ANDERS TOLVER

AN INTRODUCTION TO MARKOV CHAINS

LECTURE NOTES FOR STOCHASTIC PROCESSES

Anders Tolver
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø, Denmark

email: tolver@math.ku.dk

First printing, November 2016

ISBN: 978-87-7078-952-3

Preface

These lecture notes have been developed for the course *Stochastic Processes* at Department of Mathematical Sciences, University of Copenhagen during the teaching years 2010-2016. The material covers aspects of the theory for time-homogeneous Markov chains in discrete and continuous time on finite or countable state spaces.

The back bone of this work is the collection of examples and exercises in Chapters 2 and 3. It is my hope that all mathematical results and tools required to solve the exercises are contained in Chapters 2 and 3 and in Appendix B. The manuscript was never intended to provide complete mathematical proofs of all the main results since these may be found elsewhere in the literature. Further, inclusion of (even more) long and technical proofs would remove focus from the main goal which is to learn how to solve problems. However, any lecturer using these lecture notes should spend part of the lectures on (sketches of) proofs in order to illustrate how to work with Markov chains in a formally correct way. This may include adding a number of formal arguments not present in the lecture notes. Some exercises in Appendix C are formulated as step-by-step instructions on how to construct formal proofs of selected theoretical results. It is definitely advisable (and probably necessary) to consult other textbooks on Markov chains in order to be able to solve *all* of the exercises in Appendix C. I advise students to postpone these exercises until they feel familiar with the exercises in Chapters 2 and 3.

For further reading I can recommend the books by Asmussen [2003, Chap. 1-2], Brémaud [1999] and Lawler [2006, Chap. 1-3]. My own introduction to the topic was the lecture notes (in Danish) by Jacobsen and Keiding [1985].

Many of the exercises presented in Chapter 3 are greatly inspired by examples in Ragner Nordberg's lecture notes on *Basic Life Insurance Mathematics* (Version: September 2002). The presentation of the mathematical results on Markov chains have many similarities to various lecture notes by Jacobsen and Keiding [1985], by Nielsen, S. F., and by Jensen, S. T.

Part of this material has been used for *Stochastic Processes 2010/2011-2015/2016* at University of Copenhagen. I thank Massimiliano Tamborrino, Ketil Biering Tvermosegaard, Nina Munkholt, Rene Aakjær Jensen, Niels Olsen, Frederik Riis Mikkelsen, Mads Rystok Bisgaard, Anne Helby Pedersen, Martin Jakobsen, Camilla Nicole Schaldemose, Susanne Ditlevsen, and the students for many useful comments to this revised version. A special thanks goes to Christian Duffau-Rasmussen, Andreas Bjerre-Nielsen, Mathias Luidor Heltberg, Frederik Sørensen, Gitte Lerche Aalborg, Line Rosendahl Meldgaard Pedersen and Søren Wengel Mogensen who agreed to spend a few hours to provide me with useful feedback on the very first version of the lecture notes.

Copenhagen, Octorber 2016
Anders Tolver

Contents

1 *Introduction* 7

Motivation and some examples of Markov chains 7

About these lecture notes 10

Transition diagrams 13

Overview of exercises 14

2 *Markov chains in discrete time* 15

Definition of a Markov chain 15

Classification of states 19

Limit results and invariant probabilities 30

Absorption probabilities 37

Exercises 47

3 *Markov chains in continuous time* 67

Definition and the minimal construction of a Markov chain 67

Properties of the transition probabilities 71

Invariant probabilities and absorption 77

Birth-and-death processes 90

Exercises 97

A *Random variables and stochastic processes* 123

Probability measures 123

Random variables 124

Stochastic processes 126

B *Mathematical tools* 131

B.1 *Elementary conditional probabilities* 131

B.2 *Some formulaes for sums and series* 133

B.3 *Some results for matrices* 134

B.4 *First order differential equations* 136

B.5 *Second order linear recurrence equations* 137

B.6 *The ratio test* 138

B.7 *Integral test for convergence* 138

B.8 *How to do certain computations in* R 139

C *Proofs of selected results* 147

C.1 *Recurrence criterion 1* 147

C.2 *Number of visits to state j* 148

C.3 *Invariant distribution* 150

C.4 *Uniqueness of invariant distribution* 152

C.5 *On the ergodic theorem for discrete-time Markov chains* 153

D *Bibliography* 157

E *Index* 159

1 Introduction

Motivation and some examples of Markov chains

When my first child started in daycare, I started to register the outcome of a stochastic variable with two possible outcomes

ill: meaning that the child is not ready for daycare

ok: meaning that the child is ready for daycare

Consecutive recordings of the health state of a child made every morning is an excellent example of a sample of a discrete-time stochastic process. The sampling regime is discrete because I do not register the health state continuously at any time point but only once a day. The process is stochastic (in contrast to deterministic) because I never know with certainty whether the child will be ill or healthy on the following morning.

The sample of the health state on the first 17 days (also referred to as the sample path) is given below

ok, ok ,ok ,ok ,ok ,ill ,ill ,ill ,ill ,ill ,ok ,ok ,ok ,ok ,ok ,ok ,ill ,...

A stochastic process is a mathematical model for a sequence of random variables. The model should allow us to compute the probability of various events associated to a random phenomena. Personally, I was particularly interested in the following type of questions

- Will the child be ready for daycare tomorrow permitting me to go to work?
- Will the child be ready for daycare on Friday, where I have a meeting that can not be cancelled?
- Will the child be ready for daycare for all days during the next week?
- What is the average time between two periods of illness?

- How many days should I expect to be off work to take care of a sick child within the next year?

One of the aims of this course is to learn how to build mathematical models for random events that allows us to compute the answer to the type of questions listed above.

We are only going to deal with a very simple class of mathematical models for random events namely the class of Markov chains on a finite or countable state space. The state space is the set of possible values for the observations. Thus, for the example above the state space consists of two states: ill and ok. Below you will find an example of a Markov chain on a countably infinite state space, but first we want to discuss what kind of restrictions are put on a model by assuming that it is a Markov chain.

Within the class of stochastic processes one could say that Markov chains are characterised by the dynamical property that they never look back. The way a Markov chain continues tomorrow is affected by where it is today but independent of where it was yesterday or the day before yesterday. As long as we know the (present) value of a Markov chain, our prediction about the future behaviour of the process does not change if we get additional information about past recordings of the process.

It is clear that many random processes from real life do not satisfy the assumption imposed by a Markov chain. When we want to guess whether a child will be ready for daycare tomorrow, it probably influences our prediction whether the child has only been ill today or whether it has been ill for the past 5 days. This suggests that a Markov chain might not be a reasonable mathematical model to describe the health state of a child.

We shall now give an example of a Markov chain on an countably infinite state space. The outcome of the stochastic process is generated in a way such that the Markov property clearly holds. The state space consists of the grid of points labeled by pairs of integers. We assume that the process starts at time zero in state $(0,0)$ and that (every day) the process moves one step in one of the four directions: up, down, left, right. Each direction is chosen with equal probability $(= 1/4)$.

This stochastic process is called the (symmetric) random walk on the state space

$$\mathbb{Z} \times \mathbb{Z} = \{(i,j) | i,j \in \mathbb{Z}\}.$$

The process satisfies the Markov property because (by construction!) the *next* state is always constructed by moving one step in a random

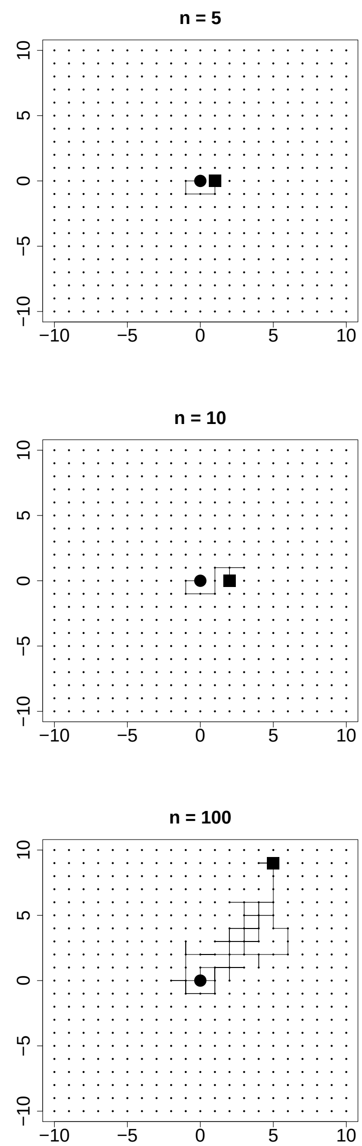


Figure 1.1: The trajectory of a symmetric random walk on $\mathbb{Z} \times \mathbb{Z}$ observed at times $n = 5, 10, 100$ when it starts in state $(0,0)$ at time $n = 0$. Note that the figure does not allow us to reconstruct the entire sample path (=sequence of visited states). It turns out that the random walk returned to state $(0,0)$ at time $n = 15$. If you do Exercise 2.4.3 you will know the probability that it will ever return to state $(0,0)$.

direction from the current state no matter how the process arrived at the current state.

The fact that the state space of the symmetric random walk on $\mathbb{Z} \times \mathbb{Z}$ is infinite makes it a more complicated but at the same time more interesting mathematical object to study. We may still (with some work) answer simple questions of the form

- What is the probability that the random walk is in state $(2, 2)$ at time $n = 10$?
- What is the probability that the random walk does not return to state $(0, 0)$ for the next time period of length 5?

However, as the state space is infinite there are a number of events that may or may not occur even if we consider an infinite time horizon. In particular, it is not obviuos

- if the random walk will ever reach (i.e. hit) state $(2, 2)$
- if the random walk will ever return to state $(0, 0)$
- what will be the average number of visits to state $(0, 0)$ if we consider at very long time horizon up to time $n = 1000$?

The last three questions have to do with the *recurrence* properties of the random walk. Much of our work on this course are centered around mathematical tools to study the long run behaviour of Markov chains on countably infinite state spaces.

So far we have only discussed mathematical models for random events that are observed at discrete time points - for instance once every day. During this course we shall also consider stochastic processes in continuous time, where the value of a random experiment is available at any time point.

The windows of my office offer an excellent view to all the cars and busses driving on Nørre Allé. The first example of a continuous-time stochastic process that comes to my mind is the number of yellow busses that have passed since I started writing this paragraph of the lecture notes (so far I have counted four!). Clearly, as I have kept my eyes fixed on Nørre Allé at any time, the value of the process is available at any time. My collected data is an example of a continous-time stochastic process. The possible values of the process is $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ hence the state space is countably infinite. The jump structure of the process is, however, very simple as only jump (upwards) of size 1 may occur. In the literature this is refered to as a counting process.

A mathematical model for the counting process of busses on Nørre Allé must describe the probability distribution for the passage times. Formally, a stochastic process includes the description of a probability space $(\Omega, \mathcal{F}, P)$ and a family of random variables (indexed by $t \in [0, \infty)$)

$$X(t) : \omega \to X(t)(\omega) \in \mathbb{N}_0.$$

For the stochastic process to be a Markov chain the distribution of $\{X(t)\}_{t \geq 0}$ must satisfy some mathematical conditions that it is hard to state and verify formally. In this course we therefore restrict our attention to continuous-time Markov chains with the property that the sample paths

$$t \to X(t)(\omega)$$

are piecewise constant. It turns out that the class of piecewise constant Markov chains on a countable state space have a nice description in terms of transition intensities.

Returning to our recordings of busses on Nørre Allé a Markov model include the describtion of the intensity, q_j, of bus arrivals given that we have already seen j busses, $j \in \mathbb{N}_0$. The interpretation of the intensity q_j is that the probability of observing the next bus (i.e. number $j + 1$) within a small time interval Δt is approximately equal to $q_j \cdot \Delta t$.

A very simple (-but probably too simple) model for the number of busses on Nørre Allé would be to assume that the intensity of future arrivals is given by the same parameter $\lambda > 0$ no matter the number of busses we have seen so far. This model is known as the *Poisson process* with intensity λ or as a *pure birth process* with intensity λ.

During the course we shall extend our study from the very simple Poisson process to Markov chains with a more complicated structure of the jumps. One of our prime examples will be the class of birth-and-death processes. A birth-and-death process is a mathematical model for a stochastic process in continuous-time that may move one step up or one step down at any time. This class of models is flexible enough to cover many interesting examples (population dynamics, queueing systems) but at the same time simple enough to allow a mathematical characterization of various important properties.

About these lecture notes

The main purpose of developing the present manuscript has been to collect a number of problems providing an easy introduction to the most basic theory of Markov chains on finite or countable state spaces. The exposition differs from most textbooks on Markov chains in that the problems take up most of the space while only a limited number of pages are devoted to the presentation of the mathematical results. This reflects my personal point of view that you should learn the stuff by working with the problems.

The core of this book is the chapters entitled *Markov chains in discrete-time* and *Markov chains in continuous-time*. They cover the main results on Markov chains on finite or countable state spaces. It is my hope that you can always easily go back to these chapters to find relevant definitions and results that hold for Markov chains. The last part of Chapters 2 and 3 present a number of problems divided into various subsections according to the size of the state space for the Markov chain. Whenever possible we have tried to put the exercises into a practical context if they deal with models that have reasonable interpretations in the real world. As the present course preceeds a course on life insurance mathematics a great number of the exercises are motivated by this particular application. However, other important examples from the vast area of applications of Markov chains have found their way to the present collection of problem, most notably from the field of queueing theory.

Clearly, the exercises vary in their difficulty and probably also in their relevance to a student just wanting to pass the course. We have made an effort to ensure that most exercises contains a mixture of (very) simple and more complicated questions. This is done because we know how frustrating it feels not to be able to get anywhere when trying to do the exercises at home. On the other hand this also implies that everyone should be able to prepare something before the exercise classes. If you show up at the exercise classes without having prepared any question for any exercise then the teaching assistant will not believe that you gave it a try. It is more likely that you will be regarded as a lazy and unambitious student ;-) If I am mistaken on this point please let me know.

I want you to remember that you are supposed to do a written exam to pass this course. Therefore my general advice to you is to use one of the problems or examples as offset anytime you work with the course. Use the questions in the problems to figure out what part of Chapters 2 or 3 that might be relevant to answer the question. Use the problems to find relevant pages from the slides used for the lectures. Do not expect it to work the other way around. It will probably take you a lot of time to read and understand the definitions and theorems of Chapters 2 or 3 and every tiny mathematical argument presented at the lectures. Even if you do manage to digest all the mathematics you will probably not find it straightforward to apply it to solve the problems. There is a huge discrepancy between reading (about) probability theory and being able to solve problems involving probability theory. In my opinion this is the main reason why courses on Markov chains (as they are often taught) have a reputation for being very difficult.

Work with the problems if you want to do well at the exam!

If you are a student who wants to learn how to answer standard questions related to Markov chains then you should read the two main chapters and do a selection of the exercises given as cross references therein. Most of the exercises are fairly easy and repetitive in their nature and you will easily be comfortable about applying the main results and definitions. For most students this will be a shot-proof way to pass the exam on an introductory course on Markov chains.

Fortunately, some students have much higher ambitions. They want to understand the mathematics behind the results in the two main chapters. They have a strong feeling that this is absolutely necessary for them in order to build up some intuition and understanding of what is really a Markov chain. Personally, I believe that the key to a succesfull introductory course on stochastic processes lies in not spending too much time on given rigorous mathematical proofs. It is lengthly and boring and does not as easily build up your intuition as if you work with some interesting examples. This point of view may not be shared by all of you. For that reason I have tried to indicate throughout the book what you should do to get a comprehensive and rigorous exposition of the mathematics. In order not to destroy the flow of the book for those who prefer the light way I have made the following choices

- short mathematical proofs and arguments are usually included in the two main chapters
- some proofs are left as exercises with a clear indication of a proof strategy
- some proofs are given in the appendices
- some proofs are not given in the lecture notes but references to secondary literature are provided

The result should be a book that you can use both as a soft and easy introduction to Markov chains or as a source to learn more of the mathematics and probability theory behind this appealing class of stochastic processes.

Transition diagrams

We advocate for visualising the dynamics of a Markov chain whenever possible. This will be done using *transition graph*s with *nodes* (or vertices) representing the states of the Markov chains and *edges* representing transitions.

For a discrete time Markov chain (at least on a finite state space) the dynamics of the chain is given by the *matrix of transition probabilities*. On the graph the transition probabilities are given as labels to the arrow representing the individual transitions. Usually, we use the convention that an edge corresponding to a zero of the transition matrix need not be drawn on the graph. Remember that a discrete-time Markov chain need not jump to a new state at every transition. This is indicated by a circular arrow on the transition diagram. As the transition probabilities for arrows pointing out from a state should always sum to one we will occationally omit the arrows from a state pointing to itself putting our faith in the readers ability to add the remaining arrows to the transition diagram. To illustrate this point complete and lazy examples of transition diagrams for the same three state discrete-time Markov chain are displayed below.

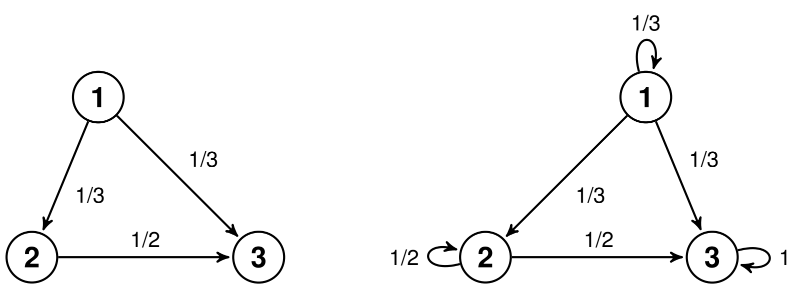


Figure 1.2: Two different versions of a transition diagram for the same discrete-time Markov chain with three states.

For a continuous-time Markov chain the dynamics is given by the time spent in each state and the distribution of the jumps whenever they occur. For a finite state space Markov chain everything is summarized in the *transition intensity matrix* with non-negative off diagonal entries and diagonals adjusted to make all rows sum to zero. The chain may be visualized by a transition diagram with nodes representing individual states and edges representing transitions. The correspondence between the transition intensity matrix and the transition diagram is obtained by labeling edges by the corresponding entry of the transition intensity matrix. In contrast to the discrete time case we (always!) omit edges of transition intensity zero. Further, there are no circular arrows from any state pointing to itself. An example of a transition diagram for a continuous-time Markov chain is given below.

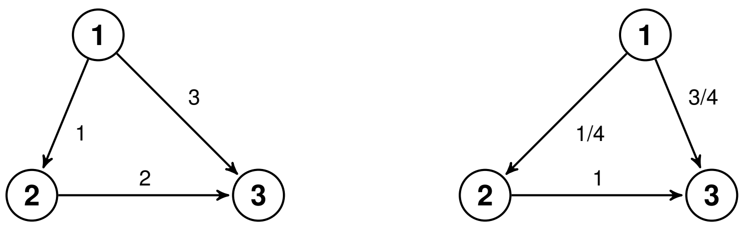


Figure 1.3: Transition diagram for a continuous time Markov chain with three states (left) and transition diagram for the corresponding discrete-time Markov chain of jumps (right).

Overview of exercises

Below we have listed some important concepts related to Markov chains and corresponding Exercises dealing with the concept. The list is not complete, in particular for Exercises covering several topics.

Communication classes
Exercises: 2.2.1, 2.3.1, 2.3.7, 3.4.2

Transience or recurrence
Exercises: 2.2.1, 2.3.7

Null-recurrence or positive recurrence
Exercises: 2.4.2, 2.4.3, 2.4.4, 2.4.6, 3.2.4, 3.4.2, 3.4.4

Periodicity
Exercises: 2.3.2, 2.3.7

Absorption probabilities
Exercises: 2.3.7, 2.3.8, 3.2.4

Invariant distribution
Exercises: 2.2.2, 2.2.4, 2.3.3, 2.3.4, 2.3.7, 2.4.1, 3.2.1, 3.2.5, 3.3.4, 3.4.3

Recurrence (=return) time
Exercises: 2.1.2, 2.3.2, 2.4.6

Markov property
Exercises: 2.2.4, 2.3.6, 3.3.1

Kolmogorov's differential equation
Exercises: 3.2.2, 3.2.3, 3.3.3

Transition probabilities
Exercises: 2.1.4, 2.3.4, 3.1.3, 3.2.1

Poisson process
Exercises: 3.1.2, 3.2.4, 3.2.6, 3.4.1, 3.5.1, 3.5.2

2

Markov chains in discrete time

Definition of a Markov chain

A **stochastic process** in discrete-time is a family, $(X(n))_{n \in \mathbb{N}_0}$, of random variables indexed by the numbers $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The possible values, S, of $X(n)$ are referred to as the **state space** of the process. In this course we consider only stochastic processes with values in a finite or countable state space. The mathematician may then think of a **random variable** , X, on S as a measurable map [1]

$$X : \quad (\Omega, \mathcal{F}) \to (S, \mathcal{P}(S))$$

where $\mathcal{P}(S)$ is the family of all subsets of S.

[1] For a brief discussion of random variables you are referred to Appendix A.

The distribution of a discrete-time stochastic process [2] with at most countable state space, S, is characterised by the point probabilities

$$P(X(n) = i_n, X(n-1) = i_{n-1}, \ldots, X(0) = i_0)$$

for $i_n, i_{n-1}, \ldots, i_0 \in S$ and $n \in \mathbb{N}_0$. From the definition of elementary conditional probabilities it follows that

[2] For more details on how to formally define the distribution of a stochastic process have a look at Appendix A.

$$\begin{aligned} & P(X(n) = i_n, \ldots, X(0) = i_0) \\ = \quad & P(X(n) = i_n | X(n-1) = i_{n-1}, \ldots, X(0) = i_0) \\ \times \quad & P(X(n-1) = i_{n-1} | X(n-2) = i_{n-2}, \ldots, X(0) = i_0) \\ \times \quad & \ldots \\ \times \quad & P(X(1) = i_1 | X(0) = i_0) \times P(X(0) = i_0). \end{aligned}$$

This is a general identity that holds for *any* discrete-time stochastic process on a countable state space. In these lecture notes we are only going to discuss the class of Markov chains to be defined below.

A discrete-time **Markov chain** on a countable set, S, is a stochastic process satisfying the **Markov property**

$$\begin{aligned} & P(X(n) = i_n | X(n-1) = i_{n-1}, \ldots, X(0) = i_0) \\ = \quad & P(X(n) = i_n | X(n-1) = i_{n-1}) \end{aligned} \tag{2.1}$$

for any $i_n, \ldots, i_0 \in S$ and $n \in \mathbb{N}$. Introducing the notation

$$P_{i,j}(n-1) = P(X(n) = j | X(n-1) = i)$$

we immediately observe that for a Markov chain the formula for the point probabilities simplifies to

$$\begin{aligned} & P(X(n) = i_n, X(n-1) = i_{n-1}, \ldots, X(0) = i_0) \\ = \quad & P_{i_{n-1},i_n}(n-1) \cdot P_{i_{n-2},i_{n-1}}(n-2) \cdot \ldots \cdot P_{i_0,i_1}(0) \cdot P(X(0) = i_0). \end{aligned}$$

We shall make a final simplification by considering only **time-homo--geneous** Markov chains for which the **transition probabilities** $P_{i,j}(n) = P_{i,j}$ do not depend on the time index $n \in \mathbb{N}$. For a discrete-time and time-homogeneous Markov chain on S we thus have that

$$P(X(n) = i_n, \ldots, X(0) = i_0) = P_{i_{n-1},i_n} \cdot \ldots \cdot P_{i_0,i_1} \cdot \phi(i_0) \tag{2.2}$$

where we use the notation

$$\phi(i_0) = P(X(0) = i_0)$$

for the **initial distribution** of $X(0)$.

Definition 1 (Homogeneous Markov chain in discrete-time) *A time-homogeneous Markov chain on a finite or countable set S is a family of random variables, $(X(n))_{n \in \mathbb{N}_0}$, on a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$P(X(n+1) = j | X(n) = i, X(n-1) = i_{n-1}, \ldots, X(0) = i_0) = P_{i,j}$$

for $j, i, i_{n-1}, \ldots, i_0 \in S$ and $n \in \mathbb{N}_0$. The distribution of the Markov chain is uniquely determined by the initial distribution and the transition probabilities

$$\begin{aligned} \phi(i) &= P(X(0) = i) \quad \leftarrow \textit{initial distribution} \\ P_{i,j} &= P(X(n+1) = j | X(n) = i) \quad \leftarrow \textit{transition probabilities}. \end{aligned}$$

□

Any probability vector $\overline{\phi} = (\phi(i))_{i \in S}$ and two-dimensional array of probabilities $P = (P_{i,j})_{i,j \in S}$ with $\sum_{j \in S} P_{i,j} = 1$ for all $i \in S$ defines the distribution of a time-homogeneous Markov chain on S through the identity (2.2). When the state space is finite we speak of the transition matrix P. As we will consider only time-homogeneous Markov chains we will throughout these lecture notes omit the phrase time-homogeneous and refer to the process simply as a Markov chain.

The dynamics of a discrete-time Markov chain with state space S is given by the array (or matrix), P, of transition probabilities. A similar representation is given by a directed graph (the **transition diagram**) with nodes representing the individual states of the chain and directed edges labeled by the probability of possible transitions.

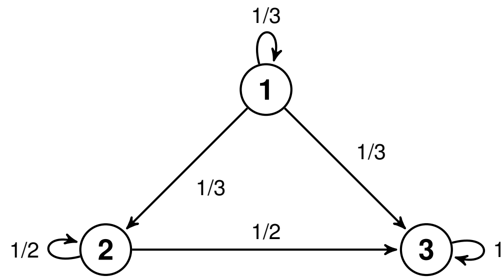


Figure 2.1: Transition diagram for a discrete-time Markov chain with three states. From state 1 there is a probability 1/3 of jumping to either of the states 1,2 or 3.

The probability of any event involving the values $X(0), \ldots, X(n)$ of a Markov chain up to time n may be obtained by splitting the event into disjoint sets of the form

$$(X(0) = i_0, \ldots, X(n) = i_n)$$

and summing up point probabilities of the form given by (2.2). For finite state space Markov chains the computation of the probability of certain events have simple representations in terms of matrix operations.

Theorem 2 (n-step transition probabilities) *For a Markov chain on a finite state space, $S = \{1, \ldots, N\}$, with transition probability matrix P and initial distribution $\overline{\phi} = (\phi(1), \ldots, \phi(N))$ (row vector) then the distribution of $X(n)$ is given by*

$$(P(X(n) = 1), \ldots, P(X(n) = N)) \quad = \quad \overline{\phi}P^n. \tag{2.3}$$

Proof: The proof goes by induction. For $n = 1$ we have

$$\begin{aligned} P(X(1) = j) &= \sum_{i \in S} P(X(0) = i, X(1) = j) \\ = \quad \sum_{i \in S} P(X(0) = i) \cdot P(X(1) = j | X(0) = i) &= \sum_{i \in S} \phi(i) P_{i,j}, \end{aligned} \tag{2.4}$$

where we use the definition of the transition probabilities to obtain the third equality. Note that the right hand side of (2.4) is exactly the j-th element of the row vector given by the matrix product $\overline{\phi}P$.

Using the Markov property we get for arbitrary $n + 1$ that

$$\begin{aligned} P(X(n+1) = j) &= \sum_{i \in S} P(X(n) = i, X(n+1) = j) \\ = \quad & \sum_{i \in S} P(X(n) = i) \cdot P(X(n+1) = j | X(n) = i) \\ = \quad & \sum_{i \in S} P(X(n) = i) \cdot P_{i,j}. \end{aligned}$$

Assuming that $P(X(n) = i)$ is given as the i-th element of the row vector $\overline{\phi}P^n$ we observe that the j-th coordinate of the row vector $(\overline{\phi}P^n)P$ coincides with the expression for $P(X(n + 1) = j)$ obtained above. In particular, by induction we deduce that the distribution of $X(n+1)$ is given by $(\overline{\phi}P^n)P = \overline{\phi}P^{n+1}$. □

Example 3 *Consider the discrete-time Markov chain with three states corresponding to the transition diagram on Figure 2.2. Assume that the initial distribution of $X(0)$ is given by*

$$\phi(1) = \phi(2) = 1/2.$$

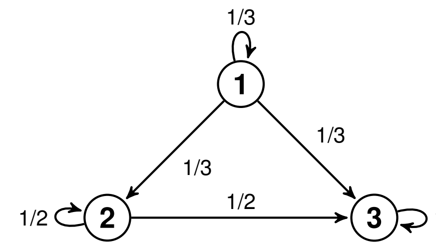


Figure 2.2: Transition diagram for Markov chain considered in Example 3.

Compute the following

1. $P(X(0)=1, X(1)=2, X(2)=3)$.
2. $P(X(2)=i)$, *for* $i=1,2,3$.
3. $P(T_3=2)$ *where*

$$T_3 = \inf\{n > 0 | X(n) = 3\}$$

is the time of the first visit to state 3.

The probability in question 1. may be computed using formula (2.2) for the point probabilities of a discrete-time Markov chain. We get

$$P(X(0)=1, X(1)=2, X(2)=3) = \phi(1)\cdot P_{1,2}\cdot P_{2,3} = \tfrac{1}{2}\cdot\tfrac{1}{3}\cdot\tfrac{1}{2} = \tfrac{1}{12}.$$

The distribution of $X(2)$ from question 2. is easily computed using the matrix formula (2.3) from Theorem 2. In particular

$$\begin{aligned} & (P(X(2)=1), P(X(2)=2), P(X(2)=3)) \\ = & (1/2, 1/2, 0)\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}^2 = (\tfrac{1}{18}, \tfrac{19}{72}, \tfrac{49}{72}). \end{aligned}$$

To answer question 3. note that there are three possibilities[3] *for the values of $X(0)$, $X(1)$, $X(2)$ for which the* first *visit to state* 3 *happens at time* 2

$$\begin{aligned} (X(0)=1, X(1)=1, X(2)=3) \\ (X(0)=1, X(1)=2, X(2)=3) \\ (X(0)=2, X(1)=2, X(2)=3). \end{aligned}$$

[3] The event $(T_3 = 2)$ expresses that the first visit to state 3 must happen exactly at time $n = 2$.

Using formula (2.2) we get

$$\begin{aligned} P(T_3=2) &= P(X(0)=1, X(1)=1, X(2)=3) \\ &+ P(X(0)=1, X(1)=2, X(2)=3) \\ &+ P(X(0)=2, X(1)=2, X(2)=3) \\ &= \phi(1)\cdot P_{1,1}\cdot P_{1,3} + \phi(1)\cdot P_{1,2}\cdot P_{2,3} + \phi(2)\cdot P_{2,2}\cdot P_{2,3} \\ &= \tfrac{1}{2}\cdot\tfrac{1}{3}\cdot\tfrac{1}{3} + \tfrac{1}{2}\cdot\tfrac{1}{3}\cdot\tfrac{1}{2} + \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{18} + \tfrac{1}{12} + \tfrac{1}{8} = \tfrac{19}{72}. \end{aligned}$$

□

Classification of states

For a discrete-time Markov chain with state space S and transition probabilities $P = (P_{i,j})_{i,j \in S}$ we say that there is a possible path from state i to state j if there is a sequence of states

$$i = i_0 \to i_1 \to \ldots \to i_n = j$$

such that for all transitions along the path we have $P_{i_{l-1},i_l} > 0$, $l = 1, \ldots, n$. We will also use the phrase that state j is **accessible** from state i.

We say that two states $i, j \in S$ **communicate** if there is a possible path from i to j and from j to i. We use the notation $i \leftrightarrow j$ when the two states i and j communicate. If we use the convention that $i \leftrightarrow i$ (i.e. that a state always communicates with itself) then the relation $\leftrightarrow$ partitions the state space, S, into disjoint **communication classes**.

A Markov chain is said to be **irreducible** if there is only one communication class.

Example 4 *For the Markov chain given by the transition diagram in Figure 2.2 above (-see page 18) it is not possible to go to state* 1 *from states* 2 *or* 3*. In particular, no other state communicate with state* 1*. From state* 2 *the Markov chain may jump to state* 3 *but it is impossible to get back to state* 2 *once the chain has jumped to state* 3*. We conclude that states* 2 *and* 3 *do not communicate. We conclude that the Markov chain is not irreducible as it may be partitioned into three communication classes* $\{1\}$, $\{2\}$ *and* $\{3\}$*. Try to give some examples of how the transition probabilities on Figure 2.1 may be modified such that we get an irreducible Markov chain.*

For more exercises on irreducibility and communication classes you are encouraged to do Exercises 2.2.1 and 2.3.1. □

Definition 5 (Closed communication classes) *A communication class,* $\mathcal{C}$*, for a Markov chain is said to be* ***closed*** *if for all* $i \in \mathcal{C}$ *it holds that*

$$\sum_{j \in \mathcal{C}} P_{i,j} = 1.$$

If $\mathcal{C}$ *has only finitely many elements then* $\mathcal{C}$ *is closed if the submatrix of transition probabilities restricted to* $\mathcal{C}$ *has all row sums equal to* 1*.* □

Example 6 *The Markov chain on Figure 2.3 has three different communication classes:* $\mathcal{C}_1 = \{1\}$, $\mathcal{C}_2 = \{2, 3\}$ *and* $\mathcal{C}_3 = \{4, 5\}$*.*

To see this observe that it is possible to make transitions between states 2 *and* 3 *and between states* 4 *and* 5*. Further, note that no other pair of states communicate. Since transitions away from* $\mathcal{C}_2$ *or* $\mathcal{C}_3$ *are not possible it follows that these classes are closed. The class* $\mathcal{C}_1$ *is not closed since the probability of staying in the class is* $1/2$ *which is strictly less than* 1*.* □

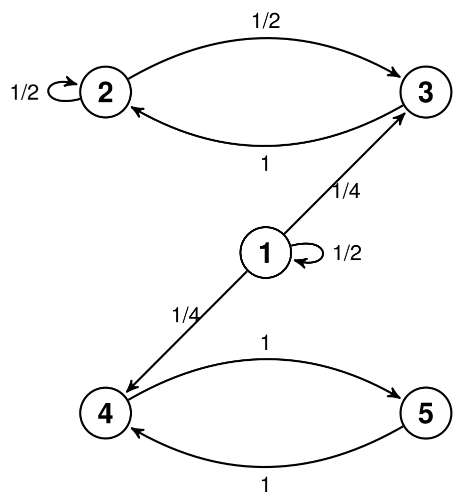


Figure 2.3: Transition diagram for a discrete-time Markov chain with five states and three communication classes.

Remark 7 (On closed classes) *The restriction of a Markov chain to a closed communication class is an irreducible Markov chain. Most results in these lecture notes are formulated for irreducible Markov chains. This means that you should just break down the analysis of a Markov chain by restricting the analysis to each closed class.* □

For any state $i \in S$ we define the **hitting time** of i by

$$T_i = \inf\{n > 0 | X(n) = i\}.$$

When the Markov chain is assumed to start in state i we will often refer to T_i above as the **return time** or the **recurrence time** to state i.

If two states i and j communicate according to the definition above then we know that

$$P(T_i < +\infty | X(0) = j) > 0 \quad \text{and} \quad P(T_j < +\infty | X(0) = i) > 0.$$

More informally: i and j communicate if it is possible (with positive probability!) to get from i to j and from j to i.

To fully understand and describe the dynamics and the long-run behaviour of a Markov chain a much more relevant question will be whether we can be sure (i.e. with probability equal to one!) that the Markov chain will move from i to j and from j to i. This leads us to the definition of transient and recurrent states (or communication classes).

Definition 8 (Recurrence and transience) *For a discrete-time Markov chain on S we say that a state $i \in S$ is **recurrent** if and only if*

$$P(T_i < +\infty | X(0) = i) \quad = \quad 1. \tag{2.5}$$

*If $P(T_i < +\infty | X(0) = i) < 1$ then i is said to be a **transient** state.* □

In clear text the meaning of (2.5) is that state i is recurrent if and only if the probability of ever returning to state i is one given that the Markov chain starts in state i at time $n = 0$.

Example 9 *Consider again the Markov chain from Example 6 given by the transition diagram in Figure 2.3. If the Markov chain is started in state* 1 *it may perform a jump away from state* 1 *(-with probability* $1/2$*) and in this case it will never return to state* 1*. This shows that*

$$P(T_1 = +\infty | X(0) = 1) \geq 1/2 > 0$$

(-or equivalently that $P(T_1 < +\infty | X(0) = 1) \leq 1/2 < 1$*) and we get from Definition 8 that state* 1 *is transient.*

For $X(0) = 4$ *or* $X(0) = 5$ *we know for sure that the Markov chain will be back again to the initial state after two steps i.e.*

$$P(T_i < +\infty | X(0) = i) \geq P(T_i = 2 | X(0) = i) = 1$$

for $i = 4, 5$ *and we have by Defintion 8 that states* 4 *and* 5 *are recurrent.*

For $X(0) = 2$ *we have that*

$$P(T_2 = 1 | X(0) = 2) = 1/2$$

and

$$P(T_2 = 2 | X(0) = 2) = 1/2$$

(-corresponding to a jump from 2 *to* 3 *followed by a jump from* 3 *to* 2*). Thus*

$$\begin{aligned} & P(T_2 < +\infty | X(0) = 2) \\ = & \sum_{n=1}^{\infty} P(T_2 = n | X(0) = 2) \\ \geq & P(T_2 = 1 | X(0) = 2) + P(T_2 = 2 | X(0) = 2) = 1 \end{aligned}$$

showing that state 2 *is recurrent.*

If the Markov chain is started in state 3 *a little more work is required. Intuitively, the Markov chain will first jump to state* 2 *and eventually it will jump back to the state* 3*. Therefore state* 3 *is recurrent. However, a formal argument requires that we show that*

$$P(T_3 < +\infty | X(0) = 3) = 1.$$

To get used to the notation let us illustrate how to do this in details.

If the Markov chain is started at $X(0) = 3$ *then the event* $(T_3 = n)$ *expresses that the* first *visit to state* 3 *(after time* 0*!) happens at time* n*. Referring to the transition diagram of the Markov chain this is only possible if the Markov chain makes a jump from state* 3 *to* 2 *(-which happens with probability* $P_{3,2} = 1$*), followed by* $n - 2$ *transitions from* 2 *to* 2 *(each* *with transition probabiliy* $= 1/2$*) and finally a jump from state* 2 *to* 3 *(*$P_{2,3} = 1/2$*). In particular, we get that*

$$P(T_3 = n|X(0) = 3) = 1 \cdot \left(\tfrac{1}{2}\right)^{n-2} \cdot \tfrac{1}{2} = \left(\tfrac{1}{2}\right)^{n-1}, \quad n \geq 2.$$

and hence

$$P(T_3 < +\infty|X(0) = 3) = \sum_{n=1}^{\infty} P(T_3 = n|X(0) = 3) = \sum_{n=2}^{\infty} \left(\tfrac{1}{2}\right)^{n-1} = 1$$

demonstrating that state 3 *is recurrent according to Definition 8.* □

Example 9 represents a simple situation where it is possible to apply Definition 8 to verify if a state is recurrent or transient. It requires that one can compute (or at least have some control over) the probability

$$P(T_i = n|X(0) = i)$$

that the *first* return of the Markov chain to state i happens exactly at time n. Another instructive example is given in Exercise 2.4.6 where the structure of possible jumps is particularly simple. We shall now give two results (Theorems 10 and 15) that turn out to be more useful for Markov chains with a more complicated jump structure.

Theorem 10 (Recurrence criterion 1) *For a discrete-time Markov chain with n-step transition probabilities* $P^n = (P^n)_{i,j \in S}$ *then state i is recurrent if and only if* $\sum_{n=1}^{\infty} (P^n)_{i,i} = +\infty$.

Sketch of proof: The following proof of Theorem 10 is reproduced based on Jacobsen and Keiding [1985]. Let

$$T_j = \inf\{n > 0 | X(n) = j\}$$

be the time of the first visit to state j. Introduce the notation

$$f_{ij}^{(n)} = P(T_j = n|X(0) = i)$$

for the probabilty that the first visit to state j happens at time n when the Markov chain is started at state $X(0) = i$. Using the elementary identity

$$f_{ij} := \sum_{n=1}^{\infty} f_{ij}^{(n)} = \sum_{n=1}^{\infty} P(T_j = n|X(0) = i) = P(T_j < +\infty|X(0) = i)$$

we have from Definition 8 that state i is recurrent if and only if $f_{ii} = 1$.

The technical part of the proof establishes the following formula

$$f_{ij} = \lim_{N \to \infty} \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{1 + \sum_{n=1}^{N} (P^n)_{j,j}} \tag{2.6}$$

where $(P^n)_{i,j}$ is the n-step transition probability from state i to j. The identity holds even when the infinite sum takes the value $+\infty$ and the conclusion from Theorem 10 follows by letting $i = j$.

For the **full proof of Theorem 10** [4] we need to establish (2.6). First note that by splitting the event $(X(n) = j)$ according to the time of the *first* visit to state j then

[4] The full proof is included for completeness of the exposition. You can just skip it and continue to read on after the proof.

$$\begin{aligned}
(P^n)_{i,j} &= P(X(n)=j|X(0)=i) = \sum_{m=1}^{n} P(X(n)=j, T_j = m|X(0)=i) \\
&= \sum_{m=1}^{n} P(X(n)=j|T_j=m)\cdot P(T_j=m|X(0)=i) \\
&= \sum_{m=1}^{n} P(X(n)=j|X(m)=j)\cdot P(T_j=m|X(0)=i) \\
&= \sum_{m=1}^{n} (P^{n-m})_{j,j} f_{ij}^{(m)}
\end{aligned}$$

where we have used the Markov property to get the second last equality.

Summing over $n = 1, \dots, N$ we get

$$\begin{aligned}
\sum_{n=1}^{N} (P^n)_{i,j} &= \sum_{n=1}^{N}\sum_{m=1}^{n} (P^{n-m})_{j,j} f_{ij}^{(m)} \\
&= \sum_{m=1}^{N}\sum_{n=m}^{N} (P^{n-m})_{j,j} f_{ij}^{(m)} \\
&= \sum_{m=1}^{N} f_{ij}^{(m)} \sum_{k=0}^{N-m} (P^k)_{j,j}.
\end{aligned}$$

This gives us an upper bound

$$\sum_{n=1}^{N} (P^n)_{i,j} \le \sum_{m=1}^{N} f_{ij}^{(m)} \cdot \underbrace{\sum_{k=0}^{N} (P^k)_{j,j}}_{\text{independent of } m}$$

and for any $M < N$ we have a lower bound

$$\sum_{n=1}^{N} (P^n)_{i,j} \ge \sum_{m=1}^{M} f_{ij}^{(m)} \sum_{k=0}^{N-m} (P^k)_{j,j} \ge \sum_{m=1}^{M} f_{ij}^{(m)} \sum_{k=0}^{N-M} (P^k)_{j,j}.$$

Dividing by $\sum_{n=0}^{N} (P^k)_{j,j}$ we get that for any $M < N$

$$\begin{aligned}
\sum_{m=1}^{M} f_{ij}^{(m)} \cdot \frac{\sum_{k=0}^{N-M} (P^k)_{j,j}}{\sum_{k=0}^{N} (P^k)_{j,j}} &\le \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{\sum_{k=0}^{N} (P^k)_{j,j}} \le \sum_{m=1}^{N} f_{ij}^{(k)} \Leftrightarrow \\
\sum_{m=1}^{M} f_{ij}^{(m)} \cdot \left(1 - \frac{\sum_{k=N-M+1}^{N} (P^k)_{j,j}}{\sum_{k=0}^{N} (P^k)_{j,j}}\right) &\le \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{1 + \sum_{k=1}^{N} (P^k)_{j,j}} \le \sum_{m=1}^{N} f_{ij}^{(k)}.
\end{aligned}$$

The right hand side tends to $f_{ij} = P(T_j < \infty | X(0) = i)$ for $N \to \infty$ and since $\sum_{k=N-M+1}^{N} (P^k)_{j,j} \leq M$ (there are M terms of size ≤ 1!) it follows that

$$\frac{\sum_{k=N-M+1}^{N} (P^k)_{j,j}}{\sum_{k=0}^{N} (P^k)_{j,j}} \overset{N \to \infty}{\to} 0$$

no matter if $\sum_{k=0}^{\infty} (P^k)_{j,j}$ is $= +\infty$ or $< +\infty$. [5] In particular we get for any M that

[5] For $\sum_{k=0}^{\infty} (P^k)_{j,j} < +\infty$ the conclusion follows from the fact that the tail of the infinite sum converges to zero.

$$\begin{aligned} \sum_{m=1}^{M} f_{ij}^{(m)} &\leq \liminf_{N \to \infty} \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{1 + \sum_{k=1}^{N} (P^k)_{j,j}} \\ &\leq \limsup_{N \to \infty} \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{1 + \sum_{k=1}^{N} (P^k)_{j,j}} \leq \sum_{m=1}^{\infty} f_{ij}^{(k)} = f_{ij}. \end{aligned}$$

Taking the limit as $M \to \infty$ we conclude that for any $i, j \in S$ then

$$f_{ij} = \lim_{N \to \infty} \frac{\sum_{n=1}^{N} (P^n)_{i,j}}{1 + \sum_{n=1}^{N} (P^n)_{j,j}}.$$

For $i = j$ this gives

$$P(T_i < \infty | X(0) = i) = f_{ii} = \begin{cases} \frac{\sum_{n=1}^{\infty} (P^n)_{i,i}}{1 + \sum_{n=1}^{\infty} (P^n)_{i,i}} < 1 & \text{, if } \sum_{n=1}^{\infty} (P^n)_{i,i} < +\infty \\ 1 & \text{, if } \sum_{n=1}^{\infty} (P^n)_{i,i} = +\infty. \end{cases}$$

□

It turns out that if j is accessible (i.e. may be reached) from a recurrent state i then j must communicate with i ($i \leftrightarrow j$) [6]. Furthermore, j will also be recurrent. The last observation is stated more formally in the following Theorem 11. The proof of Theorem 11 is fairly simple but instructive and builds on Theorem 10 above.

[6] To see this note that if $P_{i,j}^n > 0$ then there is a path from i to j of length $k \leq n$ not containing j. Clearly, it would be possible to leave state i for good through state j unless $P_{j,i}^m$ for some $m \in \mathbb{N}_0$.

Theorem 11 (Recurrence is a class property) *All states in a communication class are either all recurrent or all transient.*

Proof: Assume that states i and j communicate. Then there exist $l, m \in \mathbb{N}$ such that $P_{i,j}^l, P_{j,i}^m > 0$. Note that by the Markov property we observe that for any $k \geq 1$ then $P_{i,j}^l P_{j,j}^k P_{j,i}^m$ describes the probability of a loop of length $l + k + m$ from state i that visits state j after l and $l + k$ steps. Clearly, this is smaller than or equal to the probability of having a loop of length $l + k + m$ from state i which equals $P_{i,i}^{l+k+m}$. We therefore get the inequality

$$\sum_{n=1}^{\infty} P_{i,i}^n \geq \sum_{k=1}^{\infty} P_{i,i}^{l+k+m} \geq \sum_{k=1}^{\infty} P_{i,j}^l P_{j,j}^k P_{j,i}^m = P_{i,j}^l \cdot \left(\sum_{k=1}^{\infty} P_{j,j}^k \right) \cdot P_{j,i}^m$$

Either both states i, j are transient or we may assume (by symmetry) that j is recurrent and hence by Theorem 10 that $\sum_{n=1}^{\infty} P_{j,j}^n = +\infty$. In the latter case the inequality above shows that also $\sum_{n=1}^{\infty} P_{i,i}^n = +\infty$ and we deduce from Theorem 10 that state i is recurrent. □

Theorem 11 suggests that we first find the communication classes and then consider only one element from each class for classification into either recurrent or transient groups of states.

The interpretation of recurrence is very important: if the Markov chain is started in a recurrent state i then with probability one it will eventually return to state i. On the contrary, if a Markov chain starts in a transient state i then the probability of returning to state i is strictly less than one [7].

[7] Suppose that there is a possible loop of length k from state i to i then $(P^k)_{i,i} > 0$. **This does not imply** that state i is recurrent! Instead try to use Definition 8 or more likely Theorems 10 and 15.

Note carefully the difference between an irreducible and a recurrent Markov chain. For an irreducible Markov chain there is a path with positive probability between any two states $i \neq j$ while (as we shall see) for a recurrent Markov chain there will eventually be a transition between any two states $i \neq j$ with probability one!

As stated below it is possible to extract a bit more information from the definition of recurrence.

Theorem 12 (Number of visits to state i) *For a discrete-time Markov chain on S with initial distribution $P(X(0) = i) = 1$ consider the total number of visits to state i*

$$N_i = \sum_{n=1}^{\infty} 1(X(n) = i).$$

If i is a recurrent state then $N_i = +\infty$ with probability one. If i is a transient state then N_i follows a geometric distribution

$$P(N_i = k) = (1-q)^k q, \quad k \in \mathbb{N}_0$$

where $q = P(T_i = +\infty | X(0) = i)$ is the probability that the Markov chain never returns to state i.

In fact it may be demonstrated that if a recurrent state j is accessible from i then

$$P(N_j = +\infty | X(0) = i) = f_{ij}$$

where the notation f_{ij} is explained in Theorem 10. Further, for i is accessible from j then both $f_{ij} = f_{ji} = 1$ (-see also Jacobsen and Keiding [1985, Sætning 2.2 and 2.3]).

A formal proof of Theorem 12 may be constructed along the lines of Exercise C.2. Intuitively, the idea is to exploit that a Markov chain restarts itself any time it returns to state i. It therefore either returns infinitely often (recurrent), or it returns k times (with probability $(1-q)^k$) before it fails to return (with probability q). □

As explained in the following result it is easy to determine if a finite communication class is recurrent or transient.

Theorem 13 (Finite communication classes and recurrence) *A finite communication class is recurrent if and only if it is closed.*

Proof: For a closed communication class $\mathcal{C}$ we have for any $i \in \mathcal{C}, k \in \mathbb{N}$ that $\sum_{j\in\mathcal{C}} P_{i,j}^k = 1$. By interchanging the order of summation we therefore get that

$$+\infty = \sum_{k=1}^{\infty} \underbrace{\sum_{j\in\mathcal{C}} P_{i,j}^k}_{=1} = \sum_{j\in\mathcal{C}} \sum_{k=1}^{\infty} P_{i,j}^k.$$

Since $\mathcal{C}$ is finite this allows us to conclude that $\sum_{k=1}^{\infty} P_{i,j}^k = +\infty$ for some $j \in \mathcal{C}$. Using that i and j communicate we may choose an $m \in \mathbb{N}$ such that $P_{j,i}^m > 0$ and we conclude that

$$\begin{aligned} \sum_{n=1}^{\infty} P_{i,i}^n &\geq \sum_{k=1}^{\infty} P_{i,i}^{k+m} \\ &\geq \sum_{k=1}^{\infty} P_{i,j}^k P_{j,i}^m = \left(\sum_{k=1}^{\infty} P_{i,j}^k\right) P_{j,i}^m = +\infty. \end{aligned}$$

It follows from Theorem 10 that state i is recurrent.

If $\mathcal{C}$ is not closed there exists an $i \in \mathcal{C}$ such that $\sum_{j\in\mathcal{C}} P_{i,j} < 1$. Let $l \notin \mathcal{C}$ be any state of the Markov chain with $P_{i,l} > 0$. For $X(0) = i$ then we conclude that the probability of ever returning to state i will be less than or equal to $1 - P_{i,l} < 1$. It follows from Definition 8 and Theorem 11 that if the class $\mathcal{C}$ is not closed then it must be transient since it contains the transient state i. □

The last part of the proof of Theorem 13 shows that if a communication class is not closed then it must be transient. Note that this implication also holds for classes with infinitely many elements.

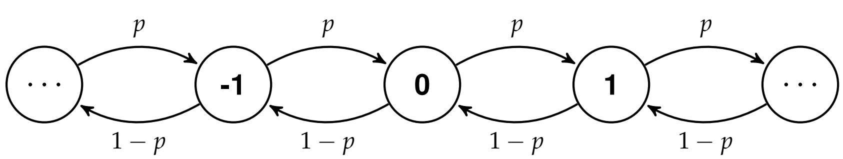


Figure 2.4: The random walk on $\mathbb{Z}$ is a famous example of a Markov chain on a countably infinite state space. It plays a central role in these lecture notes - see Example 14 and Exercise 2.4.2

Example 14 (Random walk on $\mathbb{Z}$) *The **random walk** on $\mathbb{Z}$ given by the transition diagram in Figure 2.4 is an example of a Markov chain on an infinite state space. The jump structure is very simple: from state i a jump to state $i+1$ happens with probability p, while jumps to state $i-1$ occur with probability $1-p$. Clearly, for $0 < p < 1$ there is only one communication class, i.e. the Markov chain is irreducible.*

Trying to apply Definition 8 to determine if the random walk is recurrent or transient is not easy. The first step would be to write the left hand side of (2.5) as

$$P(T_i < +\infty | X(0) = i) = \sum_{n=1}^{\infty} P(T_i = n | X(0) = i)$$

where $P(T_i = n|X(0) = i)$ *is the probability that the* first *visit to state* i *happens at time* n *given that we start the random walk at* $X(0) = i$*. Taking* $i = 0$ *then we get that* $P(T_i = n|X(0) = i) = 0$ *for* n *odd since is requires an even number of steps to return to state* 0*.*

For $n = 2$ *observe that if* $X(0) = 0$ *then there are two possible paths of length* 2 *starting and ending at state* $i = 0$*, namely* $0 \to 1 \to 0$ *and* $0 \to -1 \to 0$*. Since both paths have probabily* $p(1-p)$ *(why?) then*

$$P(T_0 = 2|X(0) = 0) = 2p(1-p).$$

For $n = 4$ *there are six different paths of length* 4 *starting and ending at state* $i = 0$

$$\begin{aligned}
0 &\to 1 \to 2 \to 1 \to 0 \\
0 &\to 1 \to 0 \to 1 \to 0 \\
0 &\to 1 \to 0 \to -1 \to 0 \\
0 &\to -1 \to 0 \to 1 \to 0 \\
0 &\to -1 \to 0 \to -1 \to 0 \\
0 &\to -1 \to -2 \to -1 \to 0.
\end{aligned}$$

To compute the probability $P(T_0 = 4|X(0) = 0)$ *we should* only *include the paths where the first visit to state* 0 *happens at time* 4 *implying that only two paths contribute to the computation (which?). We get that*

$$P(T_0 = 4|X(0) = 0) = 2p^2(1-p)^2.$$

It is not easy to find a general formula for $P(T_0 = 2k|X(0) = 0)$*. However, this is required if we want to compute* $P(T_0 < +\infty|X(0) = 0)$ *and apply directly Definition 8 to determine if the random walk is recurrent or transient.*

If we instead want to use Theorem 10 then we need expressions for

$$P^n_{i,i} = P(X(n) = i|X(0) = i).$$

Choosing $i = 0$ *then we again conclude that* $P^n_{0,0} = 0$ *for* n *odd* [8]*. Further* $P^2_{0,0} = 2p(1-p)$ *and it turns out that* $P^4_{0,0}$ *is obtained by summing all six possible loops of length* 4 *from* 0 *such that we get*

$$P^4_{0,0} = 6p^2(1-p)^2.$$

In Exercise 2.4.2 it is shown that for general $k \in \mathbb{N}$ *then*

$$P^{2k}_{0,0} = \binom{2k}{k} p^k (1-p)^k.$$

[8] Try to use Example 14 to really understand the difference between $P(T_0 = 4|X(0) = 0)$ and $P(X(4) = 0|X(0) = 0)$.

From Theorem 10 we therefore conclude that the random walk on $\mathbb{Z}$ is recurrent if and only if

$$\sum_{n=1}^{\infty} P_{0,0}^{n} = \sum_{k=1}^{\infty} P_{0,0}^{2k} = \sum_{k=1}^{\infty} \binom{2k}{k} p^k (1-p)^k = +\infty.$$

It follows (-see Exercise 2.4.2) that the random walk is recurrent if $p = 1/2$ and transient for $p \neq 1/2$.

Note that for $0 < p < 1$ there is only one closed communication class (i.e. the Markov chain is irreducible). However, the Markov chain is transient unless $p = 1/2$. This example shows that for communication classes with an infinite number of states then closed classes are not always recurrent.[9] □

[9] Example 14 shows that closed communication classes with infinitely many elements can be transient. Consequently, Theorem 13 does not hold for countably infinite classes.

In many cases it is not possible to get explicit formulaes for the n-step transition probability, $(P^n)_{i,i}$, allowing us to evaluate the sum from Theorem 10. The following result gives another way to demonstrate recurrence of a state without reference to n-step transition probabilities.

Theorem 15 (Recurrence criterion 2) *Let $(X(n))_{n\geq 0}$ be an irreducible Markov chain on S with transition probability $P = (P_{i,j})_{i,j\in S}$. Consider the system of equations*

$$\alpha(j) = \sum_{k\neq i} P_{j,k}\alpha(k), \quad j \in S, j \neq i, \tag{2.7}$$

where $i \in S$ is a fixed (but arbitrary) state.

The Markov chain is recurrent if and only if the only bounded solution to (2.7) is given by $\alpha(j) = 0, j \neq i$.

Partial proof of Theorem 15: For fixed (but arbitrary) $i \in S$ define

$$\alpha(j) = P(T_i = +\infty | X(0) = j) \quad \text{where} \quad T_i = \inf\{n \geq 0 | X(n) = i\}.$$

Note that $\alpha(j)$ is the probability of never visiting state i given that the Markov chain starts in state j. By splitting the event $(T_i = +\infty)$ according to the state of the first jump from $X(0) = j$ then it is straightforward to show that $\{\alpha(j)\}_{j\in S\setminus\{i\}}$ is a bounded solution to (2.7). One deduces that if $\alpha(j) = 0$ is the only bounded solution then we must have for all $j \neq i$ that

$$P(T_i < \infty | X(0) = j) = 1.$$

In particular, state i is recurrent since we may deduce from the Markov property that

$$\begin{aligned} P(T_i < \infty | X(0) = i) &= P_{i,i} + \sum_{j\neq i} P_{i,j} \cdot P(T_i < \infty | X(0) = j) \\ &= P_{i,i} + \sum_{j\neq i} P_{i,j} \cdot 1 = 1. \end{aligned}$$

The non-trivial part of Theorem 15 is to show that if any non-zero bounded solution to (2.7) exists then we must also have

$$P(T_i = +\infty | X(0) = j) > 0$$

for at least one $j \neq i$ implying (trivally) that the irreducible Markov chain is not recurrent. The proof of this implication may be found in Jacobsen and Keiding [1985] . □

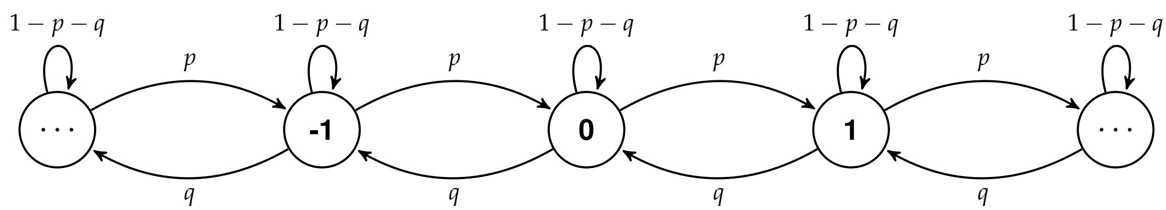


Figure 2.5: In Example 16 we study the recurrence properties of a Markov chain on $\mathbb{Z}$ which is a simple modification of the random walk from Example 14.

Example 16 (Modified random walk on $\mathbb{Z}$) *Consider the modified random walk with transition diagram given in Figure 2.5. For $p, q > 0$ (-and $p + q \leq 1$) the Markov chain is irreducible and we shall try to use Theorem 15 to determine the recurrence properties of the Markov chain* [10] *.*

Step 1: *First we need to choose some arbitrary state, i, and write down the system of equations given by (2.7) in Theorem 15. Since the transition probabilities are symmetric in all states we just use $i = 0$ and get the following system of equations*

$$\begin{aligned} \alpha(j) &= q\alpha(j-1) + (1-p-q)\alpha(j) + p\alpha(j+1), \quad |j| > 1, \\ \alpha(-1) &= q\alpha(-2) + (1-p-q)\alpha(-1) \\ \alpha(1) &= (1-p-q)\alpha(1) + p\alpha(2). \end{aligned}$$

Note that in accordance with (2.7) we are a bit careful when we write out the equations corresponding to $\alpha(j)$ where $P_{j,0} > 0$.

Step 2: *The next step is to solve the system of equations from **Step 1**. Since there are infinitely many equations (-one for each $j \neq 0$) this is only possible if the structure of the equations is very simple. Rearranging the terms we get that*

$$\alpha(j) = \frac{q}{p+q}\alpha(j-1) + \frac{p}{p+q}\alpha(j+1), \quad |j| > 1,$$

*which is a so-called **second order linear difference equation** . The general solution to equations of this form is well known and may be found using the result in Appendix B.5. In the remaining part of the Example we consider only the case $p = q$* [11]*. From Appendix B.5 we get that the general solution (-using only the equations for $\alpha(j), |j| > 1$) must have the form*

$$\begin{aligned} \alpha(j) &= c_1 + c_2 j, \quad j \geq 1 \\ \alpha(j) &= c_3 + c_4 j, \quad j \leq -1. \end{aligned}$$

[10] You are encouraged to try to write down closed form expressions for the n-step transition probabilities $(P^n)_{i,i}$. You may use some of the same ideas as in Example 14 or Exercise 2.4.2. Then try to see if you can use the recurrence criterion from Theorem 10 by determining exactly when $\sum_{n=1}^{\infty}(P^n)_{i,i} = +\infty$. This may require some work!

[11] It is a useful exercise to repeat **Step 1-4** in Example 16 for the case $q \neq p$.

Step 3: *To apply Theorem 15 we are only interested in bounded solutions to the system of equations from **Step 1**. From the general form of the solution found in **Step 2** we must have $c_2 = c_4 = 0$ to obtain a bounded solution. Plugging the resulting solution ($\alpha(j) = c_1, j \geq 1$ and $\alpha(j) = c_3, j \leq -1$) into the equations for $\alpha(1)$ and $\alpha(-1)$ we get*

$$\begin{aligned} c_3 &= qc_3 + (1-p-q)c_3 \\ c_1 &= (1-p-q)c_1 + pc_1. \end{aligned}$$

Since $0 < p, q$ the only possible values for c_1 and c_3 are zero.

Step 4: *In **Step 1-3** we have shown that the only bounded solution to (2.7) is given by $\alpha(j) = 0, j \neq 0$. In particular, it follows from Theorem 15 that for $0 < p = q$ the modified random walk is recurrent.* □

Limit results and invariant probabilities

The aim of this section is to study the limiting (also called the long-run) behaviour of a Markov chain. More precisely we want to study $P(X(n) = j)$ as $n \to \infty$ which due to the identity

$$\begin{aligned} P(X(n) = j) &= \sum_{i \in S} P(X(n) = j | X(0) = i) P(X(0) = i) \\ &= \sum_{i \in S} P(X(0) = i) P^n_{i,j} \end{aligned}$$

boils down to understanding the behaviour of the n-step transition probabilities $P^n_{i,j}$ as $n \to \infty$.

In order to discuss the long-run behaviour of a Markov chain we need to introduce the **period** of a state. To motivate the definition consider the random walk from Example 14. Assuming that the Markov chain starts in state 0 at time 0 let us consider the probabilities $P(X(n) = 0 | X(0) = 0)$ of being back in state 0 after n steps.

Clearly, $P(X(1) = 0 | X(0) = 0) = 0$ and by adding up the probability of all paths leading to $(X(2) = 0)$, we see that

$$P(X(2) = 0 | X(0) = 0) = 2p(1-p).$$

It turns out that in general we have

$$P(X(n) = 0 | X(0) = 0) = 0$$

for $n = 1, 3, 5, \ldots$ since any closed path starting and ending at state 0 must have an even number of jumps.

If we define the period of a state $i \in S$ as the greatest common divisor of the length of all possible loops starting and ending in state i, then we have demonstrated that for the random walk on $\mathbb{Z}$ then state 0 has period 2 .

The period of a Markov chain is important for the long term behaviour of $X(n)$. For the random walk on $\mathbb{Z}$ the only possible limit of

$$\lim_{n\to\infty} P(X(n) = 0|X(0) = 0)$$

is zero since every second element of the sequence is zero. Note, however, that it requires an argument to show that the limit actually exists. More generally, the period of state j turns out to be important for the study of the n-step transition probabilities $P^n_{i,j}$ as $n \to \infty$.

Definition 17 (Period of a Markov chain) *For a discrete-time Markov chain on S a loop of length n is a sequence of states $i_0, i_1, \ldots, i_n \in S$ with $i_0 = i_n$. We will speak of a possible loop if*

$$P_{i_0,i_1} \cdot P_{i_1,i_2} \cdot \ldots \cdot P_{i_{n-1},i_n} > 0.$$

Introduce

$$D_i = \{n \in \mathbb{N} |\ \text{there exists a possible loop of length } n \text{ with } i_0 = i_n = i\}$$

and define per(i) *(-the period of state i) as the largest number dividing all numbers in the set D_i.* □

Theorem 18 *All states of a communication class have the same period and we shall use the term **aperiodic** about a class of period* 1*. An irreducible, aperiodic Markov chain is a Markov-chain with one communication class of period* 1*.*

Proof: For i and j in the same communication class choose $k, l \in \mathbb{N}$ such that $P^k_{i,j} > 0$ and $P^l_{j,i} > 0$. It follows from Definition 17 that $m = l + k$ belongs to both of the sets D_i and D_j. If $n \in D_i$ (i.e. if there is a loop of length n from i to i), then there is a loop of length $l + n + k$ from j to j showing that also $n + m \in D_j$. By definition $per(j)$ divides all numbers in D_j and since both $m, n + m \in D_j$ we conclude that $per(j)$ also divides n. We have thus shown that $per(j)$ is a number that divides all $n \in D_i$. Since $per(i)$ is the greatest number dividing all $n \in D_i$ we conclude that $per(i) \geq per(j)$. The opposite inequality follows by symmetry and we conclude that $per(i) = per(j)$. □

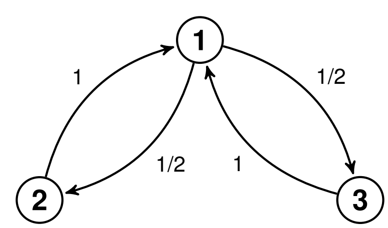


Figure 2.6: Transition diagram for a discrete-time Markov chain with three states and period 2. Direct computations in Example 19 show that $\lim_{n\to\infty} P(X(n) = j)$ only exists under certain conditions on the initial distribution $P(X(0) = i)$.

Example 19 *Figure 2.6 shows an irreducible Markov chain with period* 2*. For an arbitrary initial distribution*

$$P(X(0) = i) = \phi(i)$$

we find that

$$\begin{aligned} P(X(n) = 1) &= \begin{cases} \phi(1) & , n \text{ even} \\ \phi(2) + \phi(3) & , n \text{ odd} \end{cases} \\ P(X(n) = 2) &= \begin{cases} \frac{1}{2} \cdot \phi(1) & , n \text{ odd} \\ \frac{1}{2} \cdot (\phi(2) + \phi(3)) & , n \text{ even} \end{cases}. \end{aligned}$$

This shows that $\lim_{n\to\infty} P(X(n) = j)$ *exists if and only if*

$$\phi(1) = \phi(2) + \phi(3).$$

□

In Example 19 above we saw that the initial distribution may influence the long run behaviour for a (periodic) Markov chain. The main result on the limiting behaviour for discrete time Markov chains (Theorem 20) states that for aperiodic, recurrent Markov chains then the limit $\lim_{n\to\infty} P(X(n) = j)$ exists and do not depend on the initial distribution.

Theorem 20 *For an irreducible, recurrent and aperiodic discrete-time Markov chain then for any state i and any initial distribution it holds that*

$$\lim_{n\to\infty} P(X(n) = i) = \frac{1}{E[T_i|X(0)=i]}.$$

If $E[T_i|X(0) = i] = +\infty$ *the limit on the right hand side is defined to be* 0.

Comments on the proof: The proof of the main result may be based on the renewal theorem (-see Appendix C or Asmussen [2003, Chapter 1.2]). □

The limit in Theorem 20 involves the expectation of the return time

$$T_i = \inf\{n > 0 | X(n) = i\}$$

to state i that was introduced just before Defintion 8 on page 20 in Chapter 2 [12]. The formal definition of the expectation in Theorem 20 is

[12] We adopt the usual convention that $\inf \emptyset = +\infty$ corresponding to letting $T_i = +\infty$ if $X(n)$ never hits state i.

$$\begin{aligned} E[T_i|X(0) = i] &= \sum_{n=1}^{\infty} n \cdot P(T_i = n|X(0) = i) \\ &+ (+\infty) \cdot P(T_i = +\infty|X(0) = i) \end{aligned}$$

where

$$P(T_i = +\infty|X(0) = i)$$

describes the probability that the Markov chain never returns to state i.

If i belongs to a recurrent communication class then we know that

$$P(T_i = +\infty|X(0) = i) = 0$$

implying that we are sure to get back to state i. Note, however, that even in this case the mean return time

$$E[T_i|X(0) = i] = \sum_{n=1}^{\infty} n \cdot P(T_i = n|X(0) = i)$$

may or may not be finite.

Definition 21 (Positive recurrence and null-recurrence) *A recurrent state i is said to be **positive recurrent** if and only if the mean return time to state i is finite*

$$E[T_i|X(0) = i] < +\infty.$$

*Otherwise the recurrent state is said to be **null-recurrent** . It can be shown that all states belonging to the same recurrent class are either positive recurrent or null-recurrent* [13]. □

[13] A formal argument is given after the proof of Theorem 22 below.

It follows from Theorem 20 and Definition 21 that for an irreducible, positive recurrent and aperiodic Markov chain then the distribution of $X(n)$ has a nonzero limit. In the current formulation the result is not very useful because we are rarely able to compute $E[T_i|X(0) = i]$ implying that we can not determine if the Markov chain is positive recurrent (and much less find the limit in Theorem 20!).

From a more practical point of view we might note that by using the initial distribution $P(X(0) = i) = 1$ then Theorem 20 implies that if the Markov chain is positive recurrent then the n-step transition probabilities

$$(P^n)_{i,j} = P(X(n) = j|X(0) = i)$$

should have a limit as $n \to \infty$. Using numerical methods it might be possible to compute $(P^n)_{i,j}$ for large values of n to get an idea of the value of the limit (and whether it exists). For Markov chains on a finite state space you may use that n-step transition probabilities are obtained by multiplication of the transition matrix P by itself n times (-see Theorem 2).

It turns out that we can give a useful characterization of the limit in Theorem 20 in terms of the solution to a system of equations involving the transition probabilities. To motivative the relevant definition consider the following identity

$$\begin{aligned} & P(X(n+1) = j|X(0) = i) \\ = & \sum_{l \in S} P(X(n+1) = j, X(n) = l|X(0) = i) \\ = & \sum_{l \in S} P(X(n+1) = j|X(n) = l) \cdot P(X(n) = l|X(0) = i) \\ = & \sum_{l \in S} P(X(n) = l|X(0) = i) \cdot P_{l,j} \end{aligned}$$

which follows from the Markov property (2.1) on page 15. Assuming that the limits $\lim_{n \to \infty} P(X(n) = j|X(0) = i)$ exist then it is tempting to interchange limit and summation [14] to get

[14] The argument at least holds for Markov chains on a finite state space.

$$\begin{aligned}\pi(j) &= \lim_{n\to\infty} P(X(n+1)=j|X(0)=i) \\ &= \sum_{l\in S}\left\{\lim_{n\to\infty} P(X(n)=l|X(0)=i)\right\}\cdot P_{l,j} \\ &= \sum_{l\in S}\pi(l)P_{l,j}. \end{aligned} \tag{2.8}$$

A non-negative vector, $\overline{\pi} = (\pi(j))_{j\in S}$, solving the system of equations (2.8) for all $j \in S$ is called an **invariant measure** for the transition probabilities, $P = (P_{i,j})_{i,j\in S}$. If $\overline{\pi}$ is a probability (i.e. if all $\pi(j) \geq 0$ and $\sum_{j\in S}\pi(j) = 1$) we will speak of an **invariant distribution** for P.

We have demonstrated that (at least in certain cases) the limit of Theorem 20 must be found among the solutions to the equation (2.8) (i.e. among the class of invariant measures for P). Note that if $\overline{\pi}$ solves (2.8) then any vector obtained by multiplying $\overline{\pi}$ (coordinate wise) by a constant will also be a solution. Further there might be multiple invariant measures (not proportional!) in which case we have to be careful because only one of these can be related to the limit $\lim_{n\to\infty} P(X(n)=i)$.

The following result states that for irreducible and recurrent Markov chains then there is a unique invariant measure [15].

[15] For a Markov chain on a finite state space, $|S| < +\infty$, then the vector $(1,\ldots,1)^T$ is a (right) eigenvector with eigenvalue 1 for the transition matrix P. In particular, the transposed matrix P^T has a (right) eigenvector $\overline{v}$ with eigenvalue 1 satisfying the equation $P^T\overline{v} = \overline{v}$. With $\overline{\pi} = \overline{v}^T$ we conclude that $\overline{\pi}P = \overline{\pi}$. If $\pi(i) > 0$ and state j is accessible from i (i.e. $P^m_{i,j} > 0$ for som $m > 0$) then

$$\pi(j) = \sum_l \pi(l)P^m_{l,j} \geq \pi(i)P^m_{i,j} > 0.$$

It follows that for irreducible finite state Markov chains then we can find a solution to (2.9) with $\pi(j) > 0$ for all $j \in S$. After normalisation we have an invariant probability vector. Now, by Theorem 22 all invariant measures are proportional in particular they will all have finite mass. Consequently, irreducible, recurrent Markov chains on a finite state space will always be positive recurrent.

Theorem 22 (Existence and uniqueness of invariant measures) *For an irreducible, recurrent Markov chain, $(X(n))_{n\geq 0}$, there is a unique (up to multiplication!) invariant measure solving the equations*

$$\nu(j) = \sum_{i\in S}\nu(i)P_{i,j}, \quad j\in S. \tag{2.9}$$

The unique solution (up to multiplication) is given by

$$\nu(j) = E\left[\sum_{n=0}^{T_i-1} 1(X(n)=j)|X(0)=i\right] \tag{2.10}$$

where $i \in S$ is any fixed state. The solution can be normalized into an unique invariant probability if and only if $E[T_i|X(0)=i] < +\infty$ i.e. if and only if the Markov chain is positive recurrent. .

Sketch of proof: Note that since i is recurrent then

$$P(T_i < +\infty|X(0)=i) = 1.$$

Consequently

$$\begin{aligned}\nu(j) &= E\left[\sum_{n=0}^{T_i-1} 1(X(n)=j)|X(0)=i\right] \\ &= E\left[\sum_{n=1}^{T_i} 1(X(n)=j)|X(0)=i\right] \\ &= E\left[\sum_{n=1}^{\infty} 1(X(n)=j, X(n-1),\ldots,X(1)\neq i|X(0)=i\right]. \end{aligned}$$

Taking the sum outside the expectation and using the Markov property it now follows that $(\nu(j))_{j\in S}$ is a solution to (2.9). The same trick (-interchanging sum and expection) shows that the total mass of

$$\sum_{j\in S}\nu(j) = E[T_i|X(0)=i]$$

in particular positive recurrence (c.f. Definition 21) of i implies that the invariant measure $\{\nu(i)\}_{i\in S}$ can be normalized into an invariant distribution.

We skip the proof of the uniqueness part, however, more details may be found in questions 6.-10. of Exercise C.3 in Appendix C. □

We are allowed to use two different fixed states $i, \tilde{i} \in S$ to construct the invariant measures in Theorem 22. The total mass of the measures are $E[T_i|X(0)=i]$ and $E\left[T_{\tilde{i}}|X(0)=\tilde{i}\right]$. Since the measures are unique up to multiplication we conclude that either both measures are finite or both measures are infinite. This demonstrates that states in a recurrent communication class are either all positive recurrent or all null-recurrent which was postulated in Definition 21.

We are now in a position to gather and reformulate the above results into the main result concerning the limiting behaviour for discrete time Markov chains.

Theorem 23 *For an irreducible, aperiodic and positive recurrent Markov chain it holds that*

$$\lim_{n\to\infty} P(X(n)=j) = \pi(j) = \frac{1}{E[T_j|X(0)=j]}$$

where $\overline{\pi} = (\pi(j))_{j\in S}$ *is the unique invariant probability vector solving the system of equations*

$$\pi(j) = \sum_{i\in S}\pi(i)P_{i,j}.$$

Proof: We know from Theorem 22 that there exists an invariant distribution $\overline{\pi}$ if and only if $E[T_i|X(0)=i] < +\infty$ for some state i and that the solution is unique. The unique invariant probability may be represented as

$$\pi(j) = \frac{\nu(j)}{E[T_i|X(0)=i]}$$

where i is any fixed state in S. In particular, choosing $i=j$ then we have $\nu(j) = 1$ implying that $\pi(j) = \frac{1}{E[T_j|X(0)=j]}$. The result now follows from Theorem 20. □

The results above cover the limiting behaviour of $P(X(n)=j)$ for aperiodic and positive recurrent states j [16] A little more can be said concerning the limit for null-recurrent or transient states.

[16] Exercises 2.3.2 and 2.4.6 demonstrate the use of Theorem 23.

Theorem 24 (Limit result for null-recurrent states) *For a null-recurrent state j it holds that*

$$\lim_{n\to\infty} P(X(n)=j) = 0$$

for any choice of initial distribution.

Proof: The proof may be found in Asmussen [2003, Chapter 1]. □

Theorem 25 (Limit result for transient states) *For a transient state j it holds that*

$$\lim_{n\to\infty} P(X(n) = j) = 0$$

for any choice of initial distribution.

Proof: Since j is transient it follows from (2.6) on page 22 in the proof of Theorem 10 that $\sum_{n=1}^{\infty}(P^n)_{i,j} < +\infty$. In particular, for any $i \in S$ it holds that

$$P(X(n) = j | X(0) = i) = (P^n)_{i,j} \overset{n\to\infty}{\to} 0.$$

For an arbitrary initial distribution $\phi(i) = P(X(0) = i)$ we therefore get by dominated convergence that

$$P(X(n) = j) = \sum_{i\in S} P(X(n) = j | X(0) = i)\phi(i) \to 0,$$

for $n \to \infty$. □

The identity (2.8) may be iterated to

$$\pi(j) = \sum_{l\in S} \pi(l)(P^n)_{lj}$$

allowing us to conclude (use dominated convergence!) that $\pi(j) = 0$ for a null-recurrent or transient state j for any invariant probability vector. In particular, you can just put $\pi(j) = 0$ when you write down the system of equations for an invariant proabability which may save you a lot of time.

We finally state a result explaining the limiting behaviour of irreducible Markov chains with period larger than one (i.e. not aperiodic Markov chains).

Theorem 26 (Limit result for periodic states) *For an irreducible Markov chain with period $d > 1$ the limit $\lim_{n\to\infty} P(X(n) = i)$ does not exist for an arbitrary initial distribution. However, the average distribution over a period of length d has a limit*

$$\lim_{n\to\infty} \frac{P(X(n) = i) + P(X(n+1) = i) + \ldots + P(X(n+d-1) = i)}{d}.$$

If the limit (-denoted $\pi(i)$) describes a probability distribution (i.e. if $\sum_{i\in S} \pi(i) = 1$) then $\overline{\pi} = (\pi(i))_{i\in S}$ is a unique invariant distribution for the Markov chain.

Proof: The proof may be found in Jacobsen and Keiding [1985]. □

Example 27 (Example 19 continued) *The three state Markov chain in Example 19 (-see Figure 2.6) has period 2 and we saw that existence of a limit $\lim_{n\to\infty} P(X(n) = j)$ depends strongly on the choice of initial distribution $P(X(0) = j)$. It may be shown (try!) that the Markov chain has an invariant distribution given by $\overline{\pi} = (1/2, 1/4, 1/4)$. In particular, it follows from Theorem 26 that*

$$\lim_{n\to\infty} \frac{P(X(n) = j) + P(X(n+1) = j)}{2} = \pi(j)$$

for an arbitrary initial distribution. The message here is that to obtain a limit for a Markov chain with period d we have to average the distribution of $\{X(n)\}_{n\in\mathbb{N}_0}$ over d consecutive time points, if the result should hold independently of the initial distribution. Note however also that Example 19 *demonstrates that even for periodic Markov chains there may be other initial distributions than the invariant for which the limit* $\lim_{n\to\infty} P(X(n) = j)$ *does exists.* □

If the Markov chain is not irreducible then we may apply the results of this section to each recurrent communication class. For each of the positive recurrent classes there exists a unique invariant distribution with positive probabilities only for the states in the class. However, any convex combination of the invariant distribution on the positive recurrent subclasses constitutes an invariant probability on the entire state space of the Markov chain. In particular, in this case the invariant probability distribution is no longer unique (-see Exercise 2.3.1).

Absorption probabilities

Recurrent classes are closed: once the Markov chain enters a recurrent class then it stays there forever. Transient classes may or may not be closed but in either case we know from Theorem 25 that $\lim_{n\to\infty} P(X(n) = j) = 0$ for any transient state j. If a transient class is not closed then the Markov chain is not irreducible and there can be several other recurrent or transient classes [17].

[17] The random walk on $\mathbb{Z}$ with parameter $p \neq 1/2$ is an example of a Markov chain containing only one closed transient class. In this case the interpretation of Theorem 25 is that the probability mass gets spread out over infinitely many transient states as $n \to \infty$.

This naturally raises the following questions: if a Markov chain with multiple communication classes (i.e. not irreducible) is started in a transient state i in class $\mathcal{T}$, how many times will it visit state i before it leaves the state for good, and what is the probability that it will leave class $\mathcal{T}$ by jumping to any of the other communication classes of the Markov chain?

Theorem 28 (Absorption probabilities - finite state space) *Consider a finite state Markov chain with transition matrix P. Suppose that the states are ordered such that P can be decomposed as a block matrix*

$$P = \left(\begin{array}{c|c} \tilde{P} & 0 \\ \hline S & Q \end{array}\right)$$

where $\tilde{P}$ is the transition matrix restricted to the recurrent states. Similarly, Q is the submatrix of P restricted to the transient states, and S describes transition probabilities from transient to recurrent states. The 0 block in the upper right part of P reflects the fact that transitions from recurrent to transient states are not possible.

The ij-th entry of the matrix $M = (I - Q)^{-1}$ describes the expected number of visits to the transient state j before the Markov chain reaches one of the recurrent states given that the Markov chain starts in the transient *state i (i.e. $P(X(0)=i)=1$). Here, I denotes the identity matrix with zero off-diagonal and a diagonal of ones.*

The ij-th entry of

$$A = (I-Q)^{-1}S$$

is the probability that j is the first recurrent state reached by the Markov chain when started in the transient state i (i.e. $P(X(0)=i)=1$).

Proof: Let $\mathcal{C}$ be the set of recurrent states and let

$$\tau = \inf\{n>0|X(n)\in\mathcal{C}\}$$

be the time of the first visit to a recurrent state. Let $(\tilde{X}(n))_{n\geq 0}$ be an auxiliary Markov chain with a modified transition matrix, P_{mod}, obtained by letting all recurrent states be absorbing (i.e. $P_{i,i}=1$ for all $i\in\mathcal{C}$). Note that for any event that does only involve the time before and including τ then we may substitute the original Markov chain with $(\tilde{X}(n))_{n\geq 0}$.

The number, N_j, of visits to a transient state j before τ may be expressed as [18]

[18] There is a slight abuse of notation below, which you may notice if you are familiar with the notation of random variables.

$$N_j = \sum_{n=0}^{\infty} 1(\tilde{X}(n)=j, \tilde{X}(n-1),\ldots,\tilde{X}(0)\notin\mathcal{C}) = \sum_{n=0}^{\infty} 1(\tilde{X}(n)=j),$$

where the last equality follows from the fact that we can only have $\tilde{X}(n)=j$ if the Markov chain stays in the transient states from time 0 to $n-1$.

Now, by direct computation we have

$$P_{mod} = \left(\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array}\right) \Rightarrow P_{mod}^n = \left(\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array}\right)^n = \left(\begin{array}{c|c} I & 0 \\ \hline \left(\sum_{i=0}^{n-1} Q^i\right) S & Q^n \end{array}\right)$$

where we use the notation $Q^0 = I$. Taking expectations and using the formula (2.3) on page 17 for the distribution of $\tilde{X}(n)$ we get that

$$\begin{aligned} E[N_j|X(0)=i] &= \sum_{n=0}^{\infty} P(\tilde{X}(n)=j|\tilde{X}(0)=i) \\ &= \sum_{n=0}^{\infty} (P_{mod}^n)_{i,j} = \sum_{n=0}^{\infty} (Q^n)_{i,j}. \end{aligned}$$

Introducing the notation $R=\sum_{n=0}^{\infty}(Q^n)$ then we observe that [19]

[19] It requires a formal argument to show that all entries of R are finite: use Theorem 10 to deduce that since j is transient then $\sum_{n=0}^{\infty}(Q^n)_{j,j} < +\infty$. Then look carefully into the proof of Theorem 10 to locate an inequality ensuring that also $\sum_{n=0}^{\infty}(Q^n)_{i,j} < +\infty$ for any other state i.

$$R = I + Q(\sum_{n=0}^{\infty} Q^n) = I + QR$$

in particular we have that $R=(I-Q)^{-1}$. This demonstrates that given $X(0)=i$ then the expected number of visits to state j before absorption in a recurrent class may be found as the ij-th entry of $M = (I - Q)^{-1}$.

For a recurrent state $j \in \mathcal{C}$ then the event $(X(\tau) = j)$ involves only the behaviour of the Markov chain up to time τ. In particular, we may compute the probability $P(X(\tau) = j)$ by substituting $(X(n))_{n \geq 0}$ with $(\tilde{X}(n))_{n \geq 0}$. Consequently we have for any recurrent state $j \in \mathcal{C}$ that

$$
\begin{aligned}
& P(X(\tau) = j | X(0) = i) \\
= \; & P(\tilde{X}(\tau) = j | \tilde{X}(0) = i) \\
= \; & \sum_{n=0}^{\infty} \sum_{l \notin \mathcal{C}} P(\tilde{X}(n) = l, \tilde{X}(n+1) = j | \tilde{X}(0) = i) \\
= \; & \sum_{n=0}^{\infty} \sum_{l \notin \mathcal{C}} P(\tilde{X}(n) = l | \tilde{X}(0) = i) P(\tilde{X}(n+1) = j | \tilde{X}(n) = l) \\
= \; & \sum_{n=0}^{\infty} \sum_{l \notin \mathcal{C}} (P_{mod}^n)_{i,l} \, (P_{mod})_{l,j} \\
= \; & \sum_{n=0}^{\infty} \sum_{l \notin \mathcal{C}} (Q^n)_{i,l} \, S_{l,j} = (RS)_{i,j} = \Big((I - Q)^{-1} S \Big)_{i,j} .
\end{aligned}
$$

□

Theorem 28 is often applied to the situation where all recurrent classes are sets with only one element. In this situation all recurrent states are **absorbing states** and we have $\tilde{P} = I$. The result then gives us the absorption probabilities for each of the absorbing states.

If the Markov chain contains more than one recurrent class then Theorem 28 may be used to compute the probability that the Markov chain will end its life in each of the recurrent classes. Note that in the long-run the total probability of being absorbed in a particular recurrent class will be redistributed on individual states according to the invariant distribution restricted to the relevant class (at least if the class is aperiodic).

Example 29 (Absorption probabilities) *The purpose of the present example is to illustrate various applications of Theorem 28 on a Markov chain, $(X(n))_{n \geq 0}$, with a finite state space $S = \{1, 2, 3, 4, 5, 6\}$* [20]*. The transition diagram for the Markov chain is displayed in Figure 2.7.*

We want to answer the following questions

1. *Find the communication classes for $(X(n))_{n \geq 0}$.*

We assume that the Markov chain starts in state 5 and define the time

$$\tau = \inf\{n > 0 | X(n) \neq 3, 4, 5\}$$

of the first exit from states 3, 4, 5.

[20] To get some practice using Theorem 28 try to do Exercise 2.3.8

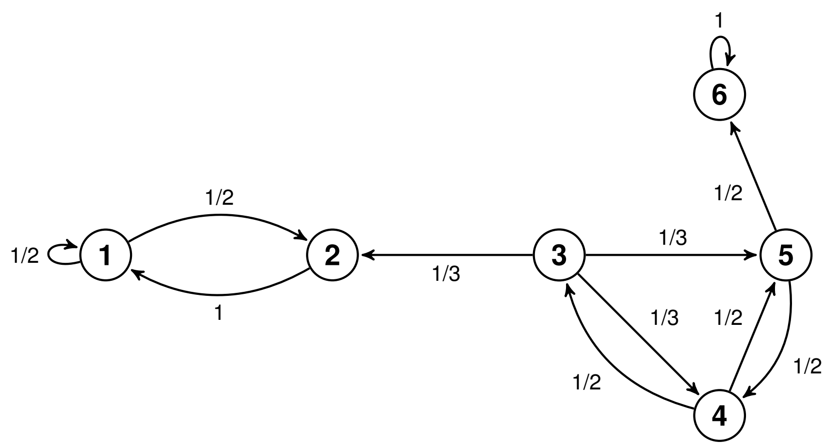


Figure 2.7: Transition diagram for the discrete-time Markov chain with six states considered in Example 29.

2. *Find $P(X(\tau) = i)$ for $i = 1, 2, 6$.*

3. *Compute the mean number of visits to state 4 before τ.*

4. *Find the probability that the first visit to state 6 happens before the first visit to state 4.*

5. *Find the probability that the first visit to state 6 happens before the first visit to state 3.*

There are three communication classes given by $C_1 = \{1,2\}$, $C_2 = \{3,4,5\}$ and $C_3 = \{6\}$. C_1 and C_3 are finite, closed classes and hence positive recurrent according to the note in the margin of page 34. C_2 is transient since there is a positive probability of leaving class C_2 without ever getting back.

The solution to questions 2. and 3. may be obtained by straightforward application of Theorem 28. We reorder the states in the order $\{1,2,6,3,4,5\}$ and decompose the transtion matrix as

$$P = \left(\begin{array}{c|c} \tilde{P} & 0 \\ \hline S & Q \end{array} \right)$$

where

$$\tilde{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad S = \begin{pmatrix} 0 & 1/3 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \quad Q = \begin{pmatrix} 0 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}.$$

Then we compute the matrices

$$A = (I - Q)^{-1} S = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/3 & 2/3 \\ 0 & 1/6 & 5/6 \end{pmatrix} \qquad M = (I - Q)^{-1} = \begin{pmatrix} 3/2 & 1 & 1 \\ 1 & 2 & 4/3 \\ 1/2 & 1 & 5/3 \end{pmatrix}.$$

From the row of A corresponding to the initial state 5 *we get that*

$$P(X(\tau)=i)=\begin{cases} 0 & ,i=1 \\ 1/6 & ,i=2 \\ 5/6 & ,i=6 \end{cases}.$$

The $(3,2)$*-th entry of M (-corresponding to states* $(5,4)$*) gives the answer to question 3. The expected number of visits to state* 4 *before* τ *is* 1 *given that we start the Markov chain at* $X(0)=5$.

By looking at the transition diagram on Figure 2.7 we see that the only way that state 4 *may be reached before state* 6 *is if the first jump goes from state* 5 *to state* 4*. Thus the answer to question 4. is* $1/2$.

The answer to question 5. requires a little more work. One possibility is to write down all possible paths of the Markov chain for which the chain will visit state 3 *before state* 6 *and add up the probabilities for all these paths. Another way to get the solution to question 5. is to consider a modified version of the Markov chain that follows the original chain* $(X(n))_{n\geq 0}$ *until the first visit of either state* 3 *or state* 6*. This may be done by changing the transition matrix*

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

to P_{mod} *as follows*

$$P_{mod}=\begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

such that states 3 *and* 6 *become absorbing. Using the decompostion in Theorem 28 on* P_{mod} *(-after reordering the states as* $\{1,2,3,6,4,5\}$ *) we get that*

$$S_{mod}=\begin{pmatrix} 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \quad Q_{mod}=\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

The 2*-th row (-corresponding to state* 5*) of*

$$A_{mod}=(I-Q_{mod})^{-1}S_{mod}=\begin{pmatrix} 0 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

describes the distribution of the value of the Markov chain at the time of the first visit to the set $\{1,2,3,6\}$. □

Theorem 28 is stated in terms of matrix operations and is therefore restricted to Markov chains on a finite state space. For Markov chains on a countable state space absorption probabilities may be found by solving a countably infinite system of equations as explained in Theorem 30.

Theorem 30 (Absorption probabilities - countable state space) *For a Markov chain on S let $\mathcal{C}$ be a recurrent class. The probabilities*

$$\alpha(j) = P(X(n) \in \mathcal{C} \text{ for some } n \geq 0 | X(0) = j), \quad j \in \mathcal{C}' \leftarrow \text{ transient states}$$

that the chain will ever visit class $\mathcal{C}$ (and stay there forever) then solves the system of equations

$$\alpha(j) = \sum_{l \in \mathcal{C}'} P_{j,l}\alpha(l) + \sum_{l \in \mathcal{C}} P_{j,l}. \tag{2.11}$$

The absorption probability $(\alpha(j))_{j \in \mathcal{C}'}$ is the smallest non-negative solution to (2.11). There is a unique bounded solution to (2.11) if and only if there is zero probability that the Markov chain stays in the transient states forever.

Comments on the proof: The first part of the theorem has a rather intuitive content. Given that the Markov chain starts at some transient state $X(0) = j$ then it may enter the closed recurrent class $\mathcal{C}$ in two different ways. It may either jump directly to a state $l \in \mathcal{C}$ (with probability $P_{j,l}$), or it may jump to another transient state $l \in \mathcal{C}'$ (with probability $P_{j,l}$) and then eventually move from state l to the class $\mathcal{C}$ (with probability $\alpha(l)$). Due to the Markov property the latter probabilities may be multiplied yielding the identity (2.11) [21].

[21] Note the we do not need to consider the possibility that the Markov chain jumps from the initial state j to some (recurrent) state $l \neq \mathcal{C} \cup \mathcal{C}'$ since then $\alpha(l) = 0$ and there will be no contribution to (2.11).

The difficult part of the proof is to characterize the absorption probability as the smallest non-negative solution to (2.11). Example 31 below describes a situation where there are multiple non-negative solutions. In particular, the latter part of the theorem is crucial if we want to compute absorption probabilities for Markov chains with infinitely many transient states.

The complete proof of Theorem 30 may be found in Jacobsen and Keiding [1985]. □

Example 31 (Random walk with absorption at zero) *In this example we consider a Markov chain, $(X(n))_{n \geq 0}$, on $\mathbb{N}_0$ with transition probabilities given by*

$$P_{i,i+1} = 1 - P_{i,i-1} = p_i \in (0,1), \quad i \geq 1, \quad P_{0,0} = 1.$$

Note that the Markov chain has two communication classes given by $\mathbb{N}$ and $\{0\}$ where the latter is an absorbing state. If we define the first hitting time of state i by

$$T_i = \inf\{n > 0 | X(n) = i\}$$

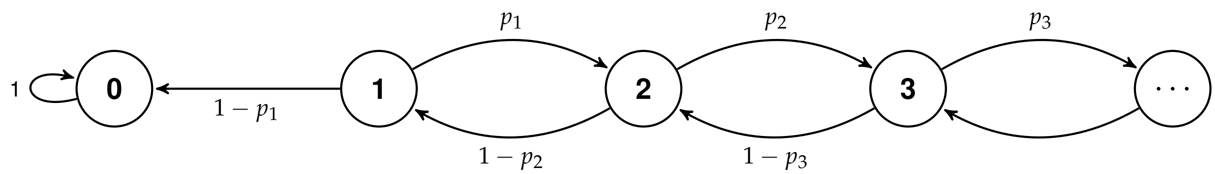


Figure 2.8: The absorption probabilities for the absorbing random walk on $\mathbb{N}_0$ considered in Example 31 may be computed using Theorem 30.

then we have for $i \geq 1$ *that*

$$P(T_i = +\infty | X(0) = i) \geq (1-p_i)(1-p_{i-1}) \cdot \ldots \cdot (1-p_1) > 0$$

which shows that state i is transient according to Definition 8.

We consider the probability

$$\alpha(i) := P(T_0 < +\infty | X(0) = i)$$

that the Markov chain will ever be absorbed in state 0 *given that* $X(0) = i$*. Using Theorem 30 then* $(a(i))_{i \geq 1}$ *may be characterised as the smallest, nonnegative solution to the system of equations given by*

$$\alpha(i) = P_{i,0} + \sum_{j \neq 0} P_{i,j}\alpha(j) = \begin{cases} (1-p_1) + p_1 \cdot \alpha(2) & ,i = 1 \\ (1-p_i) \cdot \alpha(i-1) + p_i \cdot \alpha(i+1) & ,i > 1. \end{cases}$$

One has to be very careful to get all steps of the argumentation correct when applying Theorem 30. Therefore you are encouraged to go very slowly through the remaining part of this example.

We restrict our attention to the case where all $p_i = p \in (0,1)$ *(constant). In this case (2.11) writes out as*

$$\begin{aligned} \alpha(j) &= (1-p)\alpha(j-1) + p\alpha(j+1), \quad j \geq 2 \\ \alpha(1) &= p\alpha(2) + 1 - p, \end{aligned}$$

and we may use the result on linear recurrence equations in Appendix B.5 to find the set of solutions. Referring to the notation in the Appendix B.5 we get

$$\alpha_1 = \frac{1 + \sqrt{1-4p(1-p)}}{2p}, \quad \alpha_2 = \frac{1 - \sqrt{1-4p(1-p)}}{2p}.$$

The square root is zero for $p = 1/2$ *and for* $p \neq 1/2$ *the two solutions turn out to be* 1 *and* $\frac{1-p}{p}$*.*

For $p = 1/2$ *the set of solutions is given by*

$$\alpha(i) = c_1 + c_2 \cdot i, i \geq 1,$$

with the boundary condition (corresponding to $j = 1$*)*

$$c_1 + c_2 \cdot 1 = 1/2 + 1/2(c_1 + c_2 \cdot 2)$$

implying that $c_1 = 1$. *For the solution*

$$\alpha(i) = 1 + c_2 \cdot i, \quad i \geq 1,$$

to be nonnegative we must have $c_2 \geq 0$ *and the smallest solution is obtained by letting* $c_2 = 0$. *We conclude from Theorem 30 that for* $p = 1/2$ *then* $\alpha(i) = 1, i \geq 1$, *that is the Markov chain will be absorbed at* 0 *with probability* 1.

For $p \neq 1/2$ *we get from Appendix B.5 that any solution must have the form*

$$\alpha(i) = c_1 + c_2 \cdot \left(\frac{1-p}{p}\right)^i, i \geq 1,$$

with the additional boundary condition that $\alpha(1) = (1-p) + p\alpha(2)$ *from where one may deduce that* $c_1 = 1 - c_2$. *The resulting expression for* $\alpha(i)$ *now looks like*

$$\alpha(i) = (1 - c_2) \cdot 1 + c_2 \cdot \left(\frac{1-p}{p}\right)^i, i \geq 1.$$

For $\frac{1-p}{p} > 1$ *then the solution will only be nonnegative for* $c_2 \geq 0$ *and we see that the smallest solution is actually obtained for* $c_2 = 0$ *where we have that* $\alpha(i) = 1, i \geq 1$. *We conclude from Theorem 30 that the Markov chain will be absorbed in state* 0 *with probability* 1 *if* $\frac{1-p}{p} > 1$.

For the remaining case $\frac{1-p}{p} < 1$ *we observe that* $\alpha(i) \to 1 - c_2$ *for* $i \to \infty$ *hence the solution will be nonnegative only if* $1 - c_2 \geq 0$. *On the other hand from the formula*

$$\alpha(i) = (1 - c_2) \cdot 1 + c_2 \left(\tfrac{1-p}{p}\right)^i$$

we observe that the smallest, nonnegative solution is obtained for $c_2 = 1$. *We conclude that for* $\frac{1-p}{p} < 1$ *then the probability that the Markov chain will ever be absorbed in state* 0 *given that* $X(0) = i$ *becomes*

$$\alpha(i) = \left(\tfrac{1-p}{p}\right)^i, i \geq 1.$$

□

We end this chapter by discussing how to determine the limiting behaviour for Markov chains that are *not* irreducible. We know from Theorem 25 that $\lim_{n\to\infty} P(X(n) = j) = 0$ for any transient state $j \in \mathcal{T}$. For a recurrent state j ($\in \mathcal{C}_1$ say) then we may denote by

$$\tau = \inf\{n \geq 0 | X(n) \in \mathcal{C}_1\}$$

the time of the first visit to class $\mathcal{C}_1$. By splitting the event $(X(n) = j)$ according to the time $k = 0, \ldots, n$ and the state $X(\tau) = i \in \mathcal{C}_1$ of the

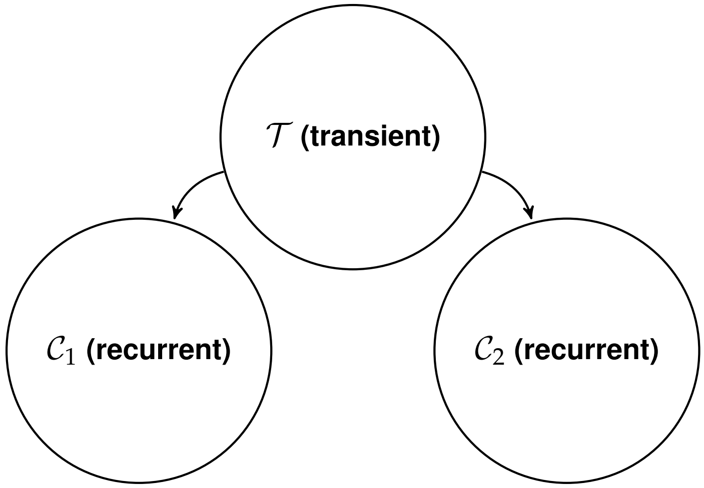


Figure 2.9: Generic picture of a Markov chain with two recurrent classes $\mathcal{C}_1$ and $\mathcal{C}_2$. The set of transient states are denoted by $\mathcal{T}$ and may consist of several communication classes.

first visit to class $\mathcal{C}_1$ then we get (j is fixed below!)

$$\begin{aligned}
& P(X(n) = j) \\
= & \sum_{k=0}^{n} \sum_{i \in \mathcal{C}_1} P(X(n) = j, \tau = k, X(\tau) = i) \\
= & \sum_{k=0}^{\infty} 1(k \le n) \sum_{i \in \mathcal{C}_1} P(X(n) = j, X(k) = i, X(k-1), \dots, X(0) \notin \mathcal{C}_1) \\
= & \sum_{k=0}^{\infty} 1(k \le n) \sum_{i \in \mathcal{C}_1} \{P(X(n) = j | X(k) = i, X(k-1), \dots, X(0) \notin \mathcal{C}_1) \\
& \qquad \times P(X(k) = i, X(k-1), \dots, X(0) \notin \mathcal{C}_1)\} \\
= & \sum_{k=0}^{\infty} 1(k \le n) \sum_{i \in \mathcal{C}_1} P(X(n) = j | X(k) = i) P(X(k) = i, X(k-1), \dots, X(0) \notin \mathcal{C}_1) \\
= & \sum_{k=0}^{\infty} 1(k \le n) \sum_{i \in \mathcal{C}_1} P_{i,j}^{n-k} P(X(k) = i, X(k-1), \dots, X(0) \notin \mathcal{C}_1) \\
= & \sum_{k=0}^{\infty} \sum_{i \in \mathcal{C}_1} \underbrace{1(k \le n) P_{i,j}^{n-k}}_{:= f_n(k,i)} P(X(\tau) = i, \tau = k).
\end{aligned}$$

If we let ν denote the measure on the discrete set $\mathbb{N}_0 \times \mathcal{C}_1$ with density

$$\nu(\{(k,i)\}) = P(X(\tau) = i, \tau = k)$$

then the formula above may be written as

$$P(x(n) = j) \quad = \sum_{(k,i) \in \mathbb{N}_0 \times \mathcal{C}_1} f_n(k,i) v(\{(k,i)\}) = \int f_n d\nu. \quad (2.12)$$

Now, if $\mathcal{C}_1$ is null-recurrent then we know from Theorem 24 that $\lim_{n\to\infty} P_{i,j}^n = 0$ for all $i, j \in \mathcal{C}_1$. In particular we have that

$$\lim_{n\to\infty} f_n(k,i) = 0, \quad (k,i) \in \mathbb{N}_0 \times \mathcal{C}_1.$$

Similarly, by Theorem 20 for $\mathcal{C}_1$ aperiodic and positive recurrent then

$$\lim_{n\to\infty} f_n(k,i) = \pi(j), \quad (k,i) \in \mathbb{N}_0 \times \mathcal{C}_1,$$

where $\overline{\pi} = (\pi(j))_{j\in\mathcal{C}_1}$ is the unique invariant distribution on $\mathcal{C}_1$.

Interchanging limit and summation in (2.12) (formally using dominated convergence) then we get

$$\lim_{n\to\infty} P(X(n) = j) = 0$$

if j is null-recurrent and

$$\lim_{n\to\infty} P(X(n) = j) = \pi(j) \sum_{(k,i)\in\mathbb{N}_0\times\mathcal{C}_1} \nu(\{(k,i)\}) = \pi(j)P(\tau < +\infty)$$

if j is aperiodic and positive recurrent. The interpretation of the latter expression is that if the Markov chain gets absorbed in class $\mathcal{C}_1$ then the absorption probability $P(\tau < +\infty)$ will in the long run be distributed on $\mathcal{C}_1$ according to the invariant probability $\overline{\pi}$ on $\mathcal{C}_1$.

We emphasize that for a positive recurrent class with period larger than one then no general limit result holds.

Exercises on Markov chains in discrete time

Markov chains with two states

2.1.1 General two state Markov chain

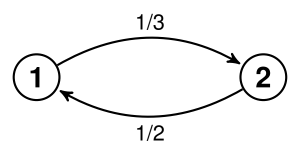


Figure 2.10: Transition diagram for Markov chain in Exercise 2.1.1 and 2.1.2. Note that loops are sometimes omitted from a transition diagram. You are supposed to deduce from the figure that $P_{1,1} = 2/3$ and $P_{2,2} = 1/2$.

We consider a Markov chain on $S = \{1, 2\}$ with transition diagram given on Figure 2.10. We assume that $X(0) = 1$.

1. Write down the transition matrix P of the Markov chain.
2. Compute $P(X(1) = 1)$ and $P(X(1) = 2)$.
3. Find the distribution of $X(2)$ and $X(3)$.
4. Compute P^2 and P^3 and compare with the results of questions 2.-3.
5. Compute also P^{10}.

Assume in the following that the initial distribution of $X(0)$ is given by $\phi(1) = \phi(2) = 1/2$.

6. Compute the distribution of $X(1)$.
7. Let $\overline{\phi} = (\phi(1), \phi(2))$ and compute $\overline{\phi}P, \overline{\phi}P^2$, and $\overline{\phi}P^3$. What did you actually compute?
8. Find the distribution of $X(5)$.
9. Find the invariant probability vector $\overline{\pi} = (\pi_1, \pi_2)$ of the Markov chain by solving the matrix equation $\overline{\pi}P = \overline{\pi}$ that may be written out as

$$\pi_1 P_{11} + \pi_2 P_{21} = \pi_1 \quad \text{and} \quad \pi_1 P_{12} + \pi_2 P_{22} = \pi_2.$$

10. Compare the results of questions 5., 8., and 9.

2.1.2 *Recurrence times*

We consider again the Markov chain of Exercise 2.1.1 given by Figure 2.10. Assume that $X(0) = 1$ and define the *recurrence time* to state 1 by

$$T_1 = \inf\{n > 0 | X(n) = 1\}.$$

The purpose of this exercise is to study the distribution of the recurrence time and its relation to the invariant probability vector of the Markov chain.

1. Find $P(T_1 = 1)$.
2. Compute $P(T_1 = 2), P(T_1 = 3)$ and find the general expression for $P(T_1 = n)$, $n \geq 2$.
3. Find the mean recurrence time $\mu_1 = E[T_1]$ to state 1.

Assume that $X(0) = 2$ and define the *recurrence time* to state 2 by

$$T_2 = \inf\{n > 0 | X(n) = 2\}.$$

4. Compute $P(T_2 = n), n \geq 2$ and the mean $\mu_2 = E[T_2]$.
5. Find $1/\mu_1$ and $1/\mu_2$.
6. Compare the results of question 5. with the invariant probability vector $\overline{\pi}$ found in question 9. of exercise 2.1.1.

2.1.3 *Two state absorbing Markov chain*

We consider in this exercise the two-state Markov chain, $(X(n))_{n \geq 0}$, given by the transition diagram on Figure 2.11. When discussing the Markov chain further we refer to the states through the following recoding: *1=alive,0=dead*.

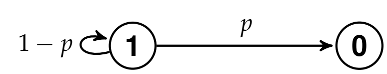


Figure 2.11: Transition diagram for Markov chain in Exercise 2.1.3

1. Write down the transition matrix for the Markov chain.
2. Assuming that $X(0) = alive$ find the probabilities $P(X(n) = alive)$ for $n \geq 1$.

Define the survival time, T, as the time of absorption in the state *dead*

$$T = \inf\{n > 0 | X(n) = dead\}.$$

3. Argue that $P(T \leq n) = P(X(n) = dead)$.
4. Find the distribution of T, i.e. $P(T = n)$ for $n \geq 1$. What is the name of the distribution of T?

   Hint: What is the name of the distribution of $\tilde{T} = T - 1$*?*
5. Compute the expected survival time $E[T]$.

2.1.4 *Transition probabilities for the two-state chain*

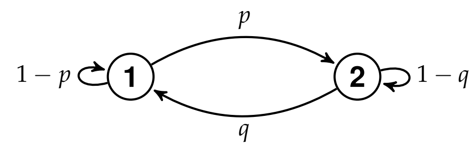


Figure 2.12: Transition diagram of a general Markov chain with two states as in Exercise 2.1.4.

Consider the general two-state Markov chain given by transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

for $p, q \in [0,1]$. The purpose of the exercise is to derive closed form expressions for the n step transition probabilities given by the matrix P^n.

1. Draw the transition diagram for the Markov chain.

2. Compute the characteristic polynomial for P given by

   $$g(\lambda) = \det(P - \lambda I)$$

   where I is the 2×2 identity matrix.

3. Argue that the equation $g(\lambda) = 0$ has two solutions

   $$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = 1 - p - q.$$

4. Find a (right) eigenvector $\overline{v} = (v_1, v_2)^t$ for P associated with the eigenvalue λ_2. This means that you should solve the system of equations $P\overline{v} = (1-p-q)\overline{v}$.

5. Show that $\overline{u} = (u_1, u_2)^t = (1,1)$ is a (right) eigenvector for P with eigenvalue 1.

6. Verify that the matrix

   $$O = \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$$

   satisfies the matrix equation

   $$PO = O \underbrace{\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}}_{=D}.$$

7. Find the inverse matrix O^{-1}.

8. Use that $P = ODO^{-1}$ to find a closed form expression for P^n and discuss the result.

Markov chains with three states

2.2.1 *Classification of states*

Consider a general three-state Markov chain as given by the transition diagram of Figure 2.13.

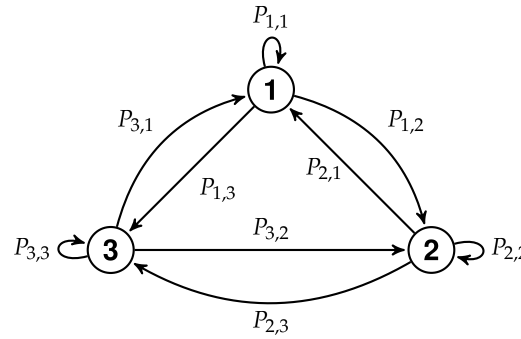


Figure 2.13: Transition diagram for a general three-state Markov chain.

1. Argue that the chain is irreducible if $P_{i,j} > 0$, for all $i \neq j$.
2. Give examples of irreducible three-state Markov chains for which $P_{i,j} = 0$ for at least one pair (i, j) of states.
3. Give examples of a three-state Markov chain with two communication classes.
4. Describe the relation between zero entries of the transition matrix P and the communication classes of the Markov chain. In each case determine if the communication classes are transient or recurrent.

2.2.2 *General three state Markov chain*

Consider a Markov chain given by the transition diagram on Figure 2.14. Assume that $X(0) = 1$ and let

$$\tau = \inf\{n > 0 | X(n) \neq 1\}$$

be the time of the first jump away from state 1.

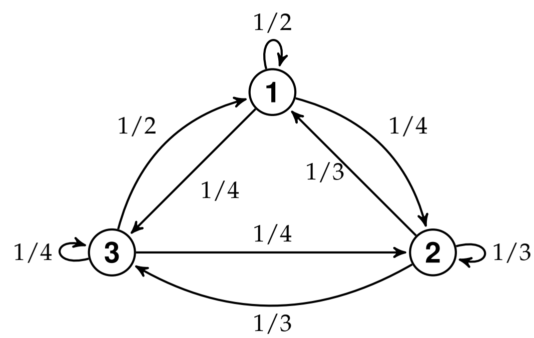


Figure 2.14: Transition diagram for the Markov chain considered in Exercise 2.2.2.

1. Find the transition matrix, P, of the chain.
2. Compute $P(X(1) = 1)$ and $P(X(2) = 1)$.
3. Use P^3, P^4, P^5 to find $P(X(n) = 1)$ for $n = 3, 4, 5$.
4. Find $P(\tau = 1), P(\tau = 2)$, and $P(\tau = 3)$. What is the name of the distribution of $\tau - 1$?
5. Write down the system of equations for the invariant distribution $\overline{\pi}$ and find $\overline{\pi}$.

2.2.3 *The one-way Markov chain*

We consider in this exercise a Markov chain given by transition diagram on Figure 2.15 under the assumption that $X(0) = 1$.

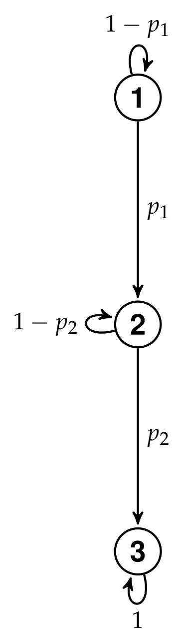


Figure 2.15: Transition diagram on the one-way Markov chain of Exercise 2.2.3.

1. Find the probabilities $P(X(1) = j)$ for $j = 1, 2, 3$.
2. Find the probabilities $P(X(2) = j)$ for $j = 1, 2, 3$.

Denote by

$$\tau_1 := \inf\{n > 0 | X(n) = 2\}$$

the time of the jump between states 1 and 2. Similarly let

$$\tau_2 = \inf\{n > \tau_1 | X(n) = 3\}$$

between the time of the jump from state 2 to state 3.

3. Find the probabilities $P(\tau_1 = k)$ for $k \geq 1$. What is the name of the distribution of τ_1?

4. Find the probabilities $P(\tau_2 = k)$ for $k = 1, 2, 3$.

5. Try to find the general formula for $P(\tau_2 = k)$ for $k \geq 1$.

6. Assuming that $p_1 = p_2$ verify that $\tau_2 - 2$ follows a negative binomial distribution.

2.2.4 *Markov property under aggregation of states*

Consider a Markov chain given by the transition diagram on Figure 2.16.

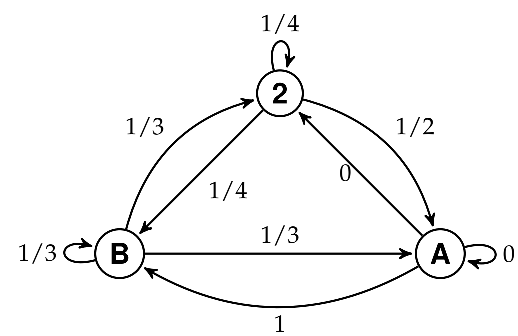


Figure 2.16: Transition diagram for the Markov chain in Exercise 2.2.4.

1. Find the transition matrix, P, for the Markov chain.

2. Write down the system of equations for the invariant distribution $\overline{\pi}$ and find $\overline{\pi}$.

Assume that the initial distribution is given by

$$P(X(0) = A) = P(X(0) = B) = P(X(0) = 2) = 1/3.$$

Use the transition matrix, P, of the chain to compute the following probabilities

3. $P(X(1) = A, X(0) = i)$ for $i \in \{A, B, 2\}$.

4. $P(X(1) = B, X(0) = i)$ for $i \in \{A, B, 2\}$.

5. $P(X(2) = 2, X(1) = A, X(0) = i)$ for $i \in \{A, B, 2\}$.

6. $P(X(2) = 2, X(1) = B, X(0) = i)$ for $i \in \{A, B, 2\}$.

Suppose that for some reason we are not able to distinguish between states A and B such that we only observe the process defined by

$$Y(n) = \begin{cases} 2 & , X(n) = 2 \\ 1 & , X(n) \in \{A, B\} \end{cases}$$

with state space $S = \{1, 2\}$

7. Use questions 3.-6. to compute $P(Y(2) = 2 | Y(1) = 1, Y(0) = 1)$.

8. Use questions 3.-6. to compute $P(Y(2) = 2|Y(1) = 1, Y(0) = 2)$.

9. Argue that $(Y(n))_{n\in\mathbb{N}_0}$ is not a Markov chain.

10. Show by an example that for certain choices of the transition probabilities for $(X(n))_{n\geq 0}$ it holds that $(Y(n))_{n\geq 0}$ is a Markov chain on $S = \{1, 2\}$.

Finite state space

2.3.1 Find the communication classes

In this Exercise we consider the Markov chain given by the transition diagram in Figure 2.17.

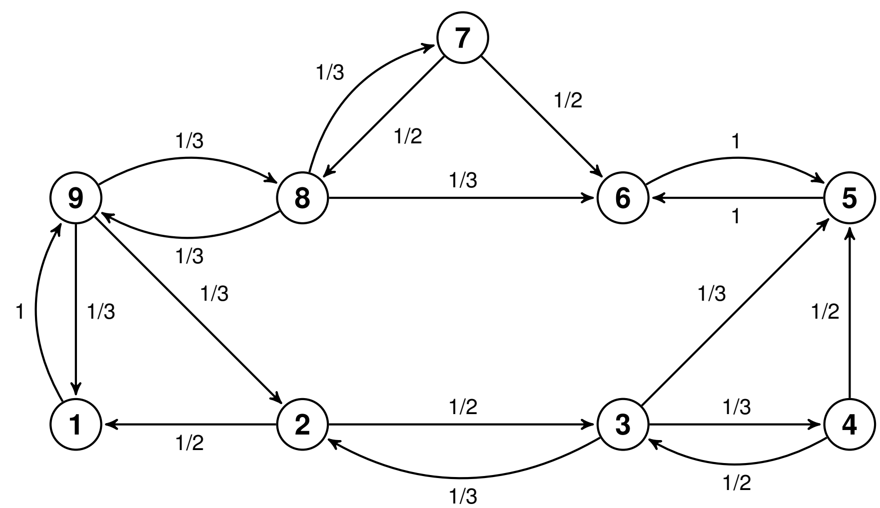


Figure 2.17: Transition diagram for the discrete-time Markov chain with seven states considered in Exercise 2.3.1.

1. Argue that 7 and 8 belong to the same communication class.

2. Show that $P^2_{2,9} > 0$ and argue that 2 and 9 belong to the same communication class.

3. Find out if states 3 and 7 communicate.

4. Determine the communication class containing state 5.

5. Find all communication classes and determine if each class is recurrent or transient.

6. Is the chain irreducible?

The **loop trick** is a useful observation to speed up the process of determining the communication classes of a Markov chain. The basic observation is that if we can find a closed path of states

$$i_0 \to i_1 \to \ldots \to i_k \to i_0$$

such that all transition probabilities along the path are positive then all states in the path belong to the same communication class.

7. Use the loop trick to find the communication classes of the Markov chain.

8. Argue that there does exist an invariant probability vector, $\overline{\pi}$, for the chain and find it.

9. Suppose that we change the transition probabilities such that

$$P_{7,6} = 0, P_{7,8} = 1, P_{2,3} = 0, P_{2,1} = 1, P_{8,7} = 1/2, P_{8,9} = 1/2, P_{8,6} = 0.$$

Show that the modified version of the Markov chain has two recurrent subclasses.

10. Find an invariant probability vector, $\overline{\pi}$, for the Markov chain described in question 9. and discuss if $\overline{\pi}$ is uniquely determined.

2.3.2 *A numeric example*

Consider the Markov chain given by the transition diagram on Figure 2.18 and assume that $X(0) = 3$.

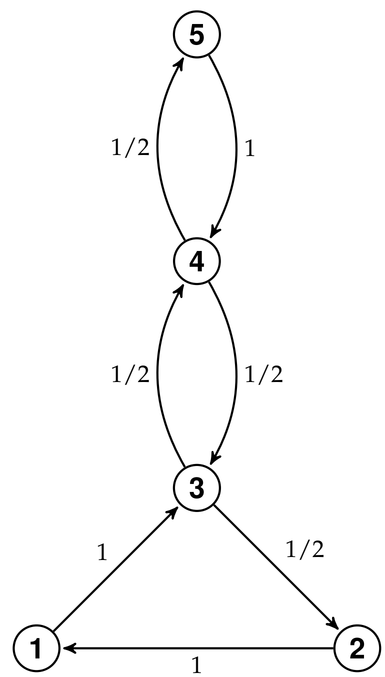


Figure 2.18: Transition diagram for Markov chain in Exercise 2.3.2.

1. Write down the transition matrix, P, of the Markov chain.

2. Find $P(X(k) = 3)$ for $k = 1, 2, 3, 4$.

3. What is the period of all recurrent communication classes of the Markov chain?

4. Compute P^2, P^4, P^8, and P^{16}.

5. Argue that the Markov chain has an invariant distribution, $\overline{\pi}$, and find this.

6. Let $T_3 := \inf\{k \geq 1 | X(k) = 3\}$ be the first time the Markov chain visits state 3. Compute $P(T_3 = k)$, $k = 1, 2, 3, 4$ and try to find the entire distribution of T_3.

7. Compute the mean, $E[T_3]$, of the return time to state 3 and compare with the invariant distribution $\overline{\pi}$ of question 5.

2.3.3 *Two component repair system*

Consider a technical device with two states *broken* and *functioning*. Suppose that every day there is a fixed probability p that the device breaks down. Every morning the state of the device is inspected and if it is broken it is replaced the following morning. Denote by $X(n)$ the state of the device on day n. Clearly, the process $(X(n))_{n\geq 0}$ is a Markov chain.

1. Find the state space and the transition matrix P and draw the transition diagram.

2. Compute the invariant probability distribution $\overline{\pi}$ and find the long term fraction of time where the device is broken.

Consider now a system consisiting of two devices (working independently of each other) that can both take the values *broken* and *functioning*. Every day there is probability p_1 and p_2 of the individual devices breaking down. Every morning the system is inspected and the following morning the broken devices (if any) are replaced. The state of the system on the morning of day n can be described by a Markov chain with the four states

(broken,broken),(broken,funct.),(funct.,broken),(funct.,funct.)

Assume throughout the exercise that no device is broken on the morning of day $n = 0$. To ease notation we recode the state space as *0=broken,1=functioning*.

3. Find the possible transitions of the four state Markov chain and draw the transition diagram of the chain without transition probabilities.

4. Compute the distribution of $X(1)$ i.e. find $P(X(1) = (i, j)), i, j = 0, 1$.

5. Find the transition matrix of the Markov chain.

6. Let $\pi_{i,j} = \lim_{n\to\infty} P(X(n) = (i, j)), i, j = 0, 1$, be the limiting distribution of $X(n)$. Show that
$$\pi_{0,0} = p_1 p_2 \pi_{1,1}.$$

7. Write down a similar equation as the one in question 6. for each of the probabilities $\pi_{1,0}, \pi_{0,1}, \pi_{1,1}$.

8. Show that the solution to the system of equations in question 6.-7. is given by
$$\pi_{i,j} = \frac{p_1^{1-i} p_2^{1-j}}{(1+p_1)(1+p_2)}, \quad i, j = 0, 1.$$

9. Suppose that it is criticial to a production company that at least one of the individual devices is functioning since otherwise the production of the company ceases and all workers are sent home. What is the long run probability that the production must be stopped and how often does it happen (on average) that both devices break down and workers are sent home?

The company now changes its policy and decides no longer to replace a broken device as long as the other is still working.

10. Draw the transition diagram (with transition probabilities) corresponding to the new replacement strategy.
11. Write down an equation for the invariant probability $\pi_{1,1}$ and show that $\pi_{0,0} = (p_1 + p_2 - p_1 p_2)\pi_{1,1}$.
12. Write down a similar equation as the one in question 11. for each of the probabilities $\pi_{0,0}, \pi_{1,0}, \pi_{0,1}$.
13. Solve the system of equations in question 11.-12.
14. Answer question 9. for the Markov chain corresponding to the new replacement strategy of the company.

2.3.4 *Random walk reflected at two barriers*

In this exercise we consider a Markov chain, $(X(n))$, on the state space $\{0, 1, \ldots, N\}$ where only transitions between neighbouring states i and $i+1$ or i and $i-1$ are possible. When the Markov chain reaches the boundary 0 it stays there with probability $1-p$ and is otherwise reflected to state 1. At the upper boundary N the chain stays with probability p as is reflected to state $N-1$ with probability $1-p$. The transition diagram is given by Figure 2.19 and we assume that $X(0) = 0$.

1. Find the transition matrix, P, of the Markov chain.
2. Compute $P(X(1) = 0)$ and $P(X(1) = 1)$. What is the name of the distribution of $X(1)$?
3. Compute $P(X(2) = k)$, for $k = 0, 1, \ldots, N$.
4. Argue that there exists an invariant probability vector, $\overline{\pi}$, and write down the system of equations that should be satisfied by $\overline{\pi} = (\overline{\pi}_0, \overline{\pi}_1, \ldots, \overline{\pi}_N)$.
5. Argue that a vector of the form

$$\pi_i = c\left(\frac{p}{1-p}\right)^i, i = 0, 1, \ldots, N,$$

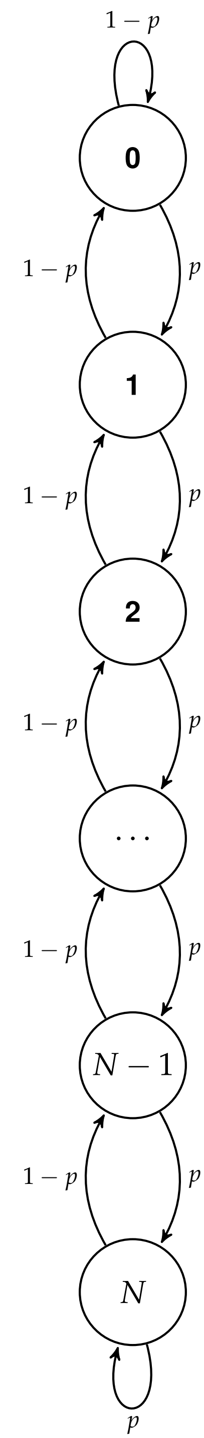


Figure 2.19: Transition diagram for the Random walk reflected at two barriers considered in Exercise 2.3.4.

satisfies the system of equations from question 4. and find the constant c that turns $\overline{\pi}$ into a probability.

The purpose of the following questions is to find a simple expression for the n step transition matrix P^n for the case of $N = 2$ where the state space is $S = \{0,1,2\}$. For this particular case the transition matrix takes the form

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

It might be that you can guess the formula for P^n by looking at the expressions for P^2, P^3, and P^4. Another possibility is to follow the strategy outlined below.

6. Compute the characteristic polynomial $g(\lambda) = \det(P - \lambda I)$ of P.

7. Verify that $g(\lambda) = 0$ has three real valued solutions λ_1, λ_2, and λ_3.

8. For each of the eigenvalues $\lambda_k, k = 1,2,3$, above find an eigenvector, v_k, for P with eigenvalue λ_k.

9. Let O be the 3×3 matrix with rows $\overline{v}_1, \overline{v}_2, \overline{v}_3$ and verify that

$$OP = \underbrace{\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}}_{=D} O.$$

10. Show that $P^n = O^{-1}D^nO$ and try to get a closed form expression of $P(X(n) = k)$ for $k \geq 1$ under the initial condition that $P(X(0) = 0) = 1$.

    Warning: maybe it is not worth spending too much time finding a closed form expression for O^{-1}.

2.3.5 Yahtzee

Yathzee is a dice game. The objective of the game is to score the most points by rolling five dice to make certain combinations. The dice can be rolled up to three times in a turn. After the first two rolls the player can save any dice that are needed to complete a combination and then re-roll the other dice. A Yahtzee is five-of-a-kind and holds the game's highest point value of 50.

The purpose of the present exercise is to compute the probability of ending up with a Yahtzee given that we use the strategy that maximizes the number-of-a-kind after each roll. To simplify the problem we consider initially in questions 1.-8. the probability of obtaining a Yathzee of five sixes. We deal with the general problem in questions 9.-17.

The problem may be put into the framework of Markov chains by defining a stochastic process as follows

- Let $X(0) = 0$, i.e. $P(X(0) = 0) = 1$.
- Roll five dice and let $X(1)$ denote the number of sixes.
- Define $X(n+1)$ recursively by the following rule.
  - If $X(n) = 5$ then $X(n+1) = 5$.
  - If $X(n) < 5$ then we let Y be the number of sixes after re-rolling the $5 - X(n)$ dice and the value at time $n+1$ may be expressed as $X(n+1) = X(n) + Y$.

1. Argue briefly that $(X(n))_{n \in \mathbb{N}_0}$ is a Markov chain and write down the set, S, of possible states for the chain.
2. Find the distribution of $X(1)$ and explain which entries of the transition matrix P that correspond to the probabilities
$$P(X(1) = k), k \in S.$$
3. Find the distribution of $X(2)$ given that $X(1) = 4$.
4. Find the distribution of $X(2)$ given that $X(1) = 3$.
5. Write down the entire transition matrix, P, of the Markov chain.
6. Compute P^2 and the probability $P(X(2) = 5)$.
7. Find the probability $P(X(3) = 5)$.
8. Use $P, P^2, P^3, \ldots$ to compute a numerical approximation to the expected number of rolls, $E\tau_5$, where
$$\tau_5 = \inf\{n > 0 | X(n) = 5\}$$
denotes the time before the Markov chain is absorbed in state 5.

The solve the original problem posed above not restricting our selves to a Yathzee of sixes we need to modify the definition of the Markov chain above. More precisely after each roll we need to allow the player to switch from saving only dice with face six to dice with another number of eyes if this is more favorable. For example if we have two-of-a-kind after n rolls (i.e. $X(n) = 2$) and the next roll results in $5 - 2 = 3$ three dice of a different kind then we let $X(n+1) = 3$ and only 2 dice are re-rolled.

The purpose of the following questions is to compute the transition matrix for the modified version, $\tilde{P}$, of the game. For all questions below compare the result to the relevant entry of the transition matrix for the original version of the game. Questions 13.-15. below are probably the most difficult.

9. Find the probabilities $P(X(1) = k), k = 3, 4, 5$.

10. Find the probabilities $P(X(2) = k|X(1) = j), k \in S$, for $j = 3, 4, 5$.

11. Find the probabilities $P(X(2) = k|X(1) = 2), k = 4, 5$.

12. Find the probabilities $P(X(2) = 5|X(1) = 1)$.

13. Find the probabilities $P(X(1) = k), k = 0, 1, 2$.

14. Find the probabilities $P(X(2) = k|X(1) = 2), k = 2, 3$.

15. Find the probabilities $P(X(2) = k|X(1) = 1), k = 1, 2, 3, 4$.

16. Write down the entire transition matrix $\tilde{P}$ and compute $\tilde{P}^2$ and $\tilde{P}^3$.

17. Find $P(X(2) = 5), P(X(3) = 5)$, and the mean $E\tau_5$ and compare with the results in questions 6.-8.

2.3.6 *Markov chain with two regimes*

Consider the 5 state Markov chain, $(X(n))_{n\in\mathbb{N}_0}$, with transition diagram given by Figure 2.20 where $p_A, p_B \in (0, 1)$.

We further define the stochastic process $(Y(n))_{n\in\mathbb{N}_0}$ defined by

$$Y(n) = \begin{cases} A & , \quad X(n) = \text{A1,A2,A3} \\ B & , \quad X(n) = \text{B1,B2}. \end{cases}$$

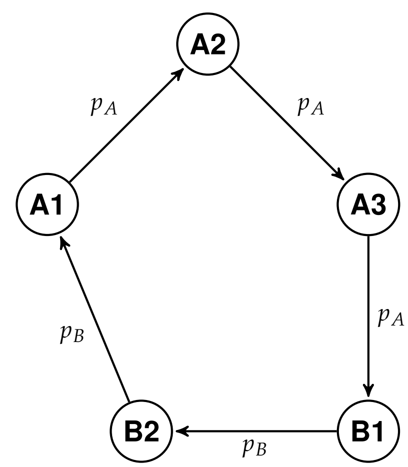


Figure 2.20: Transition diagram for the Markov chain with two regimes considered in Exercise 2.3.6. States $A1, A2, A3$ constitute one regime and states $B1, B2$ is another regime.

In the following we will study the properties of the stochastic process $\{Y(n)\}_{n\in\mathbb{N}_0}$ on the state space $S = \{A, B\}$ that keeps track of the regime.

1. Find the conditional distribution of $Y(n+1)$ given that $Y(n) = B$ and $Y(n-1) = A$.

2. Argue that the conditional distribtuion of $Y(n+1)$ given that $Y(n) = Y(n-1) = B$ is different from the result of question 1.

Assume that we know that $P(Y(0) = A) = 1$. Clearly

$$P(Y(0) = A) = \underbrace{P(X(0) = A1)}_{=\phi_{A1}} + \underbrace{P(X(0) = A2)}_{=\phi_{A2}} + \underbrace{P(X(0) = A3)}_{=\phi_{A3}}$$

but if we only observe $(Y(n))_{n\in\mathbb{N}_0}$ we do not know $\phi_{A1}, \phi_{A2}, \phi_{A3}$.

3. Let

$$\tau_B = \inf\{n > 0 | Y(n) = B\}$$

be the time of the first jump to state B. Express the probabilities $P(\tau_B = k)$ for $k = 1, 2, 3$ in terms of p_A, p_B and $\phi_{A1}, \phi_{A2}, \phi_{A3}$.

4. What should be the distribution of τ_B if $(Y(n))_{n\in\mathbb{N}_0}$ was a Markov chain on $\{A, B\}$ with initial distribution $P(Y(0) = A) = 1$ and transition matrix

$$P = \begin{pmatrix} 1-q_A & q_A \\ q_B & 1-q_B \end{pmatrix}?$$

5. Use questions 1.-4. to discuss whether $(Y(n))_{n\in\mathbb{N}_0}$ is a Markov chain on $S = \{A, B\}$.

2.3.7 *Periodicity of a Markov chain*

1. Show that states 1-3 belong to the same communication class (-see Figure 2.21 on page 59).

2. Show that states 10-12 belong to the same communication class.

3. Determine the communication class containing state 12.

4. Argue that states 1 and 6 do not belong to the same communication class.

5. Find all the disjoint communication classes in the partition of the state space. For each class determine whether the class is recurrent or transient.

6. Find the period of each communication class.

7. How would the communication classes be and what would be the period of the chain under the following changes: $P_{12,11} = 1/2 = P_{12,12}$

8. How would the communication classes be and what would be the period of the chain under the following changes: $P_{9,8} = 1/2 = P_{9,6}$

10. How would the communication classes be and what would be the period of the chain under the following changes: $P_{8,5} = P_{8,7} = P_{8,11} = 1/3$

11. Does there exist a unique invariant probability distribution for Markov chain on Figure 2.21?

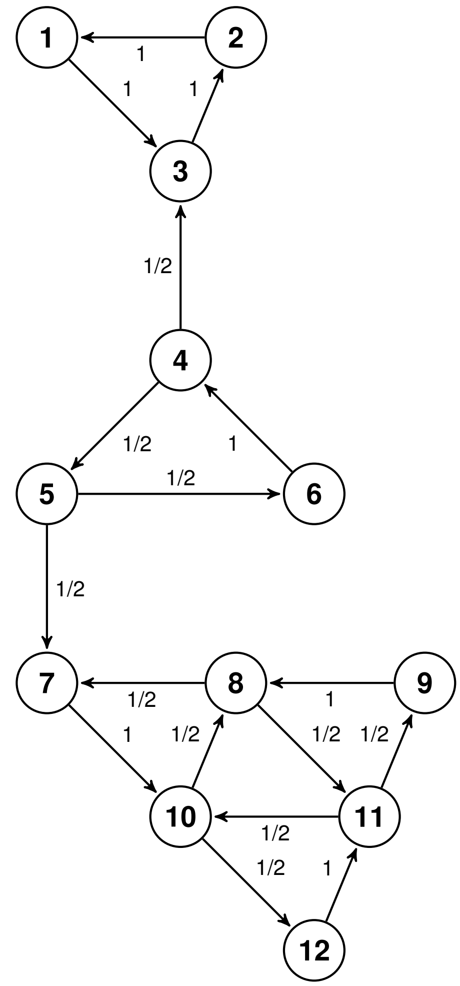


Figure 2.21: Transition diagram for questions 1.-11. of Exercise 2.3.7.

For the rest of the exercise we modify the Markov chain in Figure 2.21 by changing the following transition probabilities:

$$P_{1,1} = P_{1,3} = 1/2 = P_{12,11} = P_{12,12}.$$

12. Argue that with the modified transition probabilities then all recurrent subclasses are aperiodic.

13. Find an invariant probability vector that is concentrated on each of the recurrent subclasses.

14. Assuming that the initial distribution of the chain is given by $P(X(0) = 1) = 1$ find the limiting distribution of $X(n)$ that is find

$$\lim_{n\to\infty} P(X(n) = i), \quad i = 1, \ldots, 12.$$

15. Find the limiting distribution of $X(n)$ for the initial distribution $P(X(0) = 12) = 1$.

16. Find the limiting distribution of $X(n)$ for the initial distribution $P(X(0) = 6) = 1$.

    Hint: start by computing the probability that the first jump from state 4 *to* 3 *occurs before the first jump between states* 5 *and* 7.

2.3.8 More about absorption probabilities

Consider a Markov chain on $S = \{0, 1, 2, 3, 4, 5, 6\}$ with transition probabilities

$$\begin{aligned}
P_{0,0} &= 3/4, \quad P_{0,1} = 1/4 \\
P_{1,0} &= 1/2, \quad P_{1,1} = P_{1,2} = 1/4 \\
P_{j,0} &= P_{j,j-1} = P_{j,j} = P_{j,j+1} = 1/4, \quad j = 2,3,4,5 \\
P_{6,0} &= 1/4, \quad P_{6,5} = 1/4, \quad P_{6,6} = 1/2
\end{aligned}$$

1. Is the Markov chain irreducible?

2. Is the Markov chain aperiodic?

3. What is the long-run probability of observing the sequence of states $4 \to 5 \to 0$?

4. For $X(0) = 1$ what is the probability of reaching state 6 before state 0?

5. For $X(0) = 3$ what is the expected number of steps until the chain is in state 3 again?

6. For $X(0) = 0$ what is the expected number of steps until the chain is in state 6?

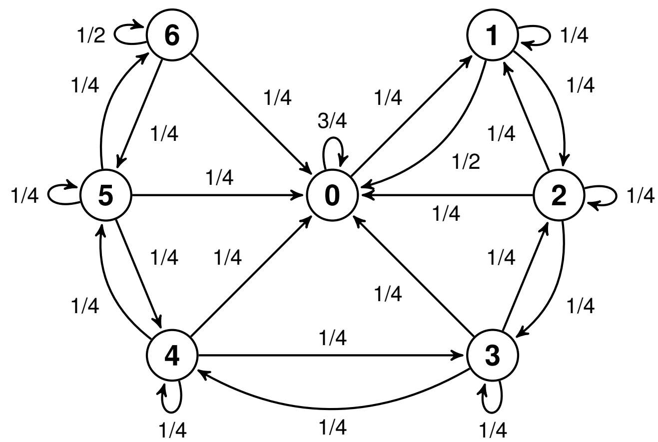


Figure 2.22: Transition diagram for the discrete-time Markov chain with seven states considered in Exercise 2.3.8.

Countable state space

2.4.1 Queueing system

Markov chains are very popular as models for the number of customers in a queueing system. In this exercise we consider the so-called single server queue. Assume that no customers are present in the queue at time 0 i.e. $P(X(0) = 0) = 1$. In each time period (=step) there is probability $p \in (0, 1)$ that a new customer arrives and probability $q \in (0, 1)$ that the service of the customer at the service desk is completed. We denote by $X(n)$ the total number of customers in the queueing system at time n and note that this is a Markov chain on $\mathbb{N}_0$. The transition probabilities of the chain is given by the infinite transition matrix $P = (P_{i,j})_{i,j\geq 0}$.

1. Find $P_{0,1}$ and $P_{0,0}$.

2. Argue that $P_{1,1} = pq + (1 - p)(1 - q)$ and find $P_{1,0}, P_{1,2}$.

3. Use question 2. to find $P_{i,i-1}, P_{i,i}, P_{i,i+1}$ for $i > 1$ and draw the transition diagram of the Markov chain.

4. Find the communication classes.

5. For a vector $\overline{\pi} = (\pi_0, \pi_1, \pi_2, \dots)$ to be an invariant distribution it must satisfy the system of equations $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{i,j}, j \geq 0$. Write out the equation for $j = 0$ and deduce that $\pi_1 = \frac{p}{q(1-p)} \pi_0$.

6. Write down the equations for $\pi_j, j \geq 1$.

In the following questions 7.-10. we assume that $p = q$.

7. Show that for $p = q$ then $\pi_j = c_0 + c_1 \cdot j, j \geq 1$, ($c_0, c_1$ constants) solves the system of equations from question 6 (for $j \geq 2$). [One may show that any solution takes this form.]

8. Find a condition on the constants c_0, c_1 that ensures that the solution π_j from question 7. is bounded for $j \geq 1$.

9. Does there exist an invariant probability vector for the chain if $p = q$?

10. Discuss whether we have showed that the chain is positive recurrent, null recurrent, or transient for $p = q$?

For the remaining part of the exercise we consider the general case where $p \neq q$.

11. Argue that $\pi_j = c_0 + c_1 \cdot \left(\frac{p(1-q)}{q(1-p)}\right)^j, j \geq 1$, solve the system of equations from question 6. (for $j \geq 2$) for any choice of the constants c_0, c_1.

12. Use the equation from question 5. to express π_0 in terms of p, q, and the two constants c_0, c_1.

13. Determine when the chain is positive recurrent and find the invariant probability vector $\overline{\pi}$.

15. Give a complete description of when (i.e. for what conditions on p and q) the chain is transient, null recurrent, or positive recurrent.

    Warning: this probably requires a little work!

2.4.2 *Random walk on $\mathbb{Z}$*

Consider the random walk, $(X(n))_{n\geq 0}$, on $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ given by the transition diagram [22]

[22] Note that the random walk on $\mathbb{Z}$ is also discussed in Example 14.

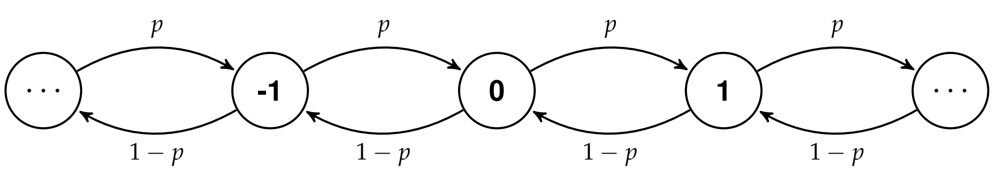


Figure 2.23: Transition diagram for a random walk on $\mathbb{Z}$ from Exercise 2.4.2.

Assume that $X(0) = 0$.

1. What is the period of the chain?

2. Find the distribution of $X(1)$ and $X(2)$.

3. Compute $P(X(2) = 0), P(X(3) = 0)$, and $P(X(4) = 0)$.

Note that $(X(2k) = 0)$ if and only if there is exactly k upward jumps and k downward jumps among the first $2k$ jumps.

4. Argue that $P(X(2k) = 0) = \binom{2k}{k} p^k (1-p)^k$, $k \geq 1$.

5. Determine if $\sum_{n=1}^{\infty} P(X(n) = 0)$ is convergent and use this to decide if the random walk on $\mathbb{Z}$ is recurrent or transient.

2.4.3 *Random walk on $\mathbb{Z}^2$*

We now generalise Exercise 2.4.2 above and consider the symmetric random walk on the pairs of integers $\mathbb{Z} \times \mathbb{Z}$. More precisely, if the chain is in state (i, j) at time n then it jumps to either of the states

$$(i, j+1), (i, j-1), (i-1, j), (i+1, j)$$

with equal probabilities $(= 1/4)$ in step $n+1$.

1. Draw (a part of) the transition diagram.
2. Argue that state $(0, 0)$ communicates with any other state and deduce that there is only one communication class.
3. Assuming that $P(X(0) = (0, 0)) = 1$ compute $P(X(n) = (0, 0))$ for $n = 1, 2, 3, 4$.
4. What is the period of the Markov chain.
5. Still assuming that $P(X(0) = (0, 0)) = 1$ argue that

$$P(X(2n) = 0) = \sum_{k=0}^{n} \frac{(2n)!}{(n-k)! \cdot (n-k)! \cdot k! \cdot k!} 4^{-2n}.$$

6. Use question 5. to determine if $\sum_{n=0}^{\infty} P(X(2n) = 0)$ is convergent and deduce if the random walk on $\mathbb{Z} \times \mathbb{Z}$ is recurrent or transient.

   Hint: Use Stirling's formula to determine the asymptotic behaviour of the terms in the sum.

2.4.4 *Random walk on $\mathbb{Z}^d$*

It is a challenging exercise to determine if the extension of the random walk in the previous exercise to $\mathbb{Z}^d$ is recurrent or transient. The dynamics of the $d-$dimensional random walk is desribed by the fact that the process moves from state $(i_1, \ldots, i_d)$ to any of the $2d$ neighbouring states given by

$$(i_1, \ldots, i_{l-1}, i_l + j, i_{l+1}, \ldots, i_d)$$

where $j \in \{-1, 1\}$ with equal probability $(= 1/(2d))$.

1. Argue that the symmetric random walk on $\mathbb{Z}^d$ has period 2.
2. Assuming that the random walk starts in state $(0, \ldots, 0)$ at time 0 argue that

$$P(X(2n) = (0, \ldots, 0)) = \sum_{k_1, \ldots, k_d \in \mathbb{N}_0 : k_1 + \ldots + k_d = n} \frac{(2n)!}{(k_1! \cdot \ldots \cdot k_d!)^2} (2d)^{-2n}.$$

3. Show that $\sum_{n=1}^{\infty} P(X(n) = (0, \ldots, 0)) < \infty$ for $d > 2$.

   Hint: This is hard. A possible solution strategy is to obtain an upper bound for the multinomial coefficient and then apply Stirling's formula.

4. Deduce from question 3. that the symmetric random walk on $\mathbb{Z}^d$ is transient for $d > 2$.

2.4.5 *Branching processes*

In this exercise we consider a model for the number, $X(n)$, of individuals in a population at time n. During each time interval (generation) each individual (independently of each other) produces a number, Z, of offsprings described by a probability distribution on $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ with density $P(Z = k) = p_k$. Note that it is also possible for an individual to die without giving birth to any offspring if $Z = 0$. The total number of individuals born in $n-$th generation, $(X(n))_{n \geq 0}$, is a Markov chain on $\mathbb{N}_0$

The important parameter for the large-time term behaviour of a branching process is the mean number of offsprings produced by an individual

$$\mu = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k p_k.$$

Not surprisingly one can prove that if $\mu > 1$ then the expected population size increases to infinity and that $\mu < 1$ implies that the population will eventually die out.

In this exercise we consider the (rather trivial) branching process with offspring distribution given by

$$p_1 = p, \quad p_0 = 1 - p,$$

with probability parameter $p \in (0, 1)$. The interpretation is that each individual gives birth to one offspring with probability p while there is a probability of $1 - p$ that no offspring is generated. Assume that we start out with a population of size $X(0) = N > 0$.

1. Find the probability $P(X(1) = N)$ and $P(X(1) = N - 1)$.

2. Argue that $X(1)$ follows a binomial distribution and find the integral parameter and the probability parameter.

3. Use 1.-2. to find the transition probabilities

$$P_{N,j} = P(X(n+1) = j | X(n) = N), j = 0, 1, \ldots, N.$$

4. Find an expression for the transition probability $P_{i,j}, i, j \in \mathbb{N}_0$.

5. Compute $EX(1)$ and give a (heuristic) argument that the population will eventually die out.

6. Show that $\sum_{n=1}^{\infty}(P^n)_{i,i} < +\infty$ for $i = 1,\dots,n$, and deduce that $\lim_{n\to\infty} P(X(n) = 0) = 1$.

   Hint: Find the communication classes and conclude that state i, $i = 1,\dots,n$, (and its communication class) is transient.

2.4.6 Positive recurrence and null-recurrence

The following exercise is greatly inspired by Exercise 2.2 in Lawler [2006].

We consider a Markov chain, $(X(n))_{n\geq 0}$, on $S = \{0,1,2,\dots\}$ with transition probabilities

$$P_{0,i} = p_i > 0, i > 0, \quad P_{i+1,i} = 1, i \geq 0, \quad P_{i,j} = 0 \text{ otherwise}$$

where $(p_i)_{i\in S}$ is a probability vector (i.e. $\sum_i p_i = 1$). Define the return time to state 0

$$T = \inf\{n > 0 | X(n) = 0\}.$$

1. Draw the transition diagram of the Markov chain.

2. Find the communication classes. Is the chain irreducible?

3. Compute
$$P(T = k | X(0) = 0), \quad k \geq 0$$
   and argue that the Markov chain is recurrent.

4. What is the condition for the Markov chain to be null-recurrent or positive recurrent?

5. Find the invariant probability vector assuming that the Markov chain is positive recurrent.

6. Consider the time of the first visit to state 10
$$T_{10} = \inf\{n > 0 | X(n) = 10\}.$$
   What is the expected return time to state 10
$$E[T_{10} | X(0) = 10]$$
   given that the Markov chain starts in state 10?

3 Markov chains in continuous time

The purpose of this chapter is to present the theory for continuous-time Markov chains on finite or countable state spaces. The result is a compromise trying to keep as many mathematical details as possible while still keeping the technical level at a suitable level for an introductory course on stochastic processes [1]. You should try to keep focus on trying to learn how to use the theory on examples and exercises and not so much on which mathematical aspects of the theory that have been left out.

[1] For a more formal and complete exposition of the theory for continuous-time Markov chains we can recommend the lecture notes by Jacobsen and Keiding [1985]. Grown up students may benefit from reading the first two chapters in Asmussen [2003].

Definition and the minimal construction of a Markov chain

A stochastic process in continuous time is a family, $(X(t))_{t\geq 0}$, of random variables indexed by the positive real line $[0,\infty)$. The possible values of $(X(t))_{t\geq 0}$, are referred to as the state space, S, of the process. In these lecture notes we shall only consider continuous-time processes on finite or countable state spaces.

Definition 32 (Homogeneous Markov chain in continuous time) [2]*A continuous-time Markov chain on a finite or countable set, S, is a family of random variables, $(X(t))_{t\geq 0}$, on a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$\begin{aligned} & P(X(t_{n+1}) = j | X(t_n) = i, X(t_{n-1}) = i_{n-1}, \dots, X(t_0) = i_0) \\ = \; & P(X(t_{n+1}) = j | X(t_n) = i) \qquad (3.1) \\ = \; & P_{i,j}(t_{n+1} - t_n) \end{aligned}$$

for $j, i, i_{n-1}, \dots, i_0 \in S$ and $t_{n+1} > t_n > \dots > t_0 \geq 0$. The distribution of the Markov chain is determined by

$$\begin{aligned} \phi(i) &= P(X(0) = i) \quad \leftarrow \quad \textit{initial distribution} \\ P_{i,j}(t) &= P(X(t+s) = j | X(s) = i) \quad \leftarrow \quad \textit{transition probabilities} \end{aligned}$$

[2] In Definition 32 the word **homogeneous** refers to the fact that the transition probabilities $P(X(t_{n+1}) = j | X(t_n) = i)$ are assumed to depend only on the time difference $t_{n+1} - t_n$. In more advanced courses you will find that for many applications of the theory it is much more flexible to allow the transition probabilities to depend on both t_{n+1} and t_n. The interpretation of the general case is that the time dynamics of the stochastic process changes (or evolves) with time.

through the identity

$$\begin{aligned} & P(X(t_{n+1}) = j, X(t_n) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_0) = i_0) \\ = \quad & P_{i,j}(t_{n+1} - t_n) \cdot P_{i_{n-1},i}(t_n - t_{n-1}) \cdot \ldots \cdot P_{i_0,i_1}(t_1 - t_0) \cdot \phi(i_0) \end{aligned} \tag{3.2}$$

□

It takes some time to digest the notation used in Definition 32. The condition (3.1) is the natural generalization of the Markov property (2.1) from Chapter 2 to stochastic processes indexed by the continuous set $[0, \infty)$. It is a consequence of more general properties of infinite product measures (-see Appendix A) that the distribution of a Markov chain in continuous time is completely determined by the **initial distribution** and the **transition probabilities**[3] given in Definition 32. Note however, that Definition 32 does not tell us what conditions the transition probabilities $(P_{i,j}(t))_{i,j \in S}$ for $t \geq 0$ must satisfy in order to define a Markov chain. This is one of the main problems with the construction of continuous-time Markov chains. We will briefly return to this discussion in Theorem 38.

[3] Note that for a continuous-time Markov chain then a transition probability, $P_{i,j}(t)$, is something that depends on time. For Markov chains in discrete time discussed in Chapter 2 then the term *transition probability* just refers to transitions one step ahead.

Instead we continue by presenting a dynamical construction of a class of stochastic processes in continuous time that turns out to satisfy the Markov property in Definition 32. This construction (the **minimal construction**) establishes a unique parametrization of a continuous-time Markov chain by **transition intensities** . Throughout these lecture notes you should think of a continuous-time Markov chain as described in terms of an initial distribution and transition intensities with sample paths constructed from the minimal construction described in Definition 33 [4].

[4] When we refer to a continuous-time Markov chain with transition intensities $Q = \{q_{i,j}\}_{i,j \in S}$ we always think of the corresponding jump process given by the minimal construction in Definition 33.

Definition 33 (The minimal construction) *Let $\overline{\phi} = (\phi(i))_{i \in S}$ be a probability vector and let $Q = (q_{i,j})_{i,j \in S}$ be real numbers with the following properties* [5]

$$\begin{aligned} q_{i,j} &\geq 0 \quad i \neq j, i, j \in S \\ q_{i,i} &= -\sum_{j \neq i} q_{i,j}. \end{aligned}$$

[5] We shall use the notation

$$q_i = \sum_{j \neq i} q_{i,j}$$

for the sum of the off-diagonal elements of the transition intensities. Note that due to the constraints in Definition 33 we have $q_i \geq 0$ and $q_i = -q_{i,i}$.

*The time-homogeneous **continuous-time Markov** chain with initial distribution $\overline{\phi}$ and **transition intensity** Q is the stochastic process $(X(t))_{t \geq 0}$ given by the following construction*

- *choose $Y(0)$ according to the initial distribution such that*

$$P(Y(0) = i) = \phi(i)$$

- *given $Y(0)$ let $\tau_1 := W_1$ follow an exponential distribution with rate parameter* [6] $-q_{Y(0),Y(0)}$ *and define $X(t) = Y(0), t \in [0, W_1)$*

[6] The exponential distribution with rate parameter $\lambda > 0$ has mean $1/\lambda$ - see also Appendix B.1.

- *given $Y(0)$ and W_1 choose $Y(1)$ such that*

$$P(Y(1) = i|Y(0)) = \frac{q_{Y(0),i}}{-q_{Y(0),Y(0)}}, i \neq Y(0)$$

Recursively, given $Y(0), \ldots, Y(n), W_1, \ldots, W_n$ [7]

- *choose W_{n+1} according to an exponential distribution with rate parameter $-q_{Y(n),Y(n)}$, let $\tau_{n+1} = \tau_n + W_{n+1}$ and define*

$$X(t) = Y(n), t \in [\tau_n, \tau_{n+1})$$

- *choose $Y(n+1)$ such that*

$$P(Y(n+1) = i|Y(0), \ldots, Y(n), W_1, \ldots, W_{n+1}) = \frac{q_{Y(n),i}}{-q_{Y(n),Y(n)}}, i \neq Y(n).$$

□

[7] Note that the recursive construction of W_{n+1} and $Y(n+1)$ given past values only makes use of the present state $Y(n)$. This is necessary for the resulting process $(X(t))_{t \geq 0}$ to satisfy the Markov property (3.1) from Definition 32. It is less obvious that the additional requirement of exponential waiting times between jumps is sufficient to get a Markov chain.

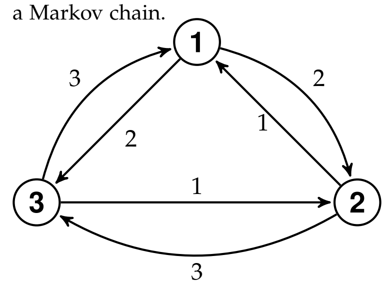


Figure 3.1: Transition diagram for Markov chain of Example 34 and Exercise 3.2.6.

Example 34 *Consider the continuous-time Markov chain on $S = \{1, 2, 3\}$ with transition intensity matrix*

$$Q = \begin{pmatrix} -4 & 2 & 2 \\ 1 & -4 & 3 \\ 3 & 1 & -4 \end{pmatrix}$$

given by the transition diagram on Figure 3.1. Let us try to mimick the minimal construction of the Markov chain assuming that the Markov chain starts in state 1 (- see Figure 3.2).

Assuming that the initial distribution is given by the vector $\overline{\phi} = (1, 0, 0)$ we have that $Y(0) = 1$ and that the waiting time, τ_1, to first jump follows an exponential distribution with rate parameter $-q_{1,1} = 4$ (i.e. mean waiting time $1/4$). At time τ_1 the Markov chain jumps to state 2 or 3 with probabilities

$$\begin{aligned} P(Y(1) = 2|Y(0), W_1) &= \frac{q_{1,2}}{-q_{1,1}} = \frac{2}{4} \\ P(Y(1) = 3|Y(0), W_1) &= \frac{q_{1,3}}{-q_{1,1}} = \frac{2}{4}. \end{aligned}$$

The waiting time, W_2, between the first and the second jump follows an exponential distribution with a parameter given by $-q_{Y(1),Y(1)}$. In this particular example the rate parameter turns out to be 4 regardless of the value of $Y(1)$. The second jump arrives at time $\tau_2 = \tau_1 + W_2$. The target, $Y(2)$, of the second jump is chosen according to the formula

$$P(Y(2) = i|Y(0), Y(1), W_1, W_2) = \frac{q_{Y(1),i}}{-q_{Y(1),Y(1)}}, \quad i \neq Y(1).$$

□

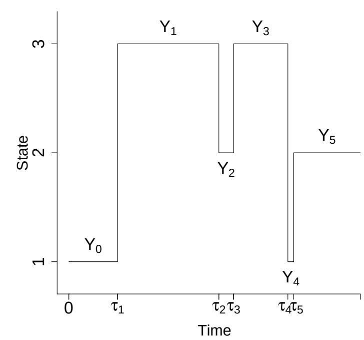


Figure 3.2: Sample path of the continuous-time Markov chain with three states in Example 34. The sequence of states, $\{Y(n)\}_{n \geq 0}$, turns out to be a discrete-time Markov chain with transition probabilities given by Definition 35. The waiting time $\tau_{n+1} - \tau_n$ between jump n and $n+1$ follows an exponential distribution with rate parameter $q_{Y(n)} = -q_{Y(n),Y(n)}$.

There are two potential problems with the minimal construction in Defintion 33 that require further attention.

Absorption If the Markov chain of the minimal construction at time τ_n jumps to a state $Y(n) = i$ with $-q_{i,i} = 0$ then we let

$$X(t) = Y(n), t \geq \tau_n$$

and we say that the Markov chain is **absorbed** at state i. Note that the construction in Definition 33 still makes sense in this case if we interpret an exponential distribution with rate parameter 0 as a random variable with probability mass 1 at $+\infty$.

Explosion There may be infinitely many jumps in finite time such that the random variable, $\tau_\infty := \lim_{n\to\infty} \tau_n$[8], is finite with positive probability

$$P(\tau_\infty < +\infty) > 0.$$

In this case the minimal construction does not assign a value to $X(t)$ for $t \geq \tau_\infty$. To ensure that the minimal construction always define a stochastic process for all $t \geq 0$ we introduce an extra state Δ and let $X(t) = \Delta$ for $t \geq \tau_\infty$. We shall refer to τ_∞ as the time of **explosion**.

[8] For the sequence of jump times we have that $\tau_1 \leq \tau_2 \leq \ldots$ in particular the limit $\tau_\infty := \lim_{n\to\infty} \tau_n$ exists as a random variable taking values in $[0, \infty]$.

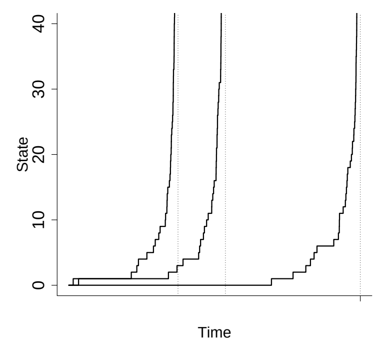


Figure 3.3: Different sample paths for a continuous-time Markov chain with state space $\mathbb{N}_0$ for which explosion is possible (-see also Examples 37 and 61). Times of explosion, τ_∞, are diplayed as dotted vertical lines.

Using the conventions above to handle the case of absorption or explosion then the minimal construction of Definition 33 always defines a continuous-time process $(X(t))_{t\geq 0}$ on the extended state space $\overline{S} = S \cup \{\Delta\}$. However, the minimal construction also defines a discrete time process $Y(0), Y(1), Y(2), \ldots$ that keeps track of the sequence of states visited by $(X(t))_{t\geq 0}$. Strictly speaking, if $(X(t))_{t\geq 0}$ is absorbed in some state i at the time, τ_m, of the m-th jump then

$$Y(m+1), Y(m+2), \ldots$$

are not defined. We shall apply the convention that in this case we let

$$Y(m+1) = Y(m+2) = \ldots = i$$

since then we have the following result

Theorem 35 (Embedded Markov chain of jumps) *For a continuous-time Markov chain with transition intensity $Q = (q_{i,j})_{i,j\in S}$ given by the minimal construction in Definition 33 then the sequence $(Y(n))_{n\in\mathbb{N}_0}$ of visited states is a discrete-time Markov chain with transition probabilities* [9]

$$P_{i,j} = \begin{cases} -\frac{q_{i,j}}{q_{i,i}} = \frac{q_{i,j}}{q_i} & i \in S \backslash A, j \notin \{i, \Delta\} \\ 0 & i \in S \backslash A, j \in \{i, \Delta\} \\ 0 & i \in A, j \neq i \\ 1 & i \in A, j = i \end{cases}$$

where $A = \{i \in S | q_{i,i} = 0\}$ is the subset of absorbing states. □

[9] The dynamics of the embedded Markov chain should strictly speaking be given by a transition matrix on the extended state space $S \cup \{\Delta\}$. To obtain this let $P_{\Delta,\Delta} = 1$ and $P_{\Delta,j} = 0$ for $j \neq \Delta$.

Example 36 *For the Markov chain in Example 34 corresponding to Figure 3.1 the embedded Markov chain of jumps has state space* $S = \{1, 2, 3\}$ *and transition probability matrix given by*

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \\ 3/4 & 1/4 & 0 \end{pmatrix}.$$

□

Example 37 (Explosion) *For the continuous time Markov chain on* $\mathbb{N}_0$ *with initial distribution* $P(X(0) = 0) = 1$ *and transition intensities*

$$q_{i,i+1} = (i+1)^2, i \in \mathbb{N}_0 \quad \text{and} \quad q_{i,j} = 0, j \neq i, i+1,$$

one may show that explosion occurs with probability one. Three examples of samples paths simulated from this Markov chain is shown in Figure 3.3. This is an example of a ***pure birth process*** *(-see also Example 61).*

Since we have $q_{i,i} \neq 0$ *for all* $i \in \mathbb{N}_0$ *then there are no absorbing states. We get from Theorem 35 that the embedded Markov chain of jumps has transition probabilities*

$$P_{i,i+1} = -\frac{q_{i,i+1}}{q_{i,i}} = -\frac{(i+1)^2}{-(i+1)^2} = 1$$

and consequently $P_{i,j} = 0$ *for* $j \neq i + 1$.

Because explosion is possible the transition intensities do not define the value of the Markov chain after the explosion time τ_∞. *The recommendation above is to let* $X(t) = \Delta$ *for* $t \geq \tau_\infty$ *where* Δ *is some arbitrary state we just introduce for completeness of the construction. With the introduction of* Δ *then the embedded Markov chain of jumps should also be defined on* $\mathbb{N}_0 \cup \{\Delta\}$. *From a practical point of view this is unimportant, but we may handle this purely technical issue by letting* $P_{\Delta,\Delta} = 1$, *and* $P_{\Delta,j} = 0$ *for* $j \neq \Delta$. □

Properties of the transition probabilities

The main ingredient in Definition 32 is the family of transition probabilities $(P_{i,j}(t))_{i,j \in S}$ for $t \geq 0$. Clearly, the transition probabilities must be non-negative

$$P_{i,j}(t) = P(X(t+s) = j | X(s) = i) \geq 0$$

and if explosion is not possible [10] (!) then

$$\begin{aligned} \sum_{j \in S} P_{i,j}(t) &= \sum_{j \in S} P(X(t+s) = j | X(s) = i) \\ &= P(X(t+s) \in S | X(s) = i) = 1. \end{aligned}$$

[10] If we consider a Markov chain where explosion is possible then

$$\sum_{j \in S} P_{i,j}(t) = 1 - P_{i,\Delta}(t) < 1.$$

Fix this by replacing S with the extended state space $\overline{S} = S \cup \{\Delta\}$.

In particular, for any fixed $t \geq 0$ then $(P_{i,j}(t))_{i,j \in S}$ must be a transition probability in the sense introduced for discrete-time Markov chains [11]. However, the transition probabilities for different time arguments must fit together in accordance with the **Chapman-Kolmogorov equations** given below.

[11] For fixed $t > 0$ then the process

$$X(t), X(2t), X(3t), \ldots$$

is a discrete-time Markov chain with transition probabilities

$$P_{i,j} = P_{i,j}(t).$$

Theorem 38 (Chapman-Kolmogorov equations) *The transition probabilities for a homogeneous continuous-time Markov chain satisfy the Chapman-Kolmogorov equations*

$$\forall s, t \geq 0, \forall i, j \in S : P_{i,j}(t+s) = \sum_{l \in S} P_{i,l}(t) \cdot P_{l,j}(s).$$

If the state space is finite ($|S| < \infty$) then $P(t) = (P_{i,j}(t))_{i,j \in S}$ may be regarded as a matrix for any fixed $t \geq 0$ and the Chapman-Kolmogorov equations may be written as a matrix equation

$$P(t+s) = P(t) \cdot P(s).$$

Proof: For $i, j \in S$, $0 \leq s, t$ and any $0 \leq u$ we have

$$\begin{aligned}
& P_{i,j}(t+s) \\
= \;& P(X(t+s+u) = j | X(u) = i) \\
= \;& \sum_{l \in S} P(X(t+s+u) = j, X(t+u) = l | X(u) = i) \\
= \;& \sum_{l \in S} P(X(t+s+u) = j | X(t+u) = l, X(u) = i) \cdot P(X(t+u) = l | X(u) = i) \\
= \;& \sum_{l \in S} P(X(t+s+u) = j | X(t+u) = l) \cdot P(X(t+u) = l | X(u) = i) \\
= \;& \sum_{l \in S} P_{l,j}((t+s+u) - (t+u)) P_{i,l}((t+u) - u) \\
= \;& \sum_{l \in S} P_{i,l}(t) \cdot P_{l,j}(s)
\end{aligned}$$

□

On this course Markov chains are usually defined in terms of the **transition intensities** , $Q = (q_{ij})_{i,j \in S}$, used in the minimal construction of a Markov chain (-see Definition 33). However, for many applications we are more interested in computing the transition probabilities

$$P(X(t+s) = j | X(s) = i) := P_{i,j}(t), \quad i, j \in S, t \geq 0$$

In the following pages we discuss various results relating the transition intensity, Q, of a continuous-time Markov chain and the transition probabilities $P(t) = (P_{i,j}(t))_{i,j \in S}$.

Theorem 39 (Infinitesimal generator of a Markov chain) *For a continuous-time Markov chain, $(X(t))_{t \geq 0}$, the transition intensities may be* *obtained from transition probabilities $P(t) = (P_{i,j}(t))_{i,j \in S}$ as the limits*

$$\lim_{t \to 0+} \frac{P_{i,i}(t) - 1}{t} = q_{i,i} \tag{3.3}$$

$$\lim_{t \to 0+} \frac{P_{i,j}(t)}{t} = q_{i,j}, \quad i \neq j. \tag{3.4}$$

Proof: The case $q_{i,i} = 0$ (i absorbing) follows immediately by observing that in this case $P_{i,i}(t) = 1$ for $t \geq 0$.

For $q_{i,i} \neq 0$ we first establish a result concerning the probability of observing at least two jumps on $[0, t]$ for $t \to 0+$. Denote by τ_2 the time of the second jump from the minimal construction in Definition 33. Splitting the event $(\tau_2 \leq t)$ according to the time and target state of the first jump we get

$$\begin{aligned}
& P(\tau_2 \leq t | X(0) = i) \\
= & \sum_{j \neq i} P(\tau_2 \leq t, Y(1) = j | X(0) = i) \\
= & \sum_{j \neq i} \int_0^t (-q_{i,i}) \exp(q_{i,i} s) \frac{q_{i,j}}{-q_{i,i}} (1 - \exp(q_{j,j}(t - s))) ds \\
\leq & \sum_{j \neq i} q_{i,j} (1 - \exp(q_{j,j} t)) \int_0^t \exp(q_{i,i} s) ds \\
\leq & (1 - \exp(q_{i,i} t)) \sum_{j \neq i} \frac{q_{i,j}}{-q_{i,i}} (1 - \exp(q_{j,j} t)).
\end{aligned}$$

For $t > 0$ the function $f_t(j) = \frac{q_{i,j}}{-q_{i,i}}(1 - \exp(q_{j,j} t))$ is dominated by $f(j) = \frac{q_{i,j}}{-q_{i,i}}$ which is integrable (-with sum one!) with respect to the counting measure on $S \backslash \{i\}$. In particular, dominated convergence gives that

$$\sum_{j \neq i} \frac{q_{i,j}}{-q_{i,i}} (1 - \exp(q_{j,j} t)) \to 0$$

for $t \to 0+$ and the computations above show that

$$\frac{P(\tau_2 \leq t | X(0) = i)}{t} \to 0 \tag{3.5}$$

for $t \to 0+$.

Note that for $X(0) = i$ then the event $(X(t) = i)$ requires that we make either zero jumps or at least two jumps on the interval $[0, t]$. Consequently, we get

$$\begin{aligned}
\frac{P_{i,i}(t)}{t} & = \frac{P(X(t) = i | X(0) = i)}{t} \\
& = \frac{P(X(t) = i, \tau_1 > t | X(0) = i)}{t} + \frac{P(X(t) = i, \tau_2 \leq t | X(0) = i)}{t}.
\end{aligned}$$

Using (3.5) and that $P(X(t) = i, \tau_1 > t | X(0) = i) = \exp(q_{i,i}t)$ we conclude that

$$\frac{P_{i,i}(t) - 1}{t} \to q_{i,i}$$

for $t \to 0+$.

For $j \neq i$ we have

$$\begin{aligned}
\frac{P_{i,j}(t)}{t} &= \frac{P(X(t) = j | X(0) = i)}{t} \\
&= \frac{P(X(t) = j, \tau_2 > t | X(0) = i)}{t} + \frac{P(X(t) = j, \tau_2 \leq t | X(0) = i)}{t} \\
&= \frac{\int_0^t \exp(q_{i,i}s) q_{i,j} \exp(q_{j,j}(t-s)) ds}{t} \\
&+ \frac{P(X(t) = j, \tau_2 \leq t | X(0) = i)}{t}
\end{aligned}$$

By continuity of the integrand the first term may be bounded by $(1-\epsilon)q_{i,j}$ and $q_{i,j}$ for any $\epsilon > 0$ provided that we choose t sufficiently small. From (3.5) the last term tends to zero as $t \to 0$ completing the proof of (3.4). □

More generally, the transition probabilities and the transition intensities are related through the **backward differential equations** (also referred to as **Kolmogorov's differential equation**).

Theorem 40 (Backward differential equations) *For a continuous-time Markov chain, $(X(t))_{t \geq 0}$, with transition intensity, $Q = (q_{i,j})_{i,j \in S}$, and transition probabilities $(P_{i,j}(t))_{i,j \in S}$ it always holds that*

$$DP_{i,j}(t) = P'_{i,j}(t) = q_{i,i}P_{i,j}(t) + \sum_{k \neq i} q_{i,k} P_{k,j}(t) \tag{3.6}$$

Proof: An intermediate step in deriving the backward differential equations of Theorem 40 is the set of **backward integral equations** (3.7) below which may be of interest in itself. The idea is to compute $P_{i,j}(t) = P(X(t+s) = j | X(s) = i)$ by conditioning on the time of the first jump in $[s, t+s]$ and the target state $k \in S$ of this first jump. Since $X(s) = i$ then by the minimal construction in Definition 33 the waiting time to the first jump follows an exponential distribution with rate parameter $q_i = -q_{i,i}$. We therefore get

$$\begin{aligned}
P_{i,j}(t) &= \delta_{i,j} \exp(q_{i,i}t) + \int_0^t \sum_{k \neq i} (-q_{i,i}) \frac{q_{i,k}}{-q_{i,i}} \exp(q_{i,i}(v)) P_{k,j}(t-v) dv \\
&= \delta_{i,j} \exp(q_{i,i}t) + \int_0^t \sum_{k \neq i} q_{i,k} \exp(q_{i,i}(t-u)) P_{k,j}(u) du, \qquad (3.7)
\end{aligned}$$

where $\delta_{i,j} = 0, i \neq j$, and $\delta_{i,i} = 1$. The contribution of the first term in (3.7) reflects the situation, where there is no jump on $[s, s+t]$.

Rewrite (3.7) as

$$P_{i,j}(t) = \exp(q_{i,i}t)\left(\delta_{i,j} + \int_0^t \sum_{k\neq i} q_{i,k}\exp(q_{i,i}(-u))P_{k,j}(u)du\right) \quad (3.8)$$

to see that $t \to P_{i,j}(t)$ is continuous. In order to deduce that

$$\int_0^t \sum_{k\neq i} q_{i,k}\exp(q_{i,i}(-u))P_{k,j}(u)du$$

is differentiable we need

$$u \to \sum_{k\neq i} q_{i,k}\exp(q_{i,i}(-u))P_{k,j}(u)$$

to be a continuous function. This may be verified formally by using dominated convergence. Basic rules for differentiation now yields

$$\begin{aligned} P'_{i,j}(t) &= q_{i,i}\exp(q_{i,i}t)\left(\delta_{i,j} + \int_0^t \sum_{k\neq i} q_{i,k}\exp(q_{i,i}(-u))P_{k,j}(u)du\right) \\ &+ \exp(q_{i,i}t)\sum_{k\neq i} q_{i,k}\exp(q_{i,i}(-t))P_{k,j}(t) \\ &= q_{i,i}P_{i,j}(t) + \sum_{k\neq i} q_{i,k}P_{k,j}(t). \end{aligned}$$

□

There also exist sets of **forward differential** and **integral equations** for continuous-time Markov chains.

Theorem 41 (Forward differential and integral equations) *For a continuous-time Markov chain, $(X(t))_{t\geq 0}$, with transition intensities, $Q = (q_{i,j})_{i,j\in S}$, and transition probabilities $P(t) = (P_{i,j}(t))_{i,j\in S}$ it holds that*

$$P_{i,j}(t) = \delta_{i,j}\exp(q_{j,j}t) + \int_0^t \sum_{l\neq j} P_{i,l}(u)q_{l,j}\exp(q_{j,j}(t-u))du$$

and

$$DP_{i,j}(t) = P_{i,j}(t)q_{j,j} + \sum_{l\neq j} P_{i,l}(t)q_{l,j}.$$

□

Comments to the proof: The full proof of the forward differential equations is not trivial. A simpler version of the proof may be given under the assumption that

$$\sum_{j\in S} P_{i,j}(t)(-q_{j,j}) < \infty,$$

but the proof is omitted here. □

Remark 42 *We have previously discussed continuous-time Markov chains where explosion may occur. Explosion refers to the fact that there may be infinitely many jumps in finite time. One can show that explosion does not happen if the condition above is satisfied, i.e. if*

$$\sum_{j \in S} P_{i,j}(t)(-q_{j,j}) < \infty.$$

It turns out that if explosion is not possible then the differential equations of Theorems 40 and 41 uniquely determines the transition probabilities $P(t) = (P_{i,j}(t))_{i,j \in S}$ subject to the initial conditions $P(0) = I$. If explosion is possible then there is no unique solution to the differential equations. The minimal solution will give the transition probabilities corresponding to the process described by the minimal construction of a Markov jump process. □

Theorem 43 (Transition probabilities for finite state space) *For a continuous-time Markov chain on a finite state space the backward differential equation may be expressed in matrix form as*

$$DP(t) = P'(t) = QP(t)$$

where $P(t) = (P_{i,j}(t))_{i,j \in S}$. Using the boundary condition $P(0) = I$ it turns out that the transition probabilities may expressed in terms of ***exponential matrices*** [12] *as*

$$P(t) = \exp(Qt), \quad t \geq 0.$$

□

[12] You may find the definition of an exponential matrix on page 135 in Appendix B. Formally, $\exp(A)$ is defined as the result of replacing a real (or complex!) number x by a matrix A in the series representation of the exponential function $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$.

Using the forward and backward differential equations it may be possible to find closed form expressions for some of the transition probabilities, $P_{i,j}(t)$, for certain values of i, j. Though Theorem 43 gives a general formula for $P(t) = \{P_{i,j}(t)\}_{i,j \in S}$ it only leads to explicit formulaes in very nice examples. There are plenty of opportunities to get familiar with applying Theorems 40, 41 and 43 if you do some of the Exercises 3.1.3, 3.2.1, 3.2.2, 3.2.3, 3.2.4, 3.2.5, 3.2.6, 3.3.3 and 3.3.4.

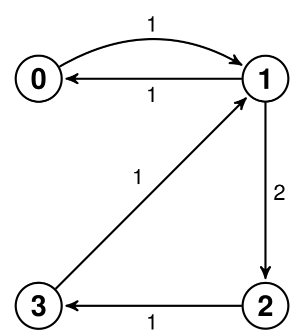


Figure 3.4: Transition diagram for a Markov chain with four states in Example 44.

Example 44 (Transition probabilities for a four state Markov chain) *We consider a continuous-time Markov chain with four states $S = \{0, 1, 2, 3\}$ and transition intensities given by*

$$Q = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -3 & 2 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

According to Theorem 40 the backward differential equations

$$QP(t) = P'(t)$$

for $P_{i,0}(t), i = 0, 1, 2, 3,$ *may be written out as*

$$\begin{aligned}(-1)\cdot P_{0,0}(t) + 1\cdot P_{1,0}(t) &= P'_{0,0}(t)\\ 1\cdot P_{0,0}(t) + (-3)\cdot P_{1,0}(t) + 2\cdot P_{2,0}(t) &= P'_{1,0}(t)\\ (-1)\cdot P_{2,0}(t) + 1\cdot P_{3,0}(t) &= P'_{2,0}(t)\\ 1\cdot P_{1,0}(t) + (-1)\cdot P_{3,0}(t) &= P'_{3,0}(t).\end{aligned}$$

It is easy to give similar expressions for $P'_{i,j}(t), j \neq 0$ *but it seems difficult to find (or guess) the formula for a general solution to the system of first order differential equations.*

The solution, however, is given in Theorem 43 in the form of an exponential matrix as $P(t) = \exp(Qt)$*. For this particular example it should actually be possible to obtain closed analytic formulaes for the exponential matrix* $\exp(Q)$*. The characteristic polynomium for* Q *becomes*

$$\begin{aligned} g(\lambda) &= \det(Q-\lambda I) = \det\begin{pmatrix} -1-\lambda & 1 & 0 & 0\\ 1 & -3-\lambda & 2 & 0\\ 0 & 0 & -1-\lambda & 1\\ 0 & 1 & 0 & -1-\lambda\end{pmatrix}\\ &= (-1)^{1+1}(-1-\lambda)\det\begin{pmatrix} -3-\lambda & 2 & 0\\ 0 & -1-\lambda & 1\\ 1 & 0 & -1-\lambda\end{pmatrix}\\ &+ (-1)^{1+2}1\det\begin{pmatrix} 1 & 2 & 0\\ 0 & -1-\lambda & 1\\ 0 & 0 & -1-\lambda\end{pmatrix}\\ &= (-1-\lambda)\cdot\left\{-(3+\lambda)(1+\lambda)^2+2\right\} - 1(1+\lambda)^2\\ &= (-1-\lambda)\cdot\left\{-3-6\lambda-3\lambda^2-\lambda-2\lambda^2-\lambda^3+2+1+\lambda\right\}\\ &= -(1+\lambda)\lambda\cdot\left\{-6-5\lambda-\lambda^2\right\}\\ &= \lambda(1+\lambda)(2+\lambda)(3+\lambda).\end{aligned}$$

Since the 4×4 *transition matrix* Q *has four distinct (real) eigenvalues (given as solutions to* $g(\lambda) = 0$*), there exists an invertible matrix* O *such that* $O^{-1}QO$ *is the diagonal matrix with entries* $0, -1, -2, -3$ [13]*. It follows that all transition probabilities* $P_{i,j}(t)$ *will be linear combinations of* $1, \exp(-t), \exp(-2t)$ *and* $\exp(-3t)$*.* □

[13] The matrix

$$O = \begin{pmatrix} 1/2 & -2/\sqrt{5} & -1/2 & -2/5\\ 1/2 & 0 & 1/2 & 4/5\\ 1/2 & 1/\sqrt{5} & 1/2 & 1/5\\ 1/2 & 0 & -1/2 & -2/5\end{pmatrix}$$

composed of normalised eigenvectors for Q as columns does the job. For more details see page 135 in Appendix B.

As demonstrated in Example 60 below the forward differential equations from Theorem 41 may be useful even when the state space S is not finite.

Invariant probabilities and absorption

The present section is devoted to the study of long run properties of a continuous-time Markov chain $(X(t))_{t\geq 0}$. Using the identity (split according to initial state)

$$P(X(t) = j) = \sum_{i \in S} \phi(i) P_{i,j}(t)$$

the problem reduces to studying the behaviour of $P_{i,j}(t)$ as $t \to \infty$. The strategy will be to borrow as much as possible from Chapter 2. The key steps are the following

- Find conditions to ensure that for arbitrary $h > 0$ then $(X(nh))_{n\in\mathbb{N}_0}$ is an irreducible, aperiodic and positive recurrent Markov chain in discrete time with transition probabilities $P(h) = (P_{i,j}(h))_{i,j\in S}$.
- Use Theorem 23 in Chapter 2 to deduce that

  $$P_{i,j}(nh) \to \pi(j) \quad \text{for} \quad n \to \infty$$

  where $\overline{\pi}P(h) = \overline{\pi}$ is the unique invariant probability for $P(h)$.
- Argue that since $h > 0$ is arbitrary then also

  $$\lim_{t\to\infty} P_{i,j}(t) = \pi(j).$$

The formal proof of the last step is straight-forward and may be found in Theorem 49 and Lemma 50 below. More work is required to get behind the first step listed above. A central point is to show that the conclusion holds provided we can find *one* $h_0 > 0$ such that $(X(nh_0))_{n\in\mathbb{N}_0}$ is irreducible and has an invariant probability $\overline{\pi}$. This is established in Theorem 48. We then continue the exposition by discussing various results related to exsistence of an invariant distribution for a continuous-time Markov chain. We finally discuss how to determine the limiting behaviour of continuous-time Markov chains with multiple communication classes.

Definition 45 (Communication classes and irreducibility) *Two states $i, j \in S$ are said to communicate if there exists $s, t > 0$ such that*

$$P_{i,j}(s) > 0 \quad \textit{and} \quad P_{j,i}(t) > 0.$$

This definition partitions the state space, S, into (disjoint) ***communication classes*** [14]*. A continuous-time Markov chain is* ***irreducible*** *if there is only one communication class.* □

[14] What do we need to show in order to formally verify this?

You are allowed to use[15] that for a continuous-time Markov chain two states $i \neq j$ communicate, if and only if there exists a sequence of states $i_1, i_2, \ldots, i_n \in S$ containing state j such that

$$q_{i,i_1} \cdot q_{i_1,i_2} \cdot \ldots \cdot q_{i_{n-1},i_n} \cdot q_{i_n,i} > 0.$$

[15] Part of the argument goes like this: Given $X(0) = i$ the event $(X(t) = j)$ may be split into a countable number of disjoint sets according to the number, n, of jumps on $[0, t]$ and the sequence of visited states along the path of length n connecting i and j. Given that $P_{i,j}(t) > 0$ at least one of these sets may have a strictly positive probability implying the existence of $i_1, \ldots, i_n$ with $i_n = j$ and $q_{i,i_1}, q_{i_1,i_2}, \ldots, q_{i_{n-1},i_n} > 0$.

This observation implies that the communication classes of a continuous-time Markov chain are the same as the communication classes of the embedded Markov chain of jumps with transition probabilities given in Theorem 35.

Since $P_{i,i}(t) > 0$ (why?) we always have that i communicate with itself. Further, it may be demonstrated that if $P_{i,j}(t_0) > 0$ for some $t_0 > 0$ then $P_{i,j}(t) > 0$ for *any* $t > 0$ [16]. In particular, we have that $P_{i,j}(h) > 0$ for all $h > 0$ if i and j belong to the same communication class $\mathcal{C}$. It is tempting to conclude that the restriction of $(X(nh))_{n\in\mathbb{N}_0}$ to a closed communication class [17] is an aperiodic [18] discrete-time Markov chain with transition probabilities $P(h) = (P_{i,j}(h))_{i,j\in C}$ and that the communication classes coincide with those of $(X(t))_{t\geq 0}$. Note however, that even for a closed class $\mathcal{C}$ then we will have that $\sum_{j\in S} P_{i,j}(h) < 1$ for all $h > 0$ and $i \in \mathcal{C}$ if explosion is possible.

[16] This requires an argument. The Chapman-Kolmogorov equations from Theorem 38 is an essential part here.

[17] Closed is defined in terms of the embedded Markov chain of jumps.

[18] It does not make sense to speak of the period of a Markov chain in continuous-time!

Definition 46 (Recurrence and transience) *We give two different (but equivalent) definitions of recurrence for continuous-time Markov chains.*

1. *An irreducible continuous-time Markov chain is recurrent if and only if the embedded discrete-time process of jumps (-see Theorem 35) is recurrent. It is transient if and only if the embedded discrete-time Markov chain of jumps is transient.*

2. *State i is transient if and only if the total time spent in state i*

$$V_i = \int_0^\infty 1(X(t) = i)dt$$

is bounded with probability 1 *(i.e. if* $P(V_i < +\infty | X(0) = i) = 1$*). For a recurrent state i then* $V_i = +\infty$ *with probability one.*

If the continuous-time Markov chain is not irreducible the definitions of recurrence and transience apply seperately to each communication class. Note that an absorbing state will always be recurrent. □

As a consequence of Definition 46 to determine if a continuous-time Markov chain is recurrent or transient we only need to study the embedded discrete-time Markov chain of jumps and use the criterions for recurrence given in Definition 8, Theorem 10, and Theorem 15 of Chapter 2 on discrete-time Markov chains.

Let us briefly discuss the equivalence between the two criterions stated in Definition 46. From the minimal construction (Definition 33) it follows that the total time, V_i, spend in state i may be expressed as

$$V_i = \sum_{n=1}^{N_i} W_n$$

where N_i is the total number of visits to state i for the embedded Markov chain of jumps and $W_1, W_2, \ldots$ are the time spent in state i at each visit. We know from Theorem 12 that

$$P(N_i = +\infty | Y(0) = i) = 1$$

if the embedded Markov chain is recurrent also implying that

$$P(V_i = +\infty | X(0) = i) = 1$$

[19]. If the embedded Markov chain of jumps is transient we know that N_i follows a geometric distribution and we deduce [20] that

$$P(V_i < +\infty | X(0) = i) = 1.$$

As indicated in the beginning of this section then the limiting behaviour of continuous-time Markov chains is intimately related to invariant probabilities.

[19] The formal argument relies on the fact that $W_1, W_2, \ldots$ are independent and exponentially distributed even when we condition on the total number, N_i, of visits to state i.

[20] In the transient case let q denote the probability that the embedded Markov chain of jumps never returns to state i. Then we actually have a closed formula for

$$P(V_i \leq v | X(0) = i)$$

given by

$$q + (1-q)(1 - \exp(qvq_{i,i}))$$

which is a truncated exponential distribution with probability mass q at zero.

Definition 47 (Invariant distribution) *A probability vector, $\overline{\pi} = (\pi(i))_{i \in S}$, is an **invariant** (or **stationary**) **distribution** for a continuous-time Markov chain if for all $t \geq 0$ and $j \in S$*

$$\pi(j) \quad = \quad \sum_{i \in S} \pi(i) P_{i,j}(t). \tag{3.9}$$

□

It is a simple consequence of the Markov property that if $\overline{\pi}$ is an invariant distribution for a continuous-time Markov chain $(X(t))_{t \geq 0}$ then

$$P(X(t) = j) = \pi(j),$$

for all $t \geq 0$ if we let $P(X(0) = j) = \pi(j)$. We say that $(X(t))_{t \geq 0}$ is a stationary process provided that it is started according to the invariant distribution.

Theorem 48 (Uniqueness of invariant distribution) *For an irreducible continuous-time Markov chain then the invariant distribution is unique if it exists.*

If for some $t_0 > 0$ there is a probability $\overline{\pi} = (\pi(i))_{i \in S}$ such that

$$\forall j \in S \quad : \quad \pi(j) = \sum_{i \in S} \pi(i) P_{i,j}(t_0)$$

then we may conclude that

1. $\forall i \in S \quad : \quad \pi(i) > 0$
2. $P(t_0)$ *is a transition probability, i.e.*

$$\forall i \in S \quad : \quad \sum_{j \in S} P_{i,j}(t_0) = 1$$

3. *$\overline{\pi}$ is an invariant distribution for the Markov chain, i.e.*

$$\forall t \geq 0, \forall j \in S \quad : \quad \pi(j) = \sum_{i \in S} \pi(i) P_{i,j}(t).$$

Proof: The proof of Theorem 48 may be constructed along the lines of Exercise C.4. □

The previous result (Theorem 48) is important. Firstly, it establishes the uniqueness of invariant distributions, and secondly it shows that we only need to find a solution to (3.9) for *one* $t_0 > 0$. Finally, it shows that $\sum_{j \in S} P_{i,j}(h) = 1$ for all $h \geq 0$ if an invariant distribution exists. In particular, explosion is not possible and we have that $(X(nh))_{n \in \mathbb{N}_0}$ is a discrete-time Markov chain for all $h > 0$.

By combining the results above we have shown the following. Suppose that $(X(t))_{t \geq 0}$ is an irreducible Markov chain and that we can find a solution to (3.9) for some $t_0 > 0$. Then $(X(nh))_{n \in \mathbb{N}_0}$ is an irreducible, aperiodic and positive recurrent (=existence of invariant distribution) Markov chain in discrete time. The following result (known as the ergodic theorem for Markov chains) determines the limiting behaviour of an irreducible Markov chain given the existence of an invariant distibution.

Theorem 49 (Limit results for transition probabilities) *For an irreducible Markov chain, $(X(t))_{t \geq 0}$, with invariant distribution $(\pi(i))_{i \in S}$ it holds for all $i, j \in S$ that*

$$\lim_{t \to \infty} P_{i,j}(t) = \pi(j).$$

Further, for any initial distribution $\overline{\phi} = (\phi(i))_{i \in S}$ and $j \in S$ it holds that

$$\lim_{t \to \infty} P(X(t) = j) = \pi(j).$$

If no invariant distribution exists then

$$\lim_{t \to \infty} P_{i,j}(t) = 0.$$

□

For the proof of Theorem 49 we need a result concerning uniform continuity of the transition probabilities $P_{i,j}(t)$ over $i \in S$.

Lemma 50 *For all $i, j \in S$ and $0 \leq t, h$ it holds that*

$$|P_{i,j}(t+h) - P_{i,j}(t)| \leq 2(1 - P_{i,i}(h)).$$

Proof: Use the Chapman-Kolmogorov equations from Theorem 38 and the triangle inequality to get

$$\begin{aligned}
|P_{i,j}(t+h) - P_{i,j}(t)| &\leq |\sum_{l\in S} P_{i,l}(h)P_{l,j}(t) - P_{i,j}(t)| \\
&= |\sum_{l\neq i} P_{i,l}(h)P_{l,j}(t) + (P_{i,i}(h)-1)P_{i,j}(t)| \\
&\leq \sum_{l\neq i} P_{i,l}(h)P_{l,j}(t) + |P_{i,i}(h)-1|P_{i,j}(t) \\
&\leq \sum_{l\neq i} P_{i,l}(h)\cdot 1 + (1-P_{i,i}(h)) = 2(1-P_{i,i}(h)).
\end{aligned}$$

□

Proof of Theorem 49: By Definition 47 then for any $h > 0$ we have that $\overline{\pi}$ is invariant for $P(h) := (P_{i,j}(h))_{i,j\in S}$. Since $P(h)$ is the transition probability for the irreducible, aperiodic [21] discrete-time Markov chain, $(X(nh))_{n\in\mathbb{N}_0}$, obtained by sampling $(X(t))_{t\geq 0}$ at grid points $h, 2h, 3h, \ldots$ then we get from Theorem 20 in Chapter 2 that

[21] It follows from 2. in Theorem 48 that $P(h)$ is a transition matrix, and aperiodicity follows from the fact that for irreducible continuous-time Markov chains then $P_{i,j}(t) > 0$ for any $t > 0$ and $i, j \in S$.

$$\lim_{n\to\infty} P_{i,j}(nh) = \pi(j). \tag{3.10}$$

For $\epsilon > 0$ use that $P_{i,i}(t) \to 1$ for $t \to 0+$ to choose $h > 0$ such that

$$\sup_{t\leq h} 2(1-P_{i,i}(t)) < \epsilon.$$

For any $0 < t < \infty$ choose an integer $n(t) \in \mathbb{N}$ such that $t \in [n(t)h, (n(t)+1)h)$ and note that by Lemma 50

$$\begin{aligned}
P_{i,j}(t) &= P_{i,j}(n(t)h) + P_{i,j}(t) - P_{i,j}(n(t)h) \\
&\leq P_{i,j}(n(t)h) + sup_{t\leq h} 2(1-P_{i,i}(t)) < P_{i,j}(n(t)h) + \epsilon.
\end{aligned}$$

Since $t > 0$ was arbitrary and as $n(t) \to \infty$ for $t \to \infty$ we conclude by referring to (3.10) that

$$\limsup_{t\to\infty} P_{i,j}(t) \leq \pi(j) + \epsilon.$$

Letting $\epsilon \to 0$ we obtain $\limsup_{t\to\infty} P_{i,j}(t) \leq \pi(j)$ and a similar approach may be used to show that $\liminf_{t\to\infty} P_{i,j}(t) \geq \pi(j)$.

For an arbitrary initial distribution it follows by dominated convergence that

$$\begin{aligned}
\lim_{t\to\infty} P(X(t)=j) &= \lim_{t\to\infty} \sum_{i\in S} P(X(t)=j, X(0)=i) \\
&= \lim_{t\to\infty} \sum_{i\in S} P(X(0)=i)P(X(t)=j|X(0)=i) \\
&= \sum_{i\in S} P(X(0)=i) \lim_{t\to\infty} P_{i,j}(t) \\
&= \pi(j)\cdot \sum_{i\in S} P(X(0)=i) = \pi(j).
\end{aligned}$$

We conclude the proof by considering the case where no invariant distribution exists. If explosion is not possible then the Markov chain $(X(nh))_{n\in\mathbb{N}_0}$ is either null-recurrent or transient implying that $\lim_{n\to\infty} P_{i,j}(nh) = 0$ according to Theorem 24 and 25 in Chapter 2. Similar arguments as in the positive recurrent case above may be used to show that then also $\lim_{t\to\infty} P_{i,j}(t) = 0$.

When explosion is possible [22] and we decide to let $X(t) = \Delta$ after explosion, then $(X(nh))_{n\in\mathbb{N}_0}$ may be regarded as a discrete-time Markov chain on an extended state space with Δ as an extra absorbing state. Here $P_{i,\Delta}(h) > 0$ for all $i \in S$ implying that $\lim_{t\to\infty} P_{i,j}(t) = 0$. □

[22] You may skip this part!

Comments to the proof: Digging into the proof of Theorem 49 it is possible to show that $\lim_{t\to\infty} P_{ij}(t) = 0$ even if the Markov chain is not irreducible as long as j does not belong to a positive recurrent communication class. □

Since Markov chains are usually specified in terms of the transition intensities we can rarely apply Definition 47 directly to find the invariant distribution. The following result gives a necessary condition for a stationary distribution expressed in terms of the transition intensities.

Theorem 51 (Necessary condition for a stationary distribution) *For a continuous-time Markov chain with transition intensities, $Q = (q_{i,j})_{i,j\in S}$, an invariant probability, $\overline{\pi} = (\pi(i))_{i\in S}$, must satisfy the system of equations*

$$\forall j \in S \quad : \quad \sum_{i\in S} \pi(i) q_{i,j} = 0 \tag{3.11}$$

or equivalently

$$\forall j \in S \quad : \quad \sum_{i\neq j} \pi(i) q_{i,j} = \pi(j)(-q_{j,j}) = \pi(j) q_j.$$

Thinking of $\overline{\pi}$ as a row vector and of Q as a matrix the system of equations has a more compact formulation as

$$\overline{\pi} Q = 0.$$

Proof: Assuming that $\overline{\pi}$ is an invariant distribution for a continuous-time Markov chain we get from the forward integral equations (-see

Theorem 41) that for $j \in S$ with $q_{j,j} \neq 0$ then

$$
\begin{aligned}
\pi(j) &= \sum_{i \in S} \pi(i) P_{i,j}(t) \\
&= \sum_{i \in S} \pi(i) \left(\delta_{i,j} \exp(q_{j,j} t) + \int_0^t \sum_{l \neq j} P_{i,l}(u) q_{l,j} \exp(q_{j,j}(t-u)) du \right) \\
&= \pi(j) \exp(q_{j,j} t) + \int_0^t \sum_{l \neq j} \sum_{i \in S} \pi(i) P_{i,l}(u) q_{l,j} \exp(q_{j,j}(t-u)) du \\
&= \pi(j) \exp(q_{j,j} t) + \int_0^t \sum_{l \neq j} \pi(l) q_{l,j} \exp(q_{j,j}(t-u)) du \\
&= \pi(j) \exp(q_{j,j} t) + \sum_{l \neq j} \pi(l) q_{l,j} \left[\frac{\exp(q_{j,j}(t-u))}{-q_{j,j}} \right]_0^t \\
&= \pi(j) \exp(q_{j,j} t) + \sum_{l \neq j} \pi(l) q_{l,j} \frac{(1 - \exp(q_{j,j} t))}{-q_{j,j}}.
\end{aligned}
$$

Multiplying both sides with $-q_{j,j}$ and rearranging terms we get

$$-q_{j,j}\pi(j)(1 - \exp(q_{j,j}t)) = \sum_{l \neq j} \pi(l) q_{l,j} (1 - \exp(q_{j,j}t))$$

and we have the desired identity

$$-q_{j,j}\pi(j) = \sum_{l \neq j} \pi(l) q_{l,j}.$$

□

From a practical point of view to find an invariant distribution for a continuous-time Markov chain start by solving the system of equations from Theorem 51. If a non-trivial solution exists (i.e. not zero in all coordinates) there will always be infinitely many solutions since multiplication by a constant does not alter the system of equations (3.11). Therefore, an important step is to check the existence of a solution that can be normalized into a probability vector of non-negative coordinates with sum one. It is very common that the coordinates of any non-zero solution sum to $+\infty$ so that no normalized solution may be found.

A probability distribution solving the system of equations from Theorem 51 will be a good candidate for an invariant distribution. However, it turns out that for Markov chains with an infinite state space additional conditions are required to ensure that we have indeed found an invariant probability. Note that the condition of the following Theorem 52 is trivially satisfied for Markov chains on a finite state space.

Theorem 52 (Sufficient condition for a stationary distribution) *If $\overline{\pi} = (\pi(i))_{i \in S}$ is a probability satisfying the condition*

$$\forall j \in S \quad : \quad \sum_{i \neq j} \pi(i) q_{i,j} = \pi(j)(-q_{j,j}) = \pi(j) q_j \tag{3.12}$$

of Theorem 51 and furthermore

$$\sum_{j \in S} \pi(j)(-q_{j,j}) < \infty \tag{3.13}$$

then $\overline{\pi} = (\pi(i))_{i \in S}$ is a unique invariant distribution for an irreducible Markov chain.

Sketch of proof: For any fixed state $j \in S$ consider the function $\mu(t) := \sum_{i \in S} \pi(i) P_{i,j}(t)$ and note that $\mu(0) = \pi(j)$. Use the backward differential equations (Theorem 40) and the assumption (3.12) to show that $\mu'(t) = 0$ and conclude that $\sum_{i \in S} \pi(i) P_{i,j}(t) = \mu(t) = \mu(0) = \pi(j)$. □

Example 53 *Let us try to find the invariant distribution for the Markov chain in Example 34. The system of equations $\overline{\pi} Q = 0$ from Theorem 51 becomes*

$$\begin{aligned} \pi(1) \cdot (-4) + \pi(2) \cdot 1 + \pi(3) \cdot 3 &= 0 \\ \pi(1) \cdot 2 + \pi(2) \cdot (-4) + \pi(3) \cdot 1 &= 0 \\ \pi(1) \cdot 2 + \pi(2) \cdot 3 + \pi(3) \cdot (-4) &= 0. \end{aligned}$$

Solving the system by expressing all coordinates in terms of $\pi(1)$ we get that $\pi(2) = 10/13 \cdot \pi(1)$ and $\pi(3) = 14/13 \cdot \pi(1)$. Looking for a solution with

$$1 = \pi(1) + \pi(2) + \pi(3) = \pi(1) \left(1 + \tfrac{10}{13} + \tfrac{14}{13}\right)$$

we get that $\overline{\pi} = (\pi(1), \pi(2), \pi(3)) = \frac{1}{37}(13, 10, 14)$.

Note that just because there is a probability vector, $\overline{\pi}$, solving the system of equations $\overline{\pi} Q = 0$ then we can not be sure that $\overline{\pi}$ is an invariant distribution for the Markov chain. However, since the Markov chain has a finite state space it is trivial to see that the additional condition (3.13) from Theorem 52 holds. We conclude that $\overline{\pi}$ is an invariant distribution for the Markov chain. Further, since the Markov chain is irreducible we get from Theorem 49 that for any initial distribution then

$$\lim_{t \to \infty} P(X(t) = j) = \pi(j), \quad j = 1, 2, 3.$$

□

It might be a bit difficult to understand the role of the additional sufficient condition given in Theorem 52 [23]. The main purpose is to

[23] It is possible to give examples demonstrating that the assumptions in Theorem 51 are not sufficient and that the assumptions in Theorem 52 are not necessary for $\overline{\pi}$ to be an invariant probability.

rule out the possibility of explosion. Note that the condition stated in Theorem 52 always holds if there is a probability vector solving (3.11) and if $(q_{j,j})$ is bounded. Below we give a necessary *and* sufficient condition for a probability $\overline{\pi} = (\pi(i))_{i \in S}$ to be an invariant distribution for a continuous-time Markov chain. This can only be done with reference to the embedded discrete-time Markov chain of jumps.

Theorem 54 *The continuous-time irreducible Markov chain $(X(t))_{t \geq 0}$ has an invariant (or stationary) distribution* [24] *if and only if the embedded discrete-time Markov chain of jumps is recurrent and there exists a probability vector $\overline{\pi} = (\pi(i))_{i \in S}$ such that (3.11) holds or written in a more compact notation such that $\overline{\pi} Q = 0$.*

[24] Theorems 54 or 52 may be used for finding invariant distributions for continuous time Markov chains. Note that for finite communication classes you only have to solve the system of equations given by (3.11) since the additional condition (3.13) of Theorem 52 always holds.

The **proof** of Theorem 54 may be found in Asmussen [2003, Chapter 2] □

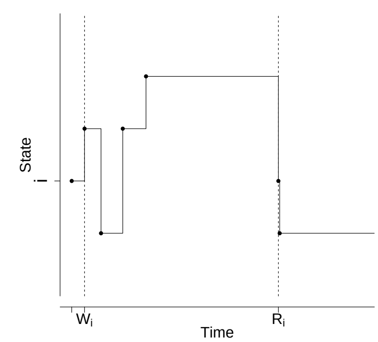
The following result characterizes the invariant distribution for an irreducible Markov chains in terms of mean recurrence (or mean return) times. The result may be regarded as a continuous-time analog to Theorem 22 in Chapter 2.

Theorem 55 (Existence of invariant distributions and positive recurrence) *For an irreducible, recurrent continuous-time Markov chain* [25] *$(X(t))_{t \geq 0}$ define the **escape time** from state i*

[25] Formally, the result does not hold in the trivial case of a continuous-time Markov chain consisting of one (absorbing) state.

$$W_i = \inf\{t \geq 0 | X(t) \neq i\}$$

*and the **return time** to state i*



$$R_i = \inf\{t > W_i | X(t) = i\}.$$

Then an invariant probability $\overline{\pi} = (\pi(i))_{i \in S}$ exists if and only if

$$E[R_i | X(0) = i] < +\infty$$

and we have that

$$\pi(i) = \frac{E[W_i | X(0) = i]}{E[R_i | X(0) = i]} = \frac{1}{q_i E[R_i | X(0) = i]}.$$

is the unique invariant probability [26].

[26] It may be demonstrated that if the mean return time is finite for some i then it is finite for any other state in the same communication class. In particular, positive recurrence is a class property.

The result is also valid when all expectations $E[R_i | X(0) = i] = +\infty$ if we take $\pi(i) = 0$ to mean that no invariant distribution exists.

*We say that a communication class is **positive recurrent** if*

$$E[R_i | X(0) = i] < \infty$$

and note that this is equivalent to existence of an invariant distribution.

□

Example 56 (Continuous-time Random Walk) *The symmetric random walk on $\mathbb{Z}$ was introduced in Example 14 as a discrete time Markov chain with transition probabilities $P_{i,i+1} = P_{i,i-1} = 1/2, i \in \mathbb{Z}$. A possible extension to a continuous-time Markov chain has transition intensities given by*

$$q_{i,i+1} = q_{i,i-1} = \lambda_{|i|},$$

where $\lambda_i > 0, i \in \mathbb{N}_0$. Note that the embedded Markov chain of jumps is exactly the symmetric random walk in discrete-time. Further, to obtain true symmetry around state 0 we have insisted that the transition intensities from states i and $-i$ are given by the same parameter $\lambda_{|i|}$.

Since the symmetric random walk in discrete-time is recurrent (-see Exercise 2.4.2) it follows by Definition 46 that the symmetric continuous-time Markov chain is recurrent regardless of the value of the parameters $\lambda_j, j \in \mathbb{N}_0$.

According to Theorem 54 an invariant probability of the symmetric random walk in continuous time exists if and only if there is a probability vector $\overline{\pi}$ solving the system of equations

$$2 \cdot \lambda_{|i|}\pi(i) = \lambda_{|i-1|}\pi(i-1) + \lambda_{|i+1|}\pi(i+1).$$

Using that we must have $\pi(i) = \pi(-i), i \in \mathbb{N}$, the equation for $i = 0$ yields that $\pi(1) = \frac{\lambda_0}{\lambda_1}\pi(0)$. By iteration it follows that in general

$$\pi(i) = \frac{\lambda_0}{\lambda_{|i|}}\pi(0), i \in \mathbb{Z}.$$

It is possible to normalize the solution into a probability if $\pi(0)$ may be chosen such that

$$1 = \sum_{j\in\mathbb{Z}} \pi(j) = \pi(0)\left(1 + 2\lambda_0 \sum_{j=1}^{\infty} \frac{1}{\lambda_j}\right).$$

We conclude that the continuous-time symmetric random walk on $\mathbb{Z}$ is positive recurrent if and only if $\sum_j \frac{1}{\lambda_j} < +\infty$ [27]. Since the mean waiting time to the next jump from state i is proportional to $\frac{1}{\lambda_{|i|}}$ we observe that positive recurrence requires the mean waiting time between jump to decay sufficiently fast with the distance to state 0. □

[27] Exercise: Given a symmetric discrete probability on $\mathbb{Z}$ with point probabilities $p(z)$. Under what conditions can we find a symmetric random walk in continuous time with invariant probability $\pi(i) = p(i)$?

Example 56 shows that while the embedded Markov chain of jumps is null-recurrent (symmetric random walk in discrete time), then the continuous-time Markov chain can be either null-recurrent or positive recurrent. In Exercise 3.4.4 we demonstrate that we can also have that the embedded Markov chain of jumps is positive recurrent but that the continuous-time Markov chain is null recurrent.

The following result applies to the case of a positive recurrent continuous-time Markov chain where the embedded Markov chain of jumps is also positive recurrent.

Theorem 57 (Time-invariant vs. event-invariant distribution) *Consider an irreducible continous-time Markov chain with transition intensity Q and assume that the invariant distribution* $\overline{\nu} = (\nu(i))_{i\in S}$ *exists. Suppose that we have also verified the existence of an invariant distribution* $\overline{\pi} = (\pi(i))_{i\in S}$ *for the embedded Markov chain of jumps. Then the following relation holds*

$$\pi(i) \quad = \quad \frac{\nu(i)q_i}{\sum_{j\in S}\nu(j)q_j}, \quad i \in S. \qquad (3.14)$$

An irreducible Markov chain $(X(t))_{t\geq 0}$ on a finite state space $|S| < +\infty$ has an invariant probability. To see this note that the embedded Markov chain of jumps is irreducible (hence closed and positive recurrent) and have a unique invariant probability $(\pi(i))_{i\in S}$. The finite (!) vector of non-negative entries given by $\nu(i) = \frac{\pi(i)}{q_i}$ may be normalised to a probability vector that satisfies the system of equations (3.12) and the assumption (3.13) from Theorem 52.

Sketch of proof: *If* $\overline{\nu}$ *is invariant for* $(X(t))_{t\geq 0}$ *then by Theorem 51*

$$\nu(j)q_j = \sum_{i\neq j}\nu(i)q_{ij} = \sum_{i\neq j}\nu(i)q_i\frac{q_{ij}}{q_i}.$$

This shows that $\tilde{\pi}(i) = \nu(i)q_i$ *defines an invariant measure for the embedded Markov chain of jumps with transition probabilties* $P_{ij} = \frac{q_{ij}}{q_i}$*. By Theorem 22 it follows that the unique invariant probability must be given by (3.14).* □

We shall conclude by discussing briefly the limiting behaviour of continuous time Markov chains with more communication classes. It is only relevant to discuss the situation where the Markov chain started in a transient state will eventually enter a recurrent class. Without loss of generality we can restrict our attention to the case where all recurrent states are absorbing. An absorbing state is a state from where the Markov chain cannot escape. State $i \in S$ is absorbing if $q_{i,i} = 0$ or equivalently if $\sum_{j\neq i} q_{i,j} = 0$. A number of interesting questions are related to an absorbing state. First of all we might want to compute the probability that the Markov chain is eventually absorbed in state i. Secondly, we could be interested in the behavior of the Markov chain until absorption for example the average time spend in any other state $j \neq i$ before being caught in state i.

Theorem 58 (Time spent in state j before absorption) *For a continuous-time Markov chain the average number of periods (visits) spent in state j before reaching an absorbing state i (i.e. with* $q_{i,i} = 0$*) may be found by studying the transition probabilities of the embedded discrete-time Markov chain of jumps. For finite state space Markov chains this computation may be carried out using Theorem 28 while you may use Theorem 30 for Markov chains on countably infinite state spaces.*

If N_j *is the mean number of visits to state j before absorption in state i then the average time spend in state j before absorption is given by* $\frac{N_j}{q_j}$*.* □

For another example illustrating the use of Theorem 58 you may have a look at Exercise 3.3.4.

Example 59 *Consider the Markov chain with state space* $S = \{1, 2, 3, 4, 5\}$ *and transition intensities given by Figure 3.5. By looking at the transition* *matrix, P, for the embedded Markov chain of jumps we immediately identify three communication classes $C_1 = \{1\}$ (absorbing), $C_2 = \{2,3\}$ (transient), and $C_3 = \{4,5\}$ (recurrent). Suppose that we start the Markov chain in state 2 how do we find the limiting distribution $\lim_{t\to\infty} P(X(t) = j)$ for $j \in S$?*

To study the behaviour of the Markov chain until the first jump to a recurrent class we consider the modified version, P_{mod}, of the transition matrix for the embedded Markov chain of jumps where recurrent states are turned into absorbing states

$$P_{mod} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Note that we have reordered the states as $\{1,4,5,2,3\}$ such that P_{mod} has a block structure as in Theorem 28

$$P_{mod} = \left(\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array}\right)$$

with

$$S = \begin{pmatrix} 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 \end{pmatrix} \quad Q = \begin{pmatrix} 0 & 1/3 \\ 1/2 & 0 \end{pmatrix}.$$

From the matrix

$$M = (I-Q)^{-1} = \begin{pmatrix} 6/5 & 2/5 \\ 3/5 & 6/5 \end{pmatrix}$$

we get that the Markov chain (started at state 2!) will on average visit state 2 a total of $6/5$ times and state 3 a total of $2/5$ times before absorption to one of the recurrent classes. We conclude from Theorem 58 that the average time spent in the transient states will be $6/5 \cdot 1/3 = 2/5$ for state 2 and $2/5 \cdot 1/2 = 1/5$ for state 3. In particular, the mean waiting time until the Markov chain jumps to a recurrent class may be found as $2/5 + 1/5 = 3/5$.

From the matrix of absorption probabilities

$$A = (I-Q)^{-1}S = \begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 1/5 & 3/5 & 1/5 \end{pmatrix}$$

we conclude that the probability that the Markov chain leaves the transient class C_2 through state 1 is $2/5$. Since state 1 is absorbing it follows that $\lim_{t\to\infty} P(X(t) = 1) = 2/5$ when $X(0) = 2$. The probability that the Markov chain will end its life in the class $C_3 = \{4,5\}$ is $1/5 + 2/5 = 3/5$. Note, however, that this probability mass will in the long run be distributed

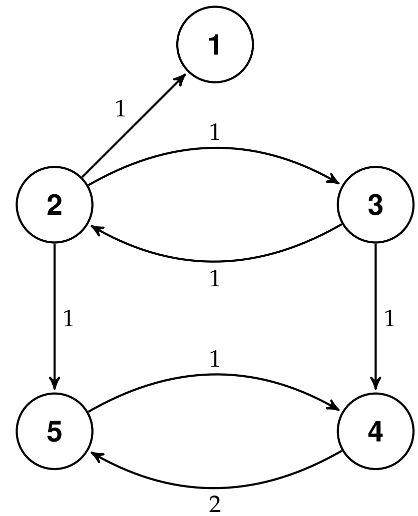


Figure 3.5: Transition diagram for the Markov chain with three communication classes studied in Example 59.

on states 4 and 5 according to the invariant distribution on C_3. The unique invariant distribution on C_3 must satisfy the equation (3.11)

$$2\pi(4) = \pi(5)$$

and we conclude that $\pi(5) = 2/3, \pi(4) = 1/3$. The limiting distribution of the Markov chain given that $X(0) = 2$ may thus be summarised as

$$\lim_{t\to\infty} P(X(t) = j) = \begin{cases} 2/5, & j = 1 \\ 0, & j = 2 \\ 0, & j = 3 \\ 3/5 \cdot 1/3 = 1/5, & j = 4 \\ 3/5 \cdot 2/3 = 2/5, & j = 5 \end{cases}$$

□

Birth-and-death processes

In this section we discuss an important class of continuous-time Markov chains on a countable state space. A **birth-and-death process** is a Markov chain on $S = \mathbb{N}_0$ that allows only jumps (upwards or downwards) of size one. Referring to our usual specification of Markov chains in terms of transition intensities this means that we assume that

$$q_{i,j} = 0, \quad i, j \in \mathbb{N}_0, |i - j| > 1$$

while the only non-zero intensities (except for the diagonal) are

$$\begin{aligned} q_{i,i+1} &= \beta_i, \quad i \in \mathbb{N}_0 \quad \leftarrow \text{ birth intensities} \\ q_{i,i-1} &= \delta_i, \quad i \in \mathbb{N} \quad \leftarrow \text{ death intensities.} \end{aligned}$$

The dynamics of a birth-and-death process is very simple. If the process is currently in state i then the waiting time to the next jump follows an exponential distribution with rate $\beta_i + \delta_i$ (i.e. mean $\frac{1}{\beta_i+\delta_i}$). At the time of the jump the process moves one step up with probability $\beta_i/(\beta_i + \delta_i)$ and one step down with probability $\delta_i/(\beta_i + \delta_i)$.

Example 60 (Transition probabilities for pure birth processes) *A Markov chain $(X(t))_{t\geq 0}$ on $\mathbb{N}_0$ with transition intensities*

$$q_{i,i+1} = -q_{i,i} = \beta_i > 0, i \in \mathbb{N}_0 \quad \textit{and} \quad q_{i,j} = 0, \quad \textit{otherwise}$$

*is called a **pure birth process** . The most famous example of a pure birth process is the **Poisson process** with constant birth intensities*

$$q_{i,i+1} = \lambda, \quad i \in \mathbb{N}_0.$$

The waiting time between jumps constitute a sequence of independent identically distributed random variables from an exponential distribution with *rate parameter λ. If $(X(t))_{t\geq 0}$ is a Poisson process with intensity λ then $X(t)$ follows a Poisson distribution with parameter λt. In Exercise 3.5.1 and 3.5.2 we study various properties of the Poisson process.*

For a general pure birth process the forward differential equations from Theorem 41 take the form

$$P'_{0,n}(t) = P_{0,n}(t) \cdot (-\beta_n) + P_{0,n-1}(t)\beta_{n-1}, \quad n \geq 1.$$

Direct computation shows that

$$y_n(t) = \exp(\beta_n t) P_{0,n}(t)$$

satisfies the differential equation

$$y'_n(t) = \beta_{n-1} \exp(\beta_n t) P_{0,n-1}(t).$$

The solution is given by

$$y_n(t) = \int_0^t \beta_{n-1} \exp(\beta_n s) P_{0,n-1}(s) ds + C$$

and from the initial condition

$$y_n(0) = \exp(\beta_n \cdot 0) P_{0,n}(0) = 0, \quad n \geq 1,$$

we (conclude that $C = 0$ and) get the recursive formula

$$P_{0,n}(t) = \exp(-\beta_n t) y_n(t) = \exp(-\beta_n t) \int_0^t \beta_{n-1} \exp(\beta_n s) P_{0,n-1}(s) ds.$$

Using that $P_{0,0}(t) = \exp(-\beta_0 t)$ the transition probabilities may be computed iteratively. If all birth intensities are different (i.e. if $\beta_i \neq \beta_j$ for $i \neq j$) then

$$P_{0,n}(t) = \beta_0 \cdot \ldots \cdot \beta_n \cdot \sum_{i=0}^{n} \exp(-\beta_i t) \frac{1}{\prod_{j=0, j\neq i}^{n} (\beta_j - \beta_i)}.$$

For the linear birth process with immigration given by $\beta_n = \beta n + \lambda$ for $\beta, \lambda > 0$ then the formula simplifies to

$$P_{0,n}(t) = \binom{\frac{\lambda}{\beta} + n - 1}{n} \left(\exp(-\beta t)\right)^{\lambda/\beta} \left(1 - \exp(-\beta t)\right)^n.$$

In particular, for $X(0) = 0$ then $X(t)$ follows a negative binomial distribution with parameter $(\lambda/\beta, \exp(-\beta t))$. □

Consider the time, τ_n, of the $n-$th jump for a birth-and-death process. If the process is absorbed before the n-th jump then we let $\tau_n = +\infty$. Clearly, $(\tau_n)_{n\geq 0}$ is increasing

$$\tau_1 \leq \tau_2 \leq \ldots \leq \tau_n$$

and we may define the variable

$$\tau_\infty = \lim_{n\to\infty} \tau_n$$

with values in $[0, +\infty]$. From a mathematical point of view it is easy to give examples of birth intensities, $(\beta_i)_{i\in\mathbb{N}_0}$, and death intensities, $(\delta_i)_{i\in\mathbb{N}}$, such that

$$P(\tau_\infty = +\infty) < 1$$

or in other words such that there is a strictly positive probability of observing an infinite number of jumps in finite time. In this situation we will say that explosion is possible or that the transition intensities allow for explosion.

Example 61 (Pure birth process with explosion) *Consider a pure birth process with $P(X(0) = 0) = 1$. Then we know that the $n-th$ jump will go from state $n-1$ to n with an expected waiting time of $1/\beta_{n-1}$. The expected time of the $n-th$ jump will hence be*

$$E[\tau_n|X(0) = 1] = \sum_{i=0}^{n-1} 1/\beta_i$$

and by monotone convergence

$$E[\tau_\infty|X(0) = 1] = \lim_{n\to\infty} E[\tau_n|X(0) = 1] = \lim_{n\to\infty} \sum_{i=0}^{n-1} 1/\beta_i.$$

In particular, if $\sum_{i=0}^{\infty} 1/\beta_i < \infty$ then τ_∞ has finite mean and we conclude that

$$P(\tau_\infty = +\infty|X(0) = 1) = 0.$$

We conclude that for a pure birth process then $\sum_{i=0}^{\infty} 1/\beta_i < \infty$ implies that there will be infinitely many jumps in finite time (=explosion) with probability 1! □

Using the recurrence criterion given in Theorem 15 we get a simple characterization of recurrent birth-and-death processes.

Theorem 62 (Birth-and-death processes: recurrence) *A birth-and-death process is recurrent if and only if*

$$\sum_{i=1}^{\infty} \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_1} = \infty. \tag{3.15}$$

Equivalently, a birth-and-death process is transient if and only if

$$\sum_{i=1}^{\infty} \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_1} < \infty. \tag{3.16}$$

Proof: We consider only the case where all $\beta_i, \delta_i > 0$. According to Definition 46 the birth-and-death process is recurrent if and only if the embedded Markov chain of jumps is recurrent. The jump chain has transition probabilities

$$P_{i,i+1} = \frac{\beta_i}{\beta_i + \delta_i}, \quad P_{i,i-1} = \frac{\delta_i}{\beta_i + \delta_i}, i \geq 1, \quad P_{0,1} = 1.$$

Take $i = 0$ in equation (2.7) for the recurrence criterion of Theorem 15 to get

$$\begin{aligned} \alpha(1) &= \sum_{k \neq 0} P_{1,k}\alpha(k) = P_{1,2} \cdot \alpha(2) \\ \alpha(j) &= \sum_{k \neq 0} P_{j,k}\alpha(k) = P_{j,j+1} \cdot \alpha(j+1) + P_{j,j-1} \cdot \alpha(j-1), \quad j > 1. \end{aligned}$$

where the last equality follows from the fact that only jumps of size one are possible. Using that $\alpha(j) = P_{j,j+1}\alpha(j) + P_{j,j-1}\alpha(j)$ we get by iteration

$$\begin{aligned} \alpha(j+1) - \alpha(j) &= \frac{P_{j,j-1}}{P_{j,j+1}} \left(\alpha(j) - \alpha(j-1)\right) \\ &= \left(\prod_{k=2}^{j} \frac{P_{k,k-1}}{P_{k,k+1}} \right) \left(\alpha(2) - \alpha(1)\right) \\ &= \left(\prod_{k=2}^{j} \frac{\frac{\delta_k}{\beta_k + \delta_k}}{\frac{\beta_k}{\beta_k + \delta_k}} \right) \left(\alpha(2) - \alpha(1)\right) \\ &= \left(\prod_{k=2}^{j} \frac{\delta_k}{\beta_k} \right) \left(\alpha(2) - \alpha(1)\right), \quad j > 1. \end{aligned}$$

Summing over $j = 2, ..., n$ and using the boundary condition $\alpha(1) = P_{1,2}\alpha(2)$ we get

$$\begin{aligned} \alpha(n+1) - \alpha(2) &= \sum_{j=2}^{n} \left(\prod_{k=2}^{j} \frac{\delta_k}{\beta_k} \right) \left(\alpha(2) - \alpha(1)\right) \Leftrightarrow \\ \alpha(n+1) &= \sum_{j=2}^{n} \left(\prod_{k=2}^{j} \frac{\delta_k}{\beta_k} \right) (1 - P_{1,2})\alpha(2) + \alpha(2). \end{aligned}$$

By letting $\alpha(2) = 0$ we see that $\alpha(i) = 0$, $i \geq 1$ is a solution to (2.7) (-remember that we assume $\delta_1, \beta_1 > 0$ which implies that $P_{1,2} \neq 1$). The only nonzero solution to (2.7) is obtained by $\alpha(2) \neq 0$ and the solution will be bounded if the coefficient

$$\sum_{j=2}^{n} \left(\prod_{k=2}^{j} \frac{\delta_k}{\beta_k} \right) (P_{1,2} - 1)$$

does not tend to $+\infty$ for $n \to \infty$. We conclude that nonzero bounded solutions exist if and only if

$$\sum_{j=2}^{\infty} \left(\prod_{k=2}^{j} \frac{\delta_k}{\beta_k} \right) < +\infty$$

which is equivalent to the sum in Theorem 62 being finite. Finally, use that Theorem 15 states that a Markov chain is transient if and only if nonzero bounded solutions to (2.7) exist. □

Theorem 63 (Birth-and-death processes: positive recurrence) *A birth-and-death process is positive recurrent if and only if*

$$\sum_{i=1}^{\infty} \frac{\beta_{i-1} \cdot \ldots \cdot \beta_0}{\delta_i \cdot \ldots \cdot \delta_1} < \infty \quad \textit{and} \quad \sum_{i=1}^{\infty} \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_1} = \infty \tag{3.17}$$

Sketch of proof of Theorem 63 We know from equation (3.11) of Theorem 51 that the invariant distribution of a continuous-time Markov chain must satisfy the system of equations

$$\forall j \in S \quad : \quad \sum_{i \neq j} \pi(i) q_{i,j} = \pi(j)(-q_{j,j}) = \pi(j) q_j.$$

For a birth-and-death process the system of equations takes the form

$$\begin{aligned} \forall j \in \mathbb{N} \quad &: \quad \pi(j-1)\beta_{j-1} + \pi(j+1)\delta_{j+1} = \pi(j)(\beta_j + \delta_j) \\ j = 0 \quad &: \quad \pi(1)\delta_1 = \pi(0)\beta_0 \end{aligned}$$

which turns out to have the solution

$$\pi(j) \quad = \quad \pi(0) \cdot \left(\prod_{i=1}^{j} \frac{\beta_{i-1}}{\delta_i} \right) \tag{3.18}$$

that can be normalized into a probability vector provided that

$$\sum_{i=1}^{\infty} \frac{\beta_{i-1} \cdot \ldots \cdot \beta_0}{\delta_i \cdot \ldots \cdot \delta_1} < \infty.$$

You are reminded of Theorem 54 which tells us that positive recurrence of a continuous-time Markov chain requires both a solution to (3.11) and that the Markov chain is demonstrated to be recurrent. This is the reason that two conditions must be given in Theorem 63. □

From Example 61 we know that a pure birth process can have infinitely many jumps on a finite time interval (i.e. explosion may occur!). It is not possible to give a simple condition on the transition intensities for a continuous-time Markov chain that determines exactly when explosion is possible. For birth-and-death processes things are little easier as we have the following result.

Theorem 64 (Explosion for a birth-and-death processes) *For a birth-and-death process with intensities*

$$q_{i,i+1} = \beta_i, \quad q_{i+1,i} = \delta_{i+1}, \quad q_{i+1,i+1} = -(\delta_{i+1} + \beta_{i+1})$$

and $\quad q_{i,j} = 0$ *otherwise,* $\quad i, j \in \mathbb{N}_0$ *then explosion is possible if and only if* [28]

$$\sum_{i=1}^{\infty} \left(\frac{1}{\beta_i} + \frac{\delta_i}{\beta_i \beta_{i-1}} + \ldots + \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_0} \right) < +\infty. \tag{3.19}$$

The inequality (3.19) is often referred to as Reuter's criterion for explosion.
Proof: *The proof of Reuter's criterion may be found in Asmussen [2003].* □

[28] From Theorem 62 we observe that if (3.19) holds then the birth-and-death process may be transient. Equivalently, for a recurrent birth-and-death process then explosion is not possible.

Example 65 (Birth-and-death process) *We consider the birth-and-death process with birth intensities*

$$\beta_i = \beta \cdot (1+i)^{\alpha}, \quad i \geq 0,$$

and death intensities

$$\delta_i = \delta \cdot i^{\alpha}, \quad i \geq 1,$$

for parameters $\alpha, \beta, \delta > 0$.

From Theorem 62 we have that the birth-and-death process is recurrent if and only if

$$\begin{aligned} +\infty &= \sum_{i=1}^{\infty} \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_1} = \sum_{i=1}^{\infty} \frac{\delta^i i^{\alpha} \cdot (i-1)^{\alpha} \cdot \ldots \cdot 1^{\alpha}}{\beta^i (i+1)^{\alpha} \cdot i^{\alpha} \cdot \ldots \cdot 2^{\alpha}} \\ &= \sum_{i=1}^{\infty} \left(\frac{\delta}{\beta} \right)^i (1+i)^{-\alpha}. \end{aligned}$$

Letting $a_i = \left(\frac{\delta}{\beta} \right)^i (1+i)^{-\alpha}$ *we get that*

$$\frac{a_{i+1}}{a_i} = \frac{\delta}{\beta} \left(\frac{i+1}{i+2} \right)^{\alpha} \overset{i \to \infty}{\to} \begin{cases} < 1 & \delta < \beta \\ 1 & \delta = \beta \\ > 1 & \delta > \beta \end{cases}$$

*From the **ratio test*** [29] *we conclude that for any value of* $\alpha > 0$ *then the birth-and-death process is recurrent for* $\delta > \beta$ *and transient for* $\beta > \delta$*. For the special case where* $\beta = \delta$ *we have that*

$$\sum_{i=1}^{\infty} \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_1} = \sum_{i=1}^{\infty} (1+i)^{-\alpha}.$$

[29] See page 138 in Appendix B.6.

*Using the **integral test*** [30] *we conclude that for* $\delta = \beta$ *then the birth-and-death process is recurrent for* $\alpha \leq 1$ *and transient for* $\alpha > 1$.

[30] See page 138 in Appendix B.7.

For the birth-and-death process to be positive recurrent we get from Theorem 63 that it must be recurrent and that we must have that

$$\sum_{i=1}^{\infty} \frac{\beta_{i-1} \cdot \ldots \cdot \beta_0}{\delta_i \cdot \ldots \cdot \delta_1} = \sum_{i=1}^{\infty} \left(\frac{\beta}{\delta} \right)^i < +\infty.$$

We conclude that the birth-and-death process is positive recurrent for $\beta < \delta$, null recurrent for $\beta = \delta, \alpha \le 1$ and transient for $\beta = \delta, \alpha > 1$ or $\beta > \delta$.

Finally, we use Theorem 64 to discuss if explosion is possible. We find that

$$\begin{aligned}
& \sum_{i=1}^{\infty} \left(\frac{1}{\beta_i} + \frac{\delta_i}{\beta_i \cdot \beta_{i-1}} + \ldots + \frac{\delta_i \cdot \ldots \cdot \delta_1}{\beta_i \cdot \ldots \cdot \beta_0} \right) \\
= & \sum_{i=1}^{\infty} \left(\frac{1}{\beta \cdot (1+i)^\alpha} + \frac{\delta \cdot i^\alpha}{\beta \cdot (1+i)^\alpha \beta \cdot i^\alpha} + \ldots + \frac{\delta \cdot i^\alpha \cdot \ldots \cdot \delta \cdot 1^\alpha}{\beta \cdot (1+i)^\alpha \cdot \ldots \cdot \beta \cdot 1^\alpha} \right) \\
= & \sum_{i=1}^{\infty} \frac{1}{\beta \cdot (1+i)^\alpha} \left(1 + \frac{\delta}{\beta} + \ldots + \frac{\delta^i}{\beta^i} \right).
\end{aligned}$$

Clearly, for $\delta > \beta$ the series is divergent ($= +\infty$) since the terms do not even go to zero as $i \to \infty$. For $\delta = \beta$ the series is

$$\sum_{i=1} \frac{1}{\beta \cdot (1+i)^\alpha} (1+i) = \sum_{i=1}^{\infty} \frac{1}{\beta \cdot (1+i)^{\alpha-1}}$$

which (according to the integral test) is convergent if and only if $\alpha - 1 > 1$ i.e. if $\alpha > 2$. We conclude that for $\beta = \delta$ and $\alpha > 2$ then explosion is possible.

For the case $\beta > \delta$ the i-th term, a_i, of the series may be bounded as follows

$$\begin{aligned}
\frac{1}{\beta \cdot (1+i)^\alpha} \le a_i &= \frac{1}{\beta \cdot (1+i)^\alpha} \sum_{j=0}^{i} \left(\tfrac{\delta}{\beta} \right)^j \\
&\le \frac{1}{\beta \cdot (1+i)^\alpha} \sum_{j=0}^{\infty} \left(\tfrac{\delta}{\beta} \right)^j = \frac{1}{\beta \cdot (1+i)^\alpha} \frac{1}{1 - \left(\frac{\delta}{\beta} \right)}.
\end{aligned}$$

Using the integral test we conclude that for $\beta > \delta$ then the series is finite and hence explosion is possible exactly if $\alpha > 1$. □

Exercises on Markov chains in continuous time

Markov chains with two states

3.1.1 Two state absorbing Markov chain

In life insurance mathematics a two state Markov chain with one absorbing state is often used to model a single life with one cause of death. This corresponds to the following transition diagram where the states are labelled as *alive* or *dead*

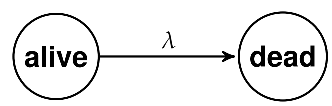


Figure 3.6: Transition diagram for Markov chain in Exercise 3.1.1.

Let $(X(t))_{t\geq 0}$ be the Markov chain given by the diagram above, and assume that the person is *alive* at time $t = 0$.

1. What is (by definition) the distribution of the first (and only) jump time, τ_1?

2. Use the general formula for conditional probabilities

$$P(A|B) = P(A \cap B)/P(B)$$

   to compute $P(\tau_1 > s + t|\tau_1 > t)$ for $s, t > 0$. Give an interpretation of the result.

3. Assuming that this is a reasonable model for the life time of a Danish woman and that the mean life duration is 80 years what is then the probability of surpassing the age of 100 years given that one has already passed the age of 80 years?

4. Find the distribution of $X(t)$.

5. Consider n single lifes given by the absorping two-state Markov chain above. Let $N(t) = \sum_{i=1}^{n} 1(X_i(t) = \text{alive})$ be the number of individuals alive at time t (i.e. $X_i(t)$ is the state of i-th person at time t). Find $E(N(t))$ and discuss what could be the distribution of $N(t)$.

3.1.2 *Two state Markov chain with equal intensities*

Let $(X(t))_{t\geq 0}$ be the Markov chain given by the transition diagram on Figure 3.7, where we assume that the transition intensities $\lambda_1 = \lambda_2 \equiv \lambda$ are the same in both states. We further assume that the initial distribution is given by $P(X(0) = 1) = 1$.

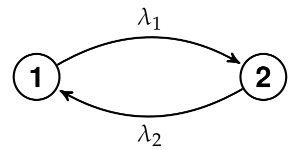


Figure 3.7: Transition diagram for Markov chain in Exercise 3.1.2 and 3.1.3.

1. Find the distribution of the jump times $\tau_1, \tau_2, \tau_3 \ldots$, by referring to results from other exercises or from Appendix B.1.

2. Let $N(t)$ denote the total number of jumps of the chain on the time interval $[0, t]$. Express the distribution of $X(t)$ (i.e. the probabilities $P(X(t) = 1)$ and $P(X(t) = 2)$) in terms of $N(t)$.

3. Find the distribution of $N(t)$ by referring to results from other exercises or from Appendix B.1 and use this to obtain a formula for the distribution of $X(t)$.

4. Find the distribution of $X(t)$ under an arbitrary initial distribution given by $P(X(0) = 1) = p \in [0, 1]$. Does the distribution of $X(t)$ depend on t?

5. Show that for any initial distribution then the limits

$$\nu_1 := \lim_{t\to\infty} P(X(t) = 1) \quad \text{and} \quad \nu_2 := \lim_{t\to\infty} P(X(t) = 2)$$

exist. Do ν_1 and ν_2 depend on the initial distribution?

3.1.3 *Transition probabilities for a two state chain*

Consider the general two-state Markov chain, $(X(t))_{t\geq 0}$, given by transition diagram on Figure 3.7. The corresponding transition matrix becomes

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}.$$

The general result says that for $i, j \in \{1, 2\}$ then the transition probabilities

$$P_{i,j}(s) := P(X(t+s) = j | X(t) = i), \quad t, s \geq 0,$$

are given by the entries of the *exponential matrix*

$$\exp(Qt) = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!},$$

where Qt is the matrix obtained by multiplying each entry of Q by t. The purpose of this exercise is to find closed form expressions for $P_{i,j}(s)$ for the general two-state Markov chain.

1. Suppose that we can find an invertible matrix U and a diagonal matrix
$$D = \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}$$
such that $Q = UDU^{-1}$. Argue that $(Qt)^n = U(D^n t^n)U^{-1}$ and deduce that
$$\exp(Qt) = U \begin{pmatrix} \exp(\delta_1 t) & 0 \\ 0 & \exp(\delta_2 t) \end{pmatrix} U^{-1}.$$

2. If U has entries u_{ij} and U^{-1} has entries $u_{ij}^{-1}, i, j \in \{1, 2\}$ write down the formula for $P_{i,j}(t)$ which is given as the $ij-$th entry of $\exp(Qt)$ from question 1. above.

The last two questions 3.-4. show that it is always possible to obtain the representation $Q = UDU^{-1}$ given in question 1. above. This implies that for a two-state Markov chain then the transition probabilities
$$P(X(t+s) = j | X(t) = i)$$
are given as linear combinations of two exponential functions $\exp(\delta_i s), i = 1, 2$.

3. Find expressions for δ_1, δ_2 (given as the eigenvalues of Q) by solving the equation
$$0 = \det(Q - \delta I) = \det \begin{pmatrix} -\lambda_1 - \delta & \lambda_1 \\ \lambda_2 & -\lambda_2 - \delta \end{pmatrix}.$$

4. For each of the eigenvalues δ_1, δ_2 find the coordinates $u_{1j}, u_{2j}, j = 1, 2$, of (right) eigenvectors for Q with eigenvalues δ_j, by solving the system of equations
$$Q \begin{pmatrix} u_{1j} \\ u_{2j} \end{pmatrix} = \delta_j \begin{pmatrix} u_{1j} \\ u_{2j} \end{pmatrix}$$

and verify that $Q = UDU^{-1}$.

Markov chains with three states

3.2.1 Model for interest rate

In this exercise we consider the three-state Markov chain with $q_{1,3} = q_{3,1} = 0$ given by the transition diagram on Figure 3.8. Note that the model does not allow jumps between states 1 and 3. The model may for instance be used to describe an interest rate that may jump between three different levels but where direct jumps from lowest to highest level do not occur.

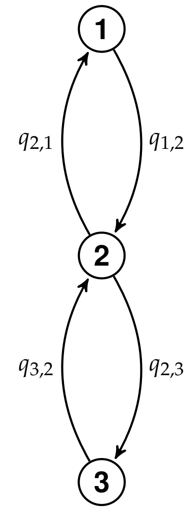


Figure 3.8: Transition diagram for the Markov chain studied in Exercise 3.2.1.

1. Write down the transition matrix, Q, for the Markov chain.
2. Find the limit distribution, $\pi(j) = \lim_{n\to\infty} P(X(t) = j)$, for the Markov chain.
3. Write down the matrix of transition probabilities, P, for the discrete-time Markov chain describing the jumps of the chain.
4. Find the invariant distribution for the discrete-time Markov chain given by P, and discuss when the probabilities of questions 2. and 4. coincide.

Assume in the following that $X(0) = 2$ and that all non-zero entries of Q are the same, i.e. $q_{1,2} = q_{2,1} = q_{2,3} = q_{3,2} = q$. Denote by $\tau_n, n \geq 1$, the time of the n-th jump of the Markov chain and let $N(t)$ be the number of jumps of the Markov chain on the interval $[0, t]$.

5. Argue that the distributions of τ_1 and $\tau_2 - \tau_1$ are exponential and find the rate parameters.
6. Using that τ_2 is the sum of the two independent random variables τ_1 and $\tau_2 - \tau_1$ show by applying the formula of Exercise 3.5.1 that τ_2 has density

$$g_2(t) = 2q(\exp(-qt) - \exp(-2qt)), t \geq 0.$$

7. Express the event $(X(t) = 2)$ in terms of events of the form $(N(t) = n)$.
8. Find a formula for the probabilities $P(X(t) = 1)$, $P(X(t) = 2)$, and $P(X(t) = 3)$ in terms of the (unknown) probabilities

$$p_n = P(N(t) = n), n \in \mathbb{N}_0.$$

Assume in the following that $X(0) = 1$ and that all non-zero entries of Q are the same, i.e. $q_{1,2} = q_{2,1} = q_{2,3} = q_{3,2} = q$.

9. Use the ideas from questions 5.-8. to express $P(X(t) = 2)$ in terms of p_n from question 8.

The computations above give us expressions for

$$P_{i,j}(s) = P(X(t+s) = j|X(t) = i)$$

for certain values of $i, j \in S = \{1, 2, 3\}$. Remember that in general the transition probability $P_{i,j}(s)$ is given as the ij-th entry of the exponential matrix $\exp(Qs)$.

10. For what values of $i, j \in S$ did we obtain expressions for $P_{i,j}(s)$ in terms of $p_n = P(N(t) = n)$ by the results of questions 7.-9?

11. Argue that $v_1 = (1, 1, 1)^T$, $v_2 = (1, 0, -1)^T$, and $v_3 = (1, -2, 1)^T$ are right eigenvectors for Q and find the corresponding eigenvalues λ_1, λ_2, and λ_3.

12. Find a matrix, O, and a diagonal matrix D such that

$$QO = OD.$$

13. Find the transition probabilites $P_{i,j}(s)$ by computing $\exp(Qs)$.

    Hint: Argue first that $\exp(Qs) = O \exp(Ds) O^{-1}$.

3.2.2 *Model with two states of health and death*

In this exercise we consider a model that may be used to describe insurances with payments depending on the state of the insured. We assume that the insured starts in state 0 (=*active0*). After a while the insured enters a more favorable state 1 (=*active1*)where she or he stays until death represented by state 2 (=*dead*). To put the model into a more practical setting we might label the states as *active0*, *active1*, and *dead*.

Assume that $X(0) = 0$ and denote by

$$T_1 = \inf\{t > 0 | X(t) = 1\}$$

the time of the jump to state 1. Further, let

$$P_{i,j}(t) = P(X(t+s) = j|X(s) = i), \quad s, t \geq 0,$$

be the transition probabilities of the Markov chain.

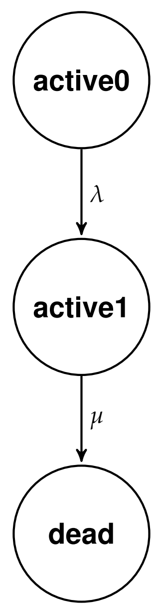


Figure 3.9: Transition diagram on the model with two states of health studied in Exercise 3.2.2.

1. Find the matrix, Q, of transition intensities and explain for what $i, j \in S = \{0, 1, 2\}$ it holds that $P_{i,j}(t) = 0, t \geq 0$.

2. Write down the backward differential equation for $P_{0,0}(t)$ and determine

$$P(X(t) = 0).$$

3. Determine $P(T_1 > t)$ and the expectation $E[T_1]$ (You do not have to redo the formal computations as the result should be known!).

The purpose of the following questions is to find a formula for the transition probabilities $P_{0,2}(t)$.

4. Find $P_{2,2}(t)$.

5. Write down the backward differential equation for $P_{1,1}(t)$ and determine $P(X(t+s) = 1|X(s) = 1)$.

6. Find $P_{1,2}(t)$.

7. Argue that $P'_{0,2}(t) = -\lambda P_{0,2}(t) + \lambda P_{1,2}(t)$.

8. Define the function

$$h(t) = \exp(\lambda t) P_{0,2}(t)$$

and deduce from question 7. that

$$h'(t) = \lambda \exp(\lambda t) P_{1,2}(t).$$

9. Use the expression for $P_{1,2}(t)$ from question 6. and the boundary condition $h(0) = 0$ to solve the differential equation from question 8. to get a formula for $h(t)$.

10. Find a closed form expression for $P_{0,2}(t)$.

3.2.3 *Model for disabilities, recoveries, and death*

A model suitable for analysing insurances with payments depending on the state of health of the insured may be given by the three state Markov chain with transition intensities indicated in Figure 3.10.

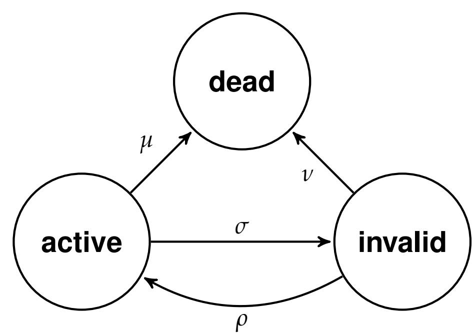


Figure 3.10: Transition diagram for Markov chain of Exercise 3.2.3.

Consider a portfolio for a person with initial state $X(0) = \textit{active}$ and denote by

$$\tau = \inf\{t > 0 | X(t) = \textit{dead}\}$$

the life length. To the insurance company it is important to know the distribution of τ. Further, if the payments depend on the state of the insured (*active*/*invalid*) it is important to study the duration of the time spend in each of the states before absorption in the final state *dead*.

To simplify the notation below we relabel the states such that $0 = \textit{active}$, $1 = \textit{invalid}$, and $2 = \textit{dead}$. As usual we denote by

$$P_{i,j}(s) = P(X(t+s) = j | X(t) = i), \quad s, t \geq 0,$$

the transition probabilities of the Markov chain.

1. Write down the transition matrix, Q.

2. For what i, j does it hold that $P_{i,j}(s) = 0$?

3. Write down the forward differential equations for the transition probabilities $P_{i,j}(s)$ for $i = 0$ (=*active*).

Assume in the following questions 4.-9. that $\nu = \mu$.

4. Use question 3. and the fact that $P_{0,0}(t) + P_{0,1}(t) + P_{0,2}(t) = 1$ to obtain a simplified differential equation for $P_{0,2}(t)$ for $\nu = \mu$.

5. Find the distribution of the survival time, τ, for $\nu = \mu$.

   Hint: First note that $P(\tau \leq t) = P_{0,2}(t)$ and then find (or guess!) the solution to the differential equation of question 4.

6. Use question 3.+5. and that $P_{0,0}(t) + P_{0,1}(t) + P_{0,2}(t) = 1$ to obtain an equation for $P'_{0,1}(t)$ that involves $P_{0,1}(t)$ but no other transition probabilities $P_{i,j}(t)$. Solve the differential equation and find $P_{0,1}(t)$.

The total time spend in the active state ($= 0$) may formally be expressed as

$$S_0 = \int_0^\infty 1(X(t) = 0)dt.$$

In a similar way we define

$$S_1 = \int_0^\infty 1(X(t) = 1)dt$$

that is the time spend in state 1 (=*invalid*). Note that we have the following formula

$$E[S_i] = E\left[\int_0^\infty 1(X(t) = i)dt\right] = \int_0^\infty P_{0,i}(t)dt$$

that may be useful for computing $E[S_i]$ when the transition probabilities are known.

7. Use the results of questions 4.-6. to obtain an expression for $E[S_i], i = 0, 1$, (-still assuming that $\mu = \nu$).

8. Compute $P(S_1 = 0)$.

9. Use question 5. and 8. to obtain an expression for $P(S_1 = 0, \tau \leq t)$ and compute the conditional probability

   $$P(S_1 = 0 | \tau \leq t)$$

   that a person dying before time t did not spend any time in the state 1 =*invalid*.

3.2.4 *Model for single life with 2 causes of death*

In the following questions let $P_{i,j}(t) = P(X(t+s) = j|X(s) = i)$ denote the transition probabilities and assume that $X(0) = alive$. For simplicity we recode the state space, S, such that: $0 = alive, 1 = dead\ 1, 2 = dead\ 2$.

1. Find the intensity matrix of the chain.
2. Determine the communication classes of the chain and argue for each class whether it is recurrent or transient.
3. For which $i, j \in S$ does it hold that $P_{i,j}(t) = 0$ or $P_{i,j}(t) = 1$?
4. Write down the backward equations for the non-constant transition probabilities.
5. Determine $P_{0,0}(t)$ (i.e. the probability of being alive at time t).
6. Find expressions for the remaining transition probabilities.
7. Assuming that $X(0) = alive$ find the probability that the process will eventually be absorbed in state *dead1*.

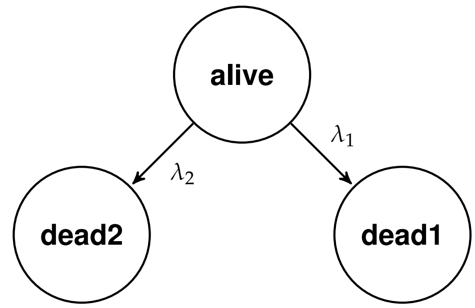


Figure 3.11: Transition diagram for Markov chain used as a model for a single life with with 2 causes of death in Exercise 3.2.4.

3.2.5 *Model with one zero in the transition matrix*

We consider a Markov chain $(X(t))_{t\geq 0}$ with transition diagram given by Figure 3.12 and assume that $P(X(0) = 3) = 1$.

To solve questions 6.+7. you might find it useful to know that the equation

$$f'(t) = \alpha f(t) + \beta \exp(\gamma t) + \delta$$

has a solution of the form

$$f(t) = c_1 \cdot \exp(\gamma t) + c_2 \cdot \exp(\alpha t) + c_3$$

for $\gamma \neq \alpha$ and c_1, c_2, c_3 suitable constants.

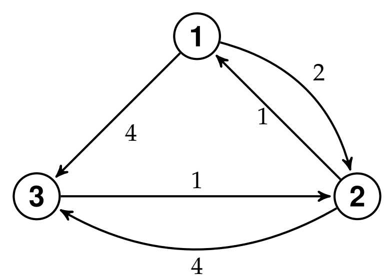


Figure 3.12: Transition diagram for Markov chain of Exercise 3.2.5.

1. Find the infinitesimal generator, Q, (=intensity matrix) for the chain.
2. Find the transition probability matrix for the Markov chain of jumps.
3. Write down the system of equations for the invariant probability $\overline{\pi}$.
4. Compute $\overline{\pi}$.
5. Write down the forward differential equations for $P_{3,j}(t), j = 1, 2, 3$.
6. Use that $P_{3,1}(t) + P_{3,2}(t) + P_{3,3}(t) = 1$ to find $P_{3,3}(t)$.

7. Find $P_{3,1}(t)$ and $P_{3,2}(t)$.

8. Let $\tau_1 = \inf\{t > 0 | X(t) = 1\}$ be the time of the first visit to state 1. Determine $E\tau_1$.

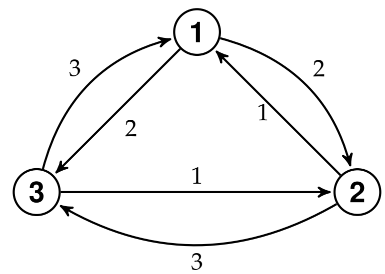


Figure 3.13: Transition diagram for Markov chain of Exercise 3.2.6.

3.2.6 A numerical example

1. Suppose that the chain starts in state 1 and let

$$\tau_1 = \inf\{t > 0 | X(t) \neq 1\}$$

be the time of the first jump. What is the mean $E\tau_1$ of τ_1?

2. Find the matrix, Q, of transition intensities and the transition matrix, P, for the jumps.

3. Argue (briefly) that $P(X(\tau_1) = 2) = 1/2$.

4. What is the distribution of the time between the first, τ_1, and the second, τ_2, jump if $X(\tau_1) = 2$.

5. What is the distribution of the time between the first and the second jump if $X(\tau_1) = 3$.

6. What is the distribution of τ_2?

7. Give an argument that the time, τ_n, of the n-th jump follows a $\Gamma-$distribution and find the parameters.

8. Find the equilibrium distribution $\overline{\pi}$ of the Markov chain.

The transition probabilities, $P(t)$, are given as the entries of the exponential matrix $\exp(Qt)$. The following question and the comments below give some more insight on the possible form of the transition probabilities. This part may be skipped if you are not familiar with complex numbers.

9. Find the characteristic polynomial $g(\lambda) = \det(Q - \lambda I)$ and show that g has one real root ($=$ 0 of course!) and two complex roots ($=-6 \pm i$).

Remark: We might continue to find a matrix O of linearly independent (column) eigenvectors for Q and compute the transition probabilities as

$$P(t) = O \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-6t}\exp(it) & 0 \\ 0 & 0 & e^{-6t}\exp(-it) \end{pmatrix} O^{-1}.$$

The eigenvectors will contain complex numbers but since we know that $P_{i,j}(t)$ are probabilities (in particular real numbers) all complex coefficients *must cancel when we compute the matrix products. Consequently, since by definition*

$$\exp(it) = \cos(t) + i \cdot \sin(t)$$

we can immediately conclude that all transition probabilities take the form

$$P_{i,j}(t) = a_{i,j} + b_{i,j} \exp(-6t) \cdot \cos(t) + c_{i,j} \exp(-6t) \cdot \sin(t),$$

for suitable real *constants* $a_{i,j}, b_{i,j}, c_{i,j}$*. We can actually use the fact that* $P_{i,j}(t) \to \pi_j$ *for* $t \to \infty$ *to conclude that* $a_{i,j} = \pi_j$*. Further, we have that*

$$\begin{aligned} 1 &= P_{i,i}(0) = a_{i,i} + b_{i,i} + c_{i,i} = \pi_i + b_{i,i} + c_{i,i} \Rightarrow b_{i,i} = 1 - \pi_i - c_{i,i} \\ 0 &= P_{i,j}(0) = a_{i,j} + b_{i,j} + c_{i,j} = \pi_j + b_{i,j} + c_{i,j} \Rightarrow b_{i,j} = -\pi_j - c_{i,j}, \quad i \neq j, \end{aligned}$$

showing that only the constants $c_{i,j}$ *need to be determined. Finally, using that* $\sum_j P_{i,j}(t) = 1$ *we get the additional constraint* $\sum_j c_{i,j} = 0$ *for any* i*. A system of equations for the remaining (6!) undetermined constants,* $c_{i,j}$*, may be obtained by the forward or backward differential equations for* $P(t)$*.*

Finite state space

3.3.1 Model for two lifes

Consider a Markov chain, $(X(t))_{t\geq 0}$, with four states given by transition diagram

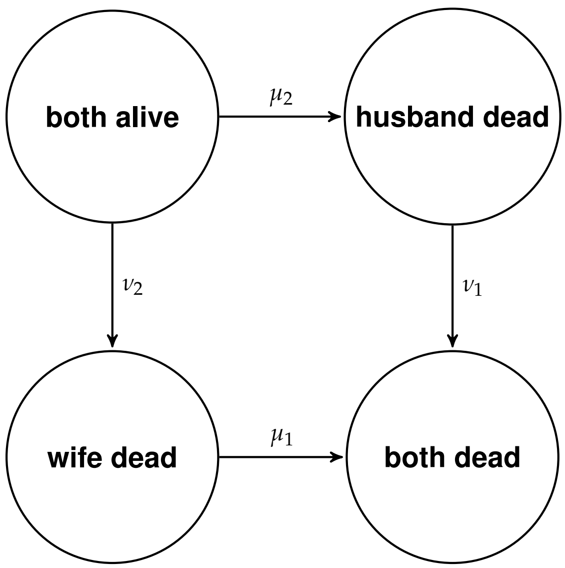


Figure 3.14: Transition diagram of model for two lifes from Exercise 3.3.1.

One may think of the model as a description of the two lifes of a married couple that wants to buy a combined life insurance and widow's pension policy.

In the following we assume that at both persons are alive at time $t = 0$.

1. Write down the transition matrix of the Markov chain and find the distribution of the first jump time, τ_1.
2. Find the probability that the husband dies before the wife.
3. Find the expected time before the last person dies.
4. Write down the backward differential equations for the transition probabilities needed to find the distribution of $X(t)$.

Consider now the stochastic process obtained by collapsing the states where one person of the couple is alive, i.e. define $(\tilde{X}(t))_{t\geq 0}$ by

$$\tilde{X}(t) = \begin{cases} 0, & X(t) = \text{both alive} \\ 1, & X(t) \in \{\text{husband dead,wife dead}\} \\ 2, & X(t) = \text{both dead} \end{cases}$$

In general $(\tilde{X}(t))_{t\geq 0}$ is not a Markov chain and we shall try to argue why.

5. Compute

$$P(X(3t) = \text{both dead}, X(2t) = \text{wife dead}, X(t) = \text{wife dead})$$

and

$$P(X(3t) = \text{both dead}, X(2t) = \text{husband dead}, X(t) = \text{husband dead})$$

using the formula

$$\begin{aligned} & P(X(3t) = k, X(2t) = j, X(t) = i) \\ = \quad & P(X(t) = i) \cdot P(X(2t) = j | X(t) = i) \cdot P(X(3t) = k | X(2t) = j). \end{aligned}$$

6. Compute $P(\tilde{X}(3t) = 2, \tilde{X}(2t) = 1, \tilde{X}(t) = 1)$ by writing the event as a disjoint union of the sets in question 5.

7. Use a similar trick as in questions 5.-6. to compute

$$P(\tilde{X}(3t) = 2, \tilde{X}(2t) = 1, \tilde{X}(t) = 0).$$

8. Write the set $(\tilde{X}(2t) = 1, \tilde{X}(t) = 1)$ as a disjoint union of events involving $(X(t))$ and use question 6. to compute

$$P(\tilde{X}(3t) = 2 | \tilde{X}(2t) = 1, \tilde{X}(t) = 1).$$

9. Use question 7. and the ideas of question 8. to compute

$$P(\tilde{X}(3t) = 2 | \tilde{X}(2t) = 1, \tilde{X}(t) = 0).$$

10. Argue that in general $\{\tilde{X}(t)\}_{t \geq 0}$ is not a Markov chain on $\{0, 1, 2\}$.

11. Under what restriction of the model parameters does it hold that $\{\tilde{X}(t)\}_{t \geq 0}$ is a Markov chain.

3.3.3 *Forward differential equations for four state chain*

Consider the Markov chain on Figure 3.15 with state space $S = \{1, 2, 3, 4\}$, where the initial distribution is given by $P(X(0) = 1) = 1$.

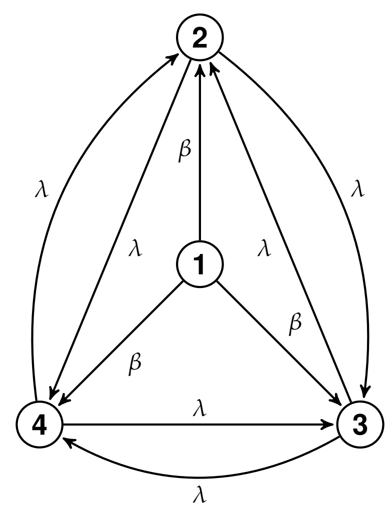


Figure 3.15: Transition diagram of four state Markov chain of Exercise 3.3.3.

1. Find the intensity matrix Q of the chain.

2. Write down the system of equations for the invariant probability, $\overline{\pi}$, of the chain and find $\overline{\pi}$.

3. Find the transition matrix P for the jumps of the chain.

4. Find the distribution of the first jump time and use this to find an expression for $P_{1,1}(t)$.

5. Write down the forward differential equation, $P'(t) = P(t)Q$, for the transition probability $P_{1,2}(t) = P(X(t) = 2 | X(0) = 1)$.

6. Solve the differential equation from question 5. by using the result from question 4. and that by symmetry we must have

$$P_{1,2}(t) = P_{1,3}(t) = P_{1,4}(t).$$

Try also to give an even simpler derivation of $P_{1,2}(t)$ refering only to symmetry but without using the differential equation.

7. Write down the forward differential equation for $P_{2,2}(t)$.

8. Using that $P_{2,1}(t) = 0$ and $\sum_{j \in S} P_{2,j}(t) = 1$ show that the equation from question 7. has a solution of the form

$$P_{2,2}(t) = c_1 + c_2 \exp(-3\lambda t)$$

and determine the constants c_1, c_2.

9. Find the remaining transition probabilities.

Hint: For an easy solution to this question start by listing transition probabilities that are zero and transition probabilities that must be the same due to symmetry. You will probably also find it useful to remember that the rows of $P(t)$ sum to one.

3.3.4 *Time to absorption*

Suppose that the Markov chain corresponding to Figure 3.16 starts in state 1, i.e. $P(X(0) = 1) = 1$.

1. Find the intensity matrix, Q, of the chain.

2. Write down the system of equations for the invariant probability $\overline{\pi}$ of the chain and find $\overline{\pi}$.

3. What is the distribution and the mean of the time to the first jump?

4. Find the transition matrix, P, for the jumps of the chain.

In the rest of the exercise we exclude the possibility that the chain can jump from state 0 to state 1. This situation corresponds to the transition diagram on Figure 3.17.

5. Write down the transition matrix for the jumps of the modified version of the chain. By convention for an absorbing state i let us put $P_{i,i} = 1$.

6. Considering only the Markov chain of jumps compute (using a computer) the expected number of times the chain will visit state 2 before absorption in state 0. Answer the same question for state 3.

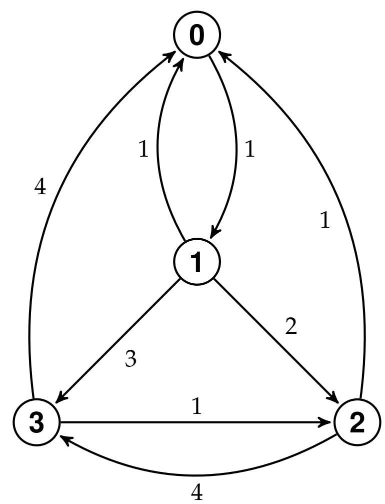


Figure 3.16: Transition diagram for the Markov chain used for question 1.-4. of Exercise 3.3.4.

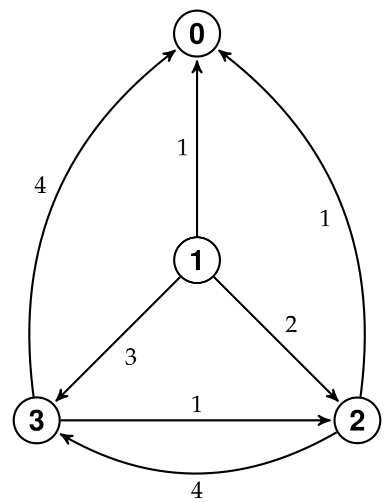


Figure 3.17: Transition diagram for the Markov chain used for question 5.-9. of Exercise 3.3.4.

7. Write down the matrix, Q, of transition intensities for the modified version of the continuous-time Markov chain.

8. Verify that $v_1 = (1,1,1,1)^T$ and $v_2 = (0,7/3,2,1)^T$ are (right) eigenvectors for Q with eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -3$.

9. Verify that $v_3 = (0,1,0,0)^T$ and $v_4 = (0,1,-2,1)^T$ are (right) eigenvectors for Q and find the corresponding eigenvalues λ_3 and λ_4.

10. Let O be the 4×4 matrix with columns given by v_i (i.e. $O = (v_1, v_2, v_3, v_4)$). Use the fact that the transition probabilities, $P(t) = (P_{i,j}(t))_{i,j\in S}$, are given by the exponential matrix

$$\exp(Qt) = O \begin{pmatrix} \exp(\lambda_1 t) & 0 & 0 & 0 \\ 0 & \exp(\lambda_2 t) & 0 & 0 \\ 0 & 0 & \exp(\lambda_3 t) & 0 \\ 0 & 0 & 0 & \exp(\lambda_4 t) \end{pmatrix} O^{-1}$$

to compute the probability $P_{1,0}(t)$.

Hint: You can use without proof that

$$O^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3/4 & 0 & 1/4 & 1/2 \\ 1 & 1 & -1/3 & -5/3 \\ -1/4 & 0 & -1/4 & 1/2 \end{pmatrix}.$$

11. Define the time of the first visit to state 0

$$T = \inf\{t > 0 | X(t) = 0\}$$

and argue that $P(T \leq t) = P(X(t) = 0)$.

12. Use (without proof) that the expectation of the nonnegative random variable T may be expressed as

$$E[T] = \int_0^\infty P(T > t)dt$$

to compute the expected time to absorption in state 0 when the chain is started at state 1 (i.e. $P(X(0) = 1) = 1$).

13. Use the following heuristic argument to compute the expected time to absorption in state 0 : First compute the expected number of the time periods where the chain visits states $1, 2$, and 3. Then multiply the expected number of visits in each state with the average waiting time before the chain jump to another state. This gives you the expected time spend in each state.

Hint: You already computed many of the necessary quantities in previous questions.

Countable state space

3.4.1 *Pure death process with constant intensity*

A birth-and-death process is a continuous time Markov chain on $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ that moves only in jumps of size one. The process may describe the size of a population and a jump between states i and $i+1$ is interpreted as a *birth* whereas jumps from i to $i-1$ corresponds to a *death*.

Denoting by $q_{i,j}$ the transition intensities from state i to state j the birth-and.death process has the following structure

$$q_{ij} = \begin{cases} \beta_i & , j = i+1, i \geq 0 \\ \delta_i & , j = i-1, i \geq 1 \\ 0 & , \text{otherwise} \end{cases}$$

for suitable nonnegative birth- and death intensities $\beta_i, \delta_i \geq 0$. Many important stochastic processes belong to the class of birth-and-death processes and may be obtained by imposing various restrictions on the birth- and death intensities.

Assume that the Markov chain $(X(t))_{t \geq 0}$ is a birth- and death process with initial distribution $\rho_i = P(X(0) = i)$.

1. Draw a part of the transition diagram for the Markov chain under the assumption that all $\beta_i, \delta_i > 0$.

2. Find under the assumption of question 1. the transition probabilities for the corresponding discrete time Markov chain of jumps for $(X(t))_{t \geq 0}$.

3. What choice of initial distribution and birth- and death intensities implies that $(X(t))_{t \geq 0}$ is a Poisson process?

The pure death process is characterized by all the birth intensities, β_i, being equal to zero. In the following we consider a pure death process with $\delta_k > 0, k \geq 1$, and initial distribution $P(X(0) = k) = 1$ for some $k \geq 2$.

4. What is the distribution of the first jump time
$$\tau_1 = \inf\{t > 0 | X(t) \neq k\}?$$

5. Find $P(X(t) = k)$.

6. Assuming that all death intensities are the same, $\delta_i = \delta > 0$, what is then the distribution of the time, τ_2, of the second jump of the chain?

7. Under the assumption of question 6. one may argue that $(X(t))_{t\geq 0}$ behaves like a modified Poisson process $(N(t))_{t\geq 0}$ with downward jumps of intensity $\delta > 0$ until the time of the $k-$th jump. Use this to compute $P(X(t) = j), j = 1, 2, \ldots, k-1$.

8. Find $P(X(t) = 0)$ under the assumption of question 6.

3.4.2 *Linear birth-and-death process*

The linear birth-and-death process is a continuous-time Markov process on $\mathbb{N}_0$ with birth intensities $\beta_i = i\beta$ and death intensities $\delta_i = i\delta$. It may be thought of as a model for a population where at any time an individual dies with intensity $\delta > 0$ and gives rise to a birth with intensity $\beta > 0$.

1. Find the communication classes of the linear birth-and-death process.

2. Assume that $P(X(0) = 1) = 1$ and let

$$\tau_1 = \inf\{t > 0 | X(t) \neq 1\}$$

   be the time of the first jump. Find the probability $P(\tau_1 > 1)$ and the distribution, $P(X(\tau_1) = i), i \geq 0$, of the chain observed just after the first jump.

3. Let $T = \inf\{t > \tau_1 | X(t) = 1)$ be the time of the first return to state 1. Use the result of question 2. to get an upper bound for the probability $P(T < +\infty | X(0) = 1)$. Discuss what you can conclude from this observation.

4. Still assuming that $P(X(0) = 1) = 1$ argue that

$$P(X(1) = 0) > \tfrac{\delta}{\delta+\beta}(1 - \exp(-(\delta + \beta))).$$

For the rest of the exercise we modify the birth intensities such that $\beta_i = i\beta + \lambda$ for some $\beta, \lambda > 0$. The resulting model has a very nice interpretation as a linear birth-and-death process with immigration intensity λ.

5. Argue that the linear birth-and-death process with immigration is irreducible.

6. Let $(Y(n))_{n\geq 0}$ be the discrete-time Markov chain of jumps. Find the transition probabilities of $(Y(n))_{n\geq 0}$.

The purpose of the following questions is to clarify for what values of the parameters that a linear birth-and-death process is transient, null-recurrent and positive recurrent. At the written exam you should directly apply the results in Chapter 3 from page 90 to answer questions 8., 11. and 12. below. Questions 7., 9. and 10. are only relevant for those of you who want to understand better how to arrive at the main results in the Section on Birth-and-death processes in Chapter 3.

7. Write down the system of equations in Theorem 15 from page 28 in Chapter 2 where you use $i = 0$ as *fixed state*.

It is rather technical to write down a complete solution to questions 8. and 11.-12. below covering all choices of the parameters β, δ, λ. To make things a bit more easy try to consider first the cases $\beta > \delta$ and $\beta < \delta$.

8. Use the system of equations from question 7. (or some other argument) to determine for what choices of $\beta, \delta, \lambda > 0$ that the linear birth-and-death process with immigration is recurrent or transient. The case $\beta = \delta < \lambda$ is particularly hard and may be skipped.

   Hint: Use that the solution to the system of equations from question 7. has the form

$$\alpha(j+1) = \alpha(1)\left\{1 + \sum_{k=1}^{j} \frac{k\delta \cdot \ldots \cdot \delta_1}{(\lambda + k\beta) \cdot \ldots \cdot 1\beta}\right\}, \quad j \geq 1.$$

   A simpler approach is just to apply a suitable result in Chapter 3.

9. Show that an invariant probability vector $\pi = (\pi_i)_{i \in \mathbb{N}_0}$ for the linear birth-and-death process with immigration must satisfy the following system of equations

$$\begin{aligned} 0 &= \delta\pi_1 - \lambda\pi_0 \\ 0 &= ((i-1)\beta + \lambda)\pi_{i-1} + (i+1)\delta\pi_{i+1} - (i\beta + i\delta + \lambda)\pi_i, i \geq 1. \end{aligned}$$

10. Verify that the vector $\nu = (\nu_i)_{i \in \mathbb{N}_0}$ where

$$\nu_i = \nu_0 \cdot \prod_{k=1}^{i} \frac{(k-1)\beta + \lambda}{k\delta}, i \geq 1,$$

   solves the system of equations from question 9.

11. Determine for what values of $\beta, \delta, \lambda > 0$ that the solution , ν, of question 10. can be normalized into a probability vector π.

12. For what choice of the parameters β, δ, λ is the birth-and-death process with immigration null-recurrent?

   Hint: If we already know that the chain is recurrent then the chain is positive recurrent if and only if there exists a probability vector solving the system of equations from question 9.

Answer the following two questions 13.-14. for the three cases $\lambda = \beta = 2\delta$, $\lambda = \beta = \delta$, and $\lambda = \beta = \delta/2$.

13. Find $\lim_{t\to\infty} P(X(t) = i)$ for $i \geq 0$ under the assumption that $P(X(0) = 1) = 1$.

3.4.3 *Queueing systems*

There is an entire branch of applied probability that deals with mathematical modeling of queueing systems. In this exercise we show by an example how continuous-time Markov chains may be used to model the number of customers in a queueing system. Throughout the exercise we assume that new customers arrive to the system according to a Poisson process with intensity $\beta > 0$ independently of the state of the system.

We consider initially the single server queue where customers are served according to the first-come-first-served queueing discipline. Upon arrival at the service desk the service time distribution is assumed to be exponential with rate $\delta > 0$ no matter how many customers are waiting in line. One can show (but you are not supposed to do so!) that under the given assumptions then the number, $(X(t))_{t\geq 0}$, of customers present in the system constitutes a continuous-time Markov chain on $\mathbb{N}_0$ with transition intensities

$$q_{i,j} = \begin{cases} \beta \quad , \quad j = i+1, i \geq 0 \\ \delta \quad , \quad j = i-1, i \geq 1 \\ 0 \quad , \quad \text{otherwise} \end{cases}$$

1. Argue that the chain is a birth-and-death process.

2. Write down the system of equations that must be satisfied for an invariant probability vector $\pi = (\pi_i)_{i\in\mathbb{N}_0}$. Find the invariant distribution, π, of the chain for the case where $\beta < \delta$.

3. Assuming that $\beta < \delta$ compute the (long run) average number of customers in the queue.

4. What is the distribution of the waiting time before arrival to the service desk if 4 customers are waiting in front of you when you arrive to the queueing system? (You are not expected to do any computations here!)

We now assume that the customers are served in their order of arrival by two servers with exponentially distributed service time distributions of (possibly different) rates $\delta_1, \delta_2 > 0$. If a customer arrives at an empty system she or he is by default served at service desk number 1. With some effort one can show that the system may be regarded as a continuous-time Markov chain on the state space

$$S = \{0, 1:0, 0:1, 2, 3, 4, \ldots\}.$$

This needs a little more explanation: state 0 means that no customers are present, state $0 : 1$ means that one customer is being served at service desk 2 while service desk 1 is vacant. Similarly, state $1 : 0$ represents the situation where service desk 1 is occupied and desk 2 is vacant. States $2, 3, 4, \ldots$ refer to situations where at least two customers are present of which two are currently being served at service desks 1 and 2.

The transition diagram (without transition intensities) is given on Figure 3.18

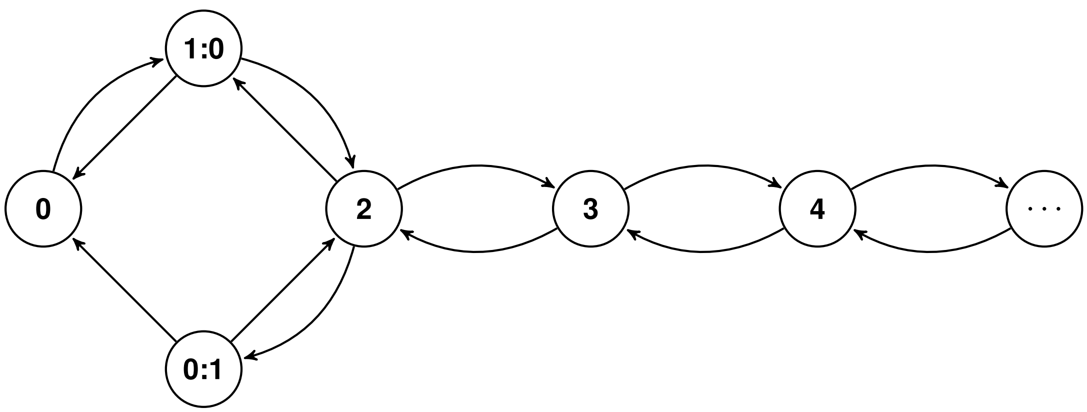


Figure 3.18: Transition diagram for the Markov chain considered in questions 5.-14. of Exercise 3.4.3.

5. What is the intensity $q_{0:1,0}$ of a jump from state $0 : 1$ to 0?

6. What is the intensity $q_{1:0,0}$ of a jump from state $1 : 0$ to 0?

7. Find the intensities $q_{0:1,2}, q_{1:0,2}, q_{0,0:1}$.

8. Argue that the intensity of a jump from state 3 to 2 equals $\delta_1 + \delta_2$.

9. Draw the transition diagram of the Markov chain with all intensities.

10. Argue very carefully that an invariant probability vector $\pi = (\pi_i)_{i \in S}$ must satisfy the system of equations

$$\begin{aligned}
0 &= \delta_1 \pi_{1:0} + \delta_2 \pi_{0:1} - \beta \pi_0 \\
0 &= \beta \pi_0 + \delta_2 \pi_2 - (\beta + \delta_1) \pi_{1:0} \\
0 &= \delta_1 \pi_2 - (\beta + \delta_2) \pi_{0:1} \\
0 &= \beta \pi_{1:0} + \beta \pi_{0:1} + (\delta_1 + \delta_2) \pi_3 - (\delta_1 + \delta_2 + \beta) \pi_2 \\
0 &= (\delta_1 + \delta_2) \pi_{i+1} + \beta \pi_{i-1} - (\beta + \delta_1 + \delta_2) \pi_i, \quad i \geq 3.
\end{aligned}$$

11. Show that for any constant c there is a vector, $\pi = (\pi_i)_{i \in S}$, with

$$\pi_i = c\left(\frac{\beta}{\delta_1+\delta_2}\right)^{i-2}, \quad i \geq 2$$

that solves the system of equations from 10 and derive expressions for $\pi_{0:1}$, $\pi_{1:0}$, and π_0.

12. For what values of the parameters $\beta, \delta_1, \delta_2$ can the solution in question 11. be normalized into an invariant probability vector? (You don't need to find a closed form expression for c to answer this question!)

13. Consider the case where $\delta_1 = \delta_2 = \delta$ and $\beta = \delta/2$. Find the invariant probability vector from questions 11.-12. and compute the (long run) average number of customers for the two-server queue.

    Hint: You can use (or verify) that $c = 3/40$.

14. Still assuming that $\delta_1 = \delta_2 = \delta$ and $\beta = \delta/2$ discuss how much the (long run) average queue length decreased by the introduction of the second server compared to the single server system (question 1.-4.).

We finally consider the case where arriving customers physically lines up in two different queues. Upon arrival a customer enters the shortest of the two lines. If there are the same number of customers in each queue any customer by default enters the queue nearest to the entrance of the building (let us call this queue number 1). If at any time the difference between the length of two queues is two the last customer in the longest queue will instantly switch to the last position in the shorter queue. The purpose of the following questions is to study the differences between the two-line queueing disciplin and the one line first-come-first-served disciplin considered in questions 5.-14.

It is possible to show that the joint number , $((X_1(t), X_2(t)))_{t \geq 0}$, of customers in the two queues is a continuous-time Markov chain on $\mathbb{N}_0 \times \mathbb{N}_0$.

15. Technically speaking the state space of the chain is much smaller than $\mathbb{N}_0 \times \mathbb{N}_0$ because a large number of the states will never be visited by the chain. What is the trimmed version, S, of the state space that represents the truly possible states of the queueing system?

16. Draw the transition diagram (with transition intensities) of the Markov chain that displays only the trimmed state space, S, from

question 15. You probably need to be careful to get all the transition intensities right in particular for jumps between states $(i, i+1)$ and (i, i) or between states $(i+1, i)$ and (i, i).

17. Argue that an invariant probability vector $\pi = (\pi_{i,j})_{(i,j) \in S}$ must satisfy the following system of equations

$$
\begin{aligned}
0 &= \delta_2 \pi_{0,1} + \delta_1 \pi_{1,0} - \beta \pi_{0,0} \\
0 &= \beta \pi_{0,0} + \delta_2 \pi_{1,1} - (\delta_1 + \beta) \pi_{1,0} \\
0 &= \delta_1 \pi_{1,1} - (\delta_2 + \beta) \pi_{0,1} \\
0 &= \beta(\pi_{i-1,i} + \pi_{i,i-1}) + (\delta_1 + \delta_2)(\pi_{i+1,i} + \pi_{i,i+1}) - (\beta + \delta_1 + \delta_2)\pi_{i,i}, \quad i \geq 1, \\
0 &= \beta \pi_{i,i} + \delta_2 \pi_{i+1,i+1} - (\beta + \delta_1 + \delta_2)\pi_{i+1,i}, \quad i \geq 1, \\
0 &= \delta_1 \pi_{i+1,i+1} - (\beta + \delta_1 + \delta_2)\pi_{i,i+1}, \quad i \geq 1.
\end{aligned}
$$

18. Verify that for any constant c there is a vector, $\pi = \{\pi_i\}_{i \in S}$, with

$$
\pi_{i,i} = c \left(\frac{\beta^2}{(\delta_1 + \delta_2)^2} \right)^i, \quad i \geq 1,
$$

that solves the system of equations from question 17 and derive expressions for the remaining coordinates of π.

Hint: Start by plugging in to the last equation of question 17. to get an expression for $\pi_{i,i+1}$ and do not try to find the constant c.

19. For what values of the parameters $\beta, \delta_1, \delta_2$ can the solution of 18. be normalized into an invariant probability vector? (You don't need to find a closed for expression for c to answer this question!)

20. Consider the case where $\delta_1 = \delta_2 = \delta$ and $\beta = \delta/2$. Find the invariant probability vector from questions 17.-18. and compute the (long run) average number of customers present in the queueing system.

21. Are there any reason to prefer one of the two suggested two-server queueing disciplines to the other from the customers point-of-view? To answer the question you may find it useful to include a discussion of your results from questions 13. and 20.

22. The total service capacity (per time unit) of a queueing system with two servers is given by the sum $\delta_1 + \delta_2$. Which of the queueing systems with two servers exploit the service capacity in the must efficient way? (Don't do any computations!)

23. Try to do some numerical computations to examine if there are any differences betweeen the two suggested two-server queueing systems when $\delta_1 \neq \delta_2$. Look at the problemer from the customers point-of-view.

Comments: The results of this exercise do not carry over to real life queueing systems for several reasons of which we shall mention a few: the unrealistic assumption of exponentially (=memoryless) distributed service times and intervals between arriving customers, the assumption of customers arriving at the same rate at all times, and the independence of the service distributions on both time and on the number of customers already present in the queueing system.

It is trival that queues build up if the (average) service capacity is lower than the average rate of arriving customers. Another important lecture you may learn by digging further into the field of queueing theory is that even for a sufficient average service capacity queues are caused by variation in interarrival times and service times. The generel message is that inducing more variation deteriorates the performance of a queueing system.

3.4.4 *Positive recurrence and null-recurrence*

We consider a Markov chain, $(X(t))_{t\geq 0}$, on $S = \{0, 1, 2, \ldots\}$ with transition intensities

$$q_{0,n} = p_n > 0, q_{n,n-1} = \delta_n > 0, n > 0, \quad q_{i,j} = 0 \text{ for any other } i \neq j$$

where $\sum_n p_n = 1$.

1. Find the transition probabilities for the embedded Markov chain of jumps.

2. Argue that $(X(t))_{t\geq 0}$ is recurrent.

So far we have demonstrated that the Markov chain $(X(t))_{t\geq 0}$ and the embedded Markov chain of jumps are always recurrent no matter the values of $p_n > 0$ and $\delta_n > 0$. The purpose of the following is to show that all four combinations of positive recurrence and null-recurrence for $(X(t))_{t\geq 0}$ and the embedded jump chain may occur.

NN $(X(t))_{t\geq 0}$ and the embedded jump chain are null-recurrent.

NP $(X(t))_{t\geq 0}$ is null-recurrent and the embedded jump chain is postive recurrent.

PN $(X(t))_{t\geq 0}$ is positive recurrent and the embedded jump chain is null-recurrent.

PP $(X(t))_{t\geq 0}$ and the embedded jump chain are positive recurrent.

Find out how the four cases listed above correspond to the four sets of parameters described in questions 3.-6. below.

3. $p_n = (1-p)p^{n-1}, 0 < p < 1$, and $\delta_n = \delta > 0$

4. $p_n = c/n^2$ and $\delta_n = \delta > 0$

5. $p_n = (1-p)p^{n-1}$ and $\delta_n = (1-p/2)^{-1}(p/2)^n$ where $0 < p < 1$

6. $p_n = c/n^2$ and $\delta_n = n(n+1)$

Hint: We did already study the embedded Markov chain of jumps in Exercise 2.4.6.

3.4.5 More examples of birth-and-death processes

We consider in this exercise four different birth-and-death processes. The purpose of this exercise is to get some experience using the results stated in Chapter 3 on birth- and death processes.

1. Show that the birth-and-death process with intensities

$$q_{i,i+1} = \beta_i = i+1, \quad q_{i+1,i} = \delta_{i+1} = 1, \quad i \geq 0$$

is transient.

2. Show that the birth-and-death process with intensities

$$q_{i,i+1} = \beta_i = i+1, \quad q_{i+1,i} = \delta_{i+1} = i+1, \quad i \geq 0$$

is null-recurrent.

3. Show that the birth-and-death process with intensities

$$q_{i,i+1} = \beta_i = i+1, \quad q_{i+1,i} = \delta_{i+1} = i+3, \quad i \geq 0$$

is positive recurrent.

4. Show that for $q < p < cq$ then the birth-and-death process with intensities

$$q_{i,i+1} = \beta_i = c^i p, \quad q_{i+1,i} = \delta_{i+1} = c^{i+1} q, \quad i \geq 0$$

is transient *and* there exists a probability vector $\overline{\pi} = (\pi(i))_{i \in S}$ solving the system

$$\sum_{i \in S} \pi(i) q_{i,j} = 0, \quad j \in S.$$

Hint: Use Appendix B.5 on linear recurrence equations.

Remark: Question 4. shows that there exists a probability vector satisfying the necessary condition of Theorem 51 for an invariant distribution. However, since the Markov chain is transient the invariant distribution does not exist. One can show using Reuter's criterion from Theorem 64 that explosion may occur for the birth-and-death process given in Question 4.

The Poisson process

3.5.1 *Basic properties of the Poisson process*

The *Poisson process* with intensity λ is the continuous-time and time-homogeneous Markov chain on $\mathbb{N}_0$, given by the transition diagram in Figure 3.19.

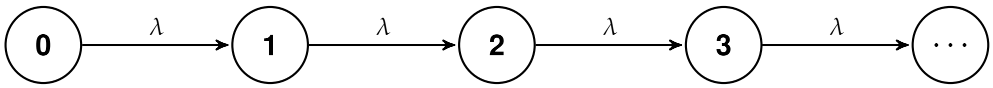


Figure 3.19: Transition diagram for the homogeneous Poisson process with intensity λ.

In particular the times between jumps are independent and exponentially distributed with density function

$$f(s) = \lambda \exp(-\lambda s), \quad s \geq 0.$$

Further, we denote by $\tau_1, \tau_2, \ldots$ the jump times of $(X(t))$ and we assume that $X(0) = 0$.

1. Compute $E[\tau_1], P(X(t) = 0)$, and $P(X(t) \geq 1)$.
2. For $0 < s < t$ compute $P(X(s) = 0, X(t) = 0)$

For non-negative independent random variables V with density g and W with density h then the density of the sum $Y = V + W$ has density given by

$$k(y) := h * g(y) := \int_0^y h(y - v) g(v) dv, \quad y \geq 0.$$

3. Find the distribution (=density) of τ_2 by using that τ_2 is the sum of two independent exponential distributions with rate parameter λ.
4. Compute $P(X(t) \geq 2)$.
5. Verify by induction that the time, τ_n, of the jump to state n follows a distribution with density

$$f_n(s) = \frac{\lambda(\lambda s)^{n-1}}{(n-1)!} \exp(-\lambda s), \quad s \geq 0.$$

6. Compute $P(X(t) \geq n)$ and $P(X(t) = n)$.

   Hint: Use without proof that

$$P(\tau_n \leq t) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$$

   or you may even try to prove the formula by induction.

7. What is the name of the distribution of τ_n and $X(t)$?

3.5.2 *Advanced exercise involving the Poisson process*

In this exercise we consider a Poisson process $(X(t))$ with intensity λ. You may use that from Exercise 3.5.1 we know the distribution of $X(t)$ i.e. the probabilities $P(X(t) = n), n \in \mathbb{N}$.

The purpose of this exercise is to study further the times of the jumps of the Poisson process which we will denote by $\tau_1, \tau_2, \tau_3, \ldots$. From the previous Exercise 3.5.1 we know the distribution of τ_n. In this exercise we consider what can be said about the distribution (=location) of the n first jump times given that we know that $X(1) = n$ i.e. that exactly n jumps occurred on the time interval $[0, 1]$.

For simplicity we consider only the distribution of τ_1 by asking the following question: given that we know that exactly one jump happened before time 1 (i.e. $X(1) = 1$) when was the most likely time on $[0, 1]$ for the jump, τ_1, to happen? Clearly, the conditional distribution of the first jump time, τ_1, given that $X(1) = 1$ is a distribution on the interval $[0, 1]$. The purpose of the following questions 1.-8. is to compute $P(a < \tau_1 \leq b | X(1) = 1)$ for $0 \leq a \leq b \leq 1$.

1. Try to argue, for instance on a suitable figure, that

$$(\tau_1 \leq b, X(1) = 1) = (X(b) = 1, X(1) = 1), \quad \text{for } 0 \leq b \leq 1.$$

2. Find the probability that $P(X(a) = 1)$, $a \geq 0$.

3. Explain how it follows from the Markov property (and the stationarity) of the Poisson process that for $s, t \geq 0$ and $i, j \in \mathbb{N}_0$ then

$$P(X(t+s) = i + j | X(s) = i) = \frac{(\lambda t)^j}{j!} \exp(-\lambda t).$$

4. Find the probability that $P(X(b) = 1, X(1) = 1)$, for $0 \leq b \leq 1$.

5. Compute $P(\tau_1 \leq b | X(1) = 1)$ using questions 1.-4.

6. Argue that for $0 \leq a \leq b \leq 1$ then

$$(a < \tau_1 \leq b, X(1) = 1) = (X(a) = 0, X(b) = 1, X(1) = 1).$$

7. Write $P(a < \tau_1 \leq b, X(1) = 1)$ as a product of three probabilities that are known from questions 1.-6. above.

8. Compute the conditional probability $P(a < \tau_1 \leq b | X(1) = 1)$.

Remark: The result shows that if exactly one jump of a Poisson process occurs on the interval $[0, 1]$ *then the (conditional) distribution of the jump follows a uniform distribution on* $[0, 1]$. *The result generalises to the case where we consider the conditional distribution of the n first jumps given* *that exactly n jumps occurred on the interval $[0, t]$. The location of the n jumps will behave as if they had been uniformly scattered over the interval $[0, t]$ independently of each other. For that reason the event times of a Poisson process is often said to describe a completely random pattern of points.*

A

Random variables and stochastic processes

Probability measures

It all begins with a **probability measure** P. You should think of a probability measure, P, on a set Ω as a function assigning a number $P(A) \in [0,1]$ to subsets $A \subset \Omega$. If you are familiar with measure theory you may correctly insist that a probability measure only assigns a probability to subsets $A \in \mathcal{F}$ in a σ-algebra $\mathcal{F}$ on Ω [1]but this point of view is not crucial for the story to come. Subsets of Ω are referred to as **events** .

[1] A σ-algebra on Ω is a class $\mathcal{F}$ of subsets of Ω such that

1. $\Omega \in \mathcal{F}$
2. $A^C \in \mathcal{F}$ if $A \in \mathcal{F}$
3. $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$ if $A_1, A_2, \ldots \in \mathcal{F}$

By definition a probability measure must have total mass equal to one (i.e. $P(\Omega) = 1$) and it must be additive over countable classes of disjoint sets, i.e.

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n),$$

provided that $A_i \cap A_j = \emptyset, i \neq j$.

It follows from the additivity above that a probability P on a countable set Ω is determined by the values $P(\{\omega\})$ from the identity

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

In this case we will refer to $P(\{\omega\})$ as the point probability at ω.

For two events A, B with $P(B) > 0$ we define the elementary **conditional probability** of A given B (notation: $P(A|B)$) as

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

Example 66 (Poisson distribution) *The* ***Poisson distribution*** *with parameter* $\lambda > 0$ *is a probability on (the countable set!)* $\mathbb{N}_0$ *given by point probabilities*

$$P(\{n\}) = \frac{\lambda^n}{n!} \exp(-\lambda), \quad n \in \mathbb{N}_0.$$

□

Example 67 (Exponential distribution) *The **exponential distribution** with rate parameter $\lambda > 0$ is a probability on $[0, \infty)$. The exponential distribution assigns probability*

$$P(A) = \int_A \lambda \exp(-\lambda t) dt$$

to (Borel measurable) subsets $A \subset [0, \infty)$. For an interval $[a, b], 0 < a < b$, we have that

$$P([a, b]) = \int_a^b \lambda \exp(-\lambda t) dt = [-\exp(-\lambda t)]_a^b = \exp(-\lambda a) - \exp(-\lambda b).$$

□

More examples of probability distributions are given in Chapter B.1.

Random variables

When we refer to a random experiment we want to emphasize that we are in a situation where we are unable to predict the outcome with certainty. There might be several reasons that we do not know the exact result of an experiment: the outcome may be affected by circumstances that we are unable to control or we may simply not have complete information allowing us to determine the result of the experiment.

The concept of a **random variable** or **stochastic variable** is used for a mathematical model of a random experiment. Formally, a random variable, X, is a function and we reflect the randomness by saying that the argument $\omega \in \Omega$ of the function is chosen according to some probability distribution, P. The outcome of the experiment is denoted by $X(\omega)$. Two different ω's will potentially give different results of the experiment reflecting the non-deterministic nature of the experiment.

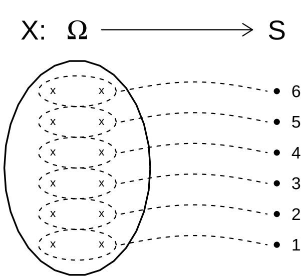


Figure A.1: Random variable with $P(X = j) = 1/6$ for $j = 1, 2, 3, 4, 5, 6$ from Example 68.

Example 68 *As a mathematical model for a random experiment we let Ω consist of 12 elements. We use a uniform distribution assigning a probability*

$$P(\{w\}) = \frac{1}{12}$$

to all $\omega \in \Omega$ and let the random variable $X : \Omega \to S = \{1, 2, 3, 4, 5, 6\}$ be given as illustrated in Figure A.

We use the notation $P(X = 6)$ for the probability that the random variable takes the value 6. We compute $P(X = 6)$ by summing $P(\{w\})$ for all ω such that $X(\omega) = 6$. Since there are exactly two ω with $X(\omega) = 6$ and since P assigns probability $\frac{1}{12}$ to all ω we conclude that $P(X = 6) = \frac{1}{6}$. It is easily checked that

$$P(X = j) = \frac{1}{6}$$

for all $j = 1, 2, 3, 4, 5, 6$*. In particular, under the suggested probability measure* P *on* Ω *the random variable* X *provides a reasonable mathematical description of the number of eyes when rolling a six-sided die.*

□

It is natural if you are a little puzzled about our choice of random variable from Example 68. Why didn't we use a one-to-one random variable X on a set Ω with only six elements and a uniform distribution, P, assigning equal probability $\frac{1}{6}$ to each ω in Ω as our model for the number of eyes of the die. An important point is that we certainly could have used this (obviously!) simpler model. However, when we are only interested in the distribution of the random variable X then the two models are equal since we have $P(X = j) = \frac{1}{6}$ in both cases. Another way to put this is that two different random variables ($X_i : \Omega_i \to S, i = 1, 2$) can be very different and still describe the same random experiment. We say that the two random variables are different representations of the same experiment.

For students familiar with measure theory a random variable may be defined (more formally) as a measurable map

$$X : (\Omega, \mathcal{F}) \to (S, \mathcal{G})$$

where $\mathcal{F}, \mathcal{G}$ are classes of subset satisfying the conditions of a σ-algebra. For a subset $A \subset S$ (-with $A \in \mathcal{G}$) the probablity that the random experiment gives a value in the set A is computed as

$$P(X \in A) \stackrel{def}{=} P(\{w \in \Omega | X(\omega) \in A\}).$$

In fact, this defines a probability on S which we will call the distribution of random variable X. Any computation that only involves the distribution of a random variable X will give exactly the same result if the computation is based on any other random variable $\tilde{X}$ with the same distribution as X. For that reason it is common to formulate results about distributions of random variables without reference to the exact representation of the random variable. Anyone may answer the following question without caring about the formal definition of the random variable X occuring in the text.

Let X be a random variable with distribution $P(X = j) = \frac{1}{6}, j = 1, \ldots, 6$. Compute the probability the X takes a value in the set $A = \{2, 4, 6\}$ of even numbers.

Remark 69 *[Canonical representation of a random variable] Suppose that we want to make a mathematical model of a random experiment with a prespecified probability distribution* Q *on some set* S*. Can we then always* *define a random variable that has distribution Q? The answer is* yes *and the solution is ridiculously simple: let $\Omega = S$ be equipped with the probability $P = Q$ and let $X : \Omega(= S) \to S$ be the identity mapping $X(\omega) = \omega$.* □

Stochastic processes

We may consider several random variables

$$X(i) : \Omega \to S, \quad i = 1, \ldots, N,$$

that are all defined on the same set Ω. If Ω is equipped with a probability P as described above, then we have a mathematical model for the outcome of a random experiment resulting in N joint values from the set S. We will speak of family of random variables, $\{X(i)\}_{i \in I}$, with index set $I = \{1, 2, \ldots, N\}$.

A stochastic process in discrete time with state space S is a family $\{X(n)\}_{n \in \mathbb{N}_0}$ of random variables $X(n) : \Omega \to S$ with index set $I = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

A stochastic process in continuous-time with state space S is a family $\{X(t)\}_{t \in [0,\infty)}$ of random variables $X(t) : \Omega \to S$ with index set $I = [0, \infty)$.

Note that there is no reason to restrict the state space S to be the same for all stochastic variables in the family.

Example 70 *Lad P be the probability on the set*

$$\Omega = \{(\omega_1, \omega_2) | \omega_1, \omega_2 = 1, 2, \ldots, 6\}$$

assigning probability $\frac{1}{36}$ to all elements $\omega \in \Omega$. Define a family of two random variables by

$$\begin{aligned} X(1)(\omega) &= X(1)(\omega_1, \omega_2) = \omega_1 \\ X(2)(\omega) &= X(2)(\omega_1, \omega_2) = \omega_1 + \omega_2. \end{aligned}$$

Note that X_1 may be used as a model for the number of eyes of a die, while X_2 is a model for the sum of the number of eyes when rolling two dice.

The example may easily be generalized by letting P be the uniform distribution with point probabilities $\frac{1}{6^k}$ on any point $\{1, \ldots, 6\}^k$. We may then consider the family of random variables

$$X(j)(\omega) = X(j)(\omega_1 + \ldots + \omega_k) = \omega_1 + \ldots + \omega_j, \quad j =, \ldots, k,$$

representing the sum of the first j dice when rolling up to k six-sided dice.

We may compute the probability $P(X(1) = 3, X(2) = 7)$ by summing the point probabilities $P(\{\omega\})$ of all ω such that $X(1)(\omega) = 3$ and $X(2)(\omega) = 7$. Only one element ($\omega = (3, 4)$) is contained in the set

$$(X(1) = 3, X(2) = 7) = \{\omega \in \Omega | X(1)(\omega) = 3, X(2)(\omega) = 7\}$$

in particular we have that

$$P(X(1) = 3, X(2) = 7) = P((3,4)) = \tfrac{1}{36}.$$

For a more challenging exercise try to use a similar technique to compute the probability that you must roll a six-sided die three times before the sum of the eyes equals or exceeds 6. □

Example 71 *Example 70 involved a finite family of random variables. In this example we consider an infinite family* $\{X(n)\}_{n\in\mathbb{N}_0}$ *of random variables on a set* S *indexed by* $\mathbb{N}_0$. *This is what we have previously referred to as a discrete-time stochastic process with state space* S. *It is not trivial what set* Ω *and what probability measure* P *on* Ω *that we should use and how to define the random variables* $X(n) : \Omega \to S$ *in order to define stochastic process with some interesting properties. Therefore we may use a more indirect approach.*

Often we are more interested in the probability distribution of the stochastic process which is a probability measure Q *defined on the infinite product set*

$$S^{\mathbb{N}_0} = \{(s_0, s_1, \ldots) | s_n \in S, n \in \mathbb{N}_0\}.$$

If we can come up with a probability measure Q *on* $S^{\mathbb{N}_0}$ *then we may use the canonical representation from Remark 69. Formally, we define a discrete-time stochastic process on* $\Omega = S^{\mathbb{N}_0}$ *equipped with the probability distribution* $P = Q$ *by letting*

$$X(n)(\{s_i\}_i) = s_n.$$

Since $\{X(n)\}_{n\in\mathbb{N}_0}$ *is just the identity mapping on* Ω *then we get that*

$$P(X(0) \in A_0, X(1) \in A_1, \ldots) = Q(A_0 \times A_1 \times \ldots)$$

for $A_0, A_1, \ldots$ *being subsets of* S.

The example shows that existence of a discrete-time stochastic process on S *with some prespecified properties of the distribution boils down to verifying that the description characterizes a probability measure on the infinite product space* $S^{\mathbb{N}_0}$. □

Example 72 *[Product measures] Consider a family of probability measures* μ_i *on* S_i *indexed by* $i \in I$. *For any finite subset* $I_0 \subset I$ *there is a unique probability measure* $\otimes_{i\in I_0}\mu_i$ *on the product space* $\times_{i\in I_0} S_i$ *defined by*

$$\otimes_{i\in I_0}\mu_i(\times_{i\in I_0} A_i) = \prod_{i\in I_0} \mu_i(A_i),$$

where $A_i \subset S_i$. *If the sets* S_i *are sufficiently well-behaved then it follows by the* ***Kolmogorov extension theorem*** *that there is a unique probability measure on the* infinite *product set* $S^I = \{\{s_i\}_{i\in I} | s_i \in S_i, i \in I\}$. *In* *particular, using the canonical representation of stochastic variables there exists a stochastic process indexed by I such that*

$$P(X(i_1) \in A_{i_1}, \ldots, X(i_n) \in A_{i_n}) = \mu_{i_1}(A_{i_1}) \cdot \ldots \cdot \mu_{i_n}(A_{i_n})$$

for any finite subset $i_1, \ldots, i_n \in I$ *and* $A_{i_j} \subset S_{i_j}$*. We say that the family* $(X_i)_{i \in I}$ *of random variables are independent and that the marginal distributions of* $X(i)$ *are given by* μ_i*.* □

Example 73 *[Poisson process] The exponential distribution with rate parameter* $\lambda > 0$ *is a probability,* μ*, on* $[0, \infty)$ *given by the density*

$$f(x) = 1_{[0,\infty)} \lambda \exp(-\lambda x).$$

We may use Example 72 to construct a family $\{W(n)\}_{n \in \mathbb{N}}$ *of independent identically distributed random variables all following an exponential distribution with rate parameter* λ*.*

The family $\{W(n)\}_{n \in \mathbb{N}}$ *may be used to define other stochastic processes via transformation. In the following we explain in detail the construction of the* Poisson process*. We think of* $W(n)$ *as random waiting times between the occurence a given type of event. The n-th event thus happens at time*

$$T(n) = W(1) + \ldots + W(n).$$

The Poisson process is the continuous-time stochastic process $\{N(t)\}_{t \geq 0}$ *where* $N(t)$ *keeps track of the number of events that has happened up to (and including) time t. Formally, we may write*

$$N(t) = \sum_{n=1}^{\infty} 1(T(n) \leq t).$$

As a consequence of the representation of $\{N(t)\}_{t \geq 0}$ *in terms of the waiting times* $\{W(n)\}_{n \in \mathbb{N}}$ *we conclude that* $N_0 = 0$ *and that the sample paths defined as*

$$t \to N(t)(\omega)$$

are piecewise constant and moves upwards in jump of size one (-see Figure B.7 on page 145). □

The following is rather technical but absolutely essential if you want to get a deeper understanding of how to work with stochastic processes. The construction in Example 73 made use of a family of independent exponentially distributed waiting times to construct a family of random variables $\{N(t)\}_{t \geq 0}$ that we refer to as the Poisson process. The Poisson process takes values in $\mathbb{N}_0$ and hence has a distribution Q on the (countably) infinite product set $\mathbb{N}_0^{[0,\infty)}$. If we use the canonical representation from Remark 69 we get another representation of a stochastic process $(\tilde{N}(t))_{t \geq 0}$ with the same distribution as the Poisson process. Any computation related to the distribution of a Poisson process is redundant to the choice of representation. Note however that by the construction in Example 73 we obtained that the sample paths

$$t \to N_t(\omega)$$

are piecewise constant. There is nothing to ensure that the canonical representation

$$t \to \tilde{N}_t(\omega)$$

has similar nice properties. In general, sample path properties are not determined solely by the distribution of a stochastic process. In many situations its is convenient to work with specific representations of stochastic processes. This may be required for a given mathematical proof strategy to work even though the final result we are establishing may be formulated entirely in terms of the distribution of the stochastic process.

B
Mathematical tools

This chapter contains some mathematical results that might be useful to solve the exercises on Markov chains from Chapters 2 and 3.

B.1 Elementary conditional probabilities

For two events (=sets) A, B with $P(B) > 0$ the conditional probability of $A|B$ is