] Jones, P. and Smith, P. (2001). *Stochastic Processes: An Introduction*. Arnold Texts in Statistics. Hodder Arnold. (Cited on pages v and vi.)

[37] Kallenberg, O. (2002). *Foundations of Modern Probability*. Probability and its Applications. Springer-Verlag, New York, second edition. (Cited on page 47.)

[38] Karlin, S. and Taylor, H. (1981). *A Second Course in Stochastic Processes*. Academic Press Inc., New York. (Cited on pages 158, 169, and 338.)

[39] Karlin, S. and Taylor, H. (1998). *An introduction to stochastic modeling*. Academic Press, Inc., San Diego, CA, third edition. (Cited on pages v, vi, 228, 229, 233, 288, and 408.)
[40] Kato, T. (1995). *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer-Verlag, Berlin. Reprint of the 1980 edition. (Cited on pages 276, 277, and 644.)
[41] Kijima, M. (1997). *Markov processes for stochastic modeling*. Stochastic Modeling Series. Chapman & Hall, London. (Cited on page 219.)
[42] Laforgia, A. (1991). Bounds for modified Bessel functions. *J. Comput. Appl. Math.*, 34(3):263–267. (Cited on page 693.)
[43] Lalley, S. (2013). Continuous time Markov chains. https://galton.uchicago.edu/~lalley/Courses/313/ContinuousTime.pdf. Accessed: 2015-12-10. (Cited on page 337.)
[44] Léon, C. and Perron, F. (2004). Optimal Hoeffding bounds for discrete reversible Markov chains. *The Annals of Applied Probability*, pages 958–970. (Cited on page 608.)
[45] Levin, D., Peres, Y., and Wilmer, E. (2009). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI. With a chapter by James G. Propp and David B. Wilson. (Cited on pages 229, 268, 631, and 636.)
[46] Lezaud, P. (1998). Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867. (Cited on pages 274 and 641.)
[47] Lukacs, E. (1970). *Characteristic Functions*. Hafner Publishing Co., New York. Second edition, revised and enlarged. (Cited on page 43.)
[48] Markov, A. A. (1909). Recherches sur un cas remarquable d'épreuves dépendantes. *Acta Math.*, 33:87–104. (Cited on pages 2 and 119.)
[49] Mecke, J. (1967). Stationäre zufällige Masse auf lokalkompakten Abelschen Gruppen. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, 9:36–58. (Cited on page 396.)
[50] Medhi, J. (2010). *Stochastic Processes*. New Age Science Limited, Tunbridge Wells, UK, third edition. (Cited on pages v and vi.)
[51] MOE and UCLES (2020). Mathematics (Syllabus 9758), Singapore-Cambridge General Certificate of Education Advanced Level Higher 2 (2022). https://www.seab.gov.sg/docs/default-source/national-examinations/syllabus/alevel/2022syllabus/9758_y22_sy.pdf. Accessed: 2022-12-28. (Cited on page 17.)
[52] MOE and UCLES (2022). Mathematics (Syllabus 4048), Singapore-Cambridge General Certificate of Education Ordinary Level (2022). https://www.seab.gov.sg/docs/default-source/national-examinations/syllabus/olevel/2022syllabus/4048_y22_sy.pdf. Accessed: 2024-04-20. (Cited on page 17.)
[53] Muni Toke, I. (2015). The order book as a queueing system: average depth and influence of the size of limit orders. *Quantitative Finance*, 15(5):795–808. (Cited on pages 353 and 675.)
[54] Neveu, J. (1977). Processus ponctuels. In *Ecole d'été de Probabilités de Saint-Flour VI*, volume 598 of *Lecture Notes in Mathematics*, pages 249–445. Springer-Verlag. (Cited on page 389.)
[55] Norris, J. (1998). *Markov Chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. Reprint of 1997 original. (Cited on pages v, 136, 239, and 310.)
[56] Olaleye, T., Sowunmi, F., Abiola, S., Salako, M., and Eleyoowo, I. (2009). A Markov chain approach to the dynamics of vehicular traffic characteristics in Abeokuta metropolis. *Research Journals of Applied Sciences, Engineering and Technology*, 1(3):160–166. (Cited on page 147.)
[57] Pitman, J. (1999). *Probability*. Springer. (Cited on page 9.)
[58] Pólya, G. (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassenetz. *Mathematische Annalen*, 84:149–160. (Cited on page 476.)
[59] Pólya, G. and Szegö, G. (1998). *Problems and Theorems in Analysis I*. Springer. Reprint of the 1978 Edition. (Cited on page 666.)

[60] Privault, N. (2008). Potential theory in classical probability. In Franz, U. and Schürmann, M., editors, *Quantum potential theory*, volume 1954 of *Lecture Notes in Math.*, pages 3–59. Springer, Berlin. (Cited on pages 159 and 167.)
[61] Privault, N. (2009). *Stochastic analysis in discrete and continuous settings: with normal martingales*, volume 1982 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. (Cited on pages 390, 399, and 696.)
[62] Privault, N. (2011). Generalized Bell polynomials and the combinatorics of Poisson central moments. *Electron. J. Combin.*, 18(1):Research Paper 54, 10. (Cited on page 396.)
[63] Privault, N. (2012). Moments of Poisson stochastic integrals with random integrands. *Probab. Math. Stat.*, 32(2):227–239. (Cited on page 397.)
[64] Privault, N. (2023). Notes on Stochastic Finance. Course notes, https://personal.ntu.edu.sg/nprivault/MA5182/introduction2.pdf, 1265 pages. Accessed: 2023-08-30. (Cited on pages 108 and 109.)
[65] Redner, S. (2001). *A guide to first-passage processes*. Cambridge University Press, Cambridge. (Cited on pages 108, 109, 466, and 468.)
[66] Ross, S. (1996). *Stochastic Processes*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. (Cited on pages v and vi.)
[67] Schalekamp, F. and van Zuylen, A. (2009). Rank aggregation: Together we're strong. In *2009 Proceedings of the Eleventh Workshop on Algorithm Engineering and Experiments (ALENEX09)*, pages 38–51. SIAM. (Cited on page 267.)
[68] Schütz, G. and Trimper, S. (2008). Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk. *Phys. Rev. E (3)*. (Cited on page 145.)
[69] Sellke, S., Shroff, N., and Bagchi, S. (2008). Modeling and automated containment of worms. *IEEE Transactions on dependable and secure computing*, 5(2):71–86. (Cited on page 283.)
[70] Serfozo, R. (2009). *Basics of applied stochastic processes*. Probability and its Applications (New York). Springer-Verlag, Berlin. (Cited on pages 258, 604, and 611.)
[71] Sigman, K. (2014). Continuous-time Markov chains. https://www.columbia.edu/~ks20/stochastic-I/stochastic-I-CTMC.pdf. Accessed: 2014-12-09. (Cited on page 337.)
[72] Slivnyak, I. (1962). Some properties of stationary flows of homogeneous random events. *Theory Probab. Appl.*, 7(3):336–341. (Cited on page 396.)
[73] Steele, J. (2001). *Stochastic Calculus and Financial Applications*, volume 45 of *Applications of Mathematics*. Springer-Verlag, New York. (Cited on pages v and 378.)
[74] Stroock, D. (2011). *Probability theory, an analytic view*. Cambridge University Press, Cambridge, second edition. (Cited on page 46.)
[75] Vinay, S. and Kok, P. (2019). Statistical analysis of quantum-entangled-network generation. *Phys. Rev. E*, 99:042313. (Cited on page 191.)
[76] Wang, Z., Stocker, A., and Lee, D. (2012). Optimal neural tuning curves for arbitrary stimulus distributions: Discrimax, infomax and minimum l^p loss. In *Advances in Neural Information Processing Systems (NIPS 2012)*, volume 25, pages 1–9. (Cited on page 712.)
[77] Williams, R. J. (1992). Asymptotic variance parameters for the boundary local times of reflected Brownian motion on a compact interval. *J. Appl. Probab.*, 29(4):996–1002. (Cited on pages 466 and 467.)
[78] Wolfer, G. and Kontorovich, A. (2021). Statistical estimation of ergodic Markov chain kernel over discrete state space. *Bernoulli*, 27(1):532–553. (Cited on pages 277 and 648.)
[79] Wu, L. (2000). A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields*, 118(3):427–438. (Cited on page 399.)

Index

Author index

This book provides an undergraduate introduction to discrete and continuous-time Markov chains and their applications. A large focus is placed on the first step analysis technique and its applications to average hitting times and ruin probabilities. Classical topics such as recurrence and transience, stationary and limiting distributions, as well as branching processes, are also covered. Two major examples (gambling processes and random walks) are treated in detail from the beginning, before the general theory itself is presented in the subsequent chapters. An introduction to discrete-time martingales and their relation to ruin probabilities and mean exit times is also provided, and the book includes a chapter on spatial Poisson processes with some recent results on moment identities and deviation inequalities for Poisson stochastic integrals. The concepts presented are illustrated by examples, together with 172 exercises, 40 problems, and their complete solutions.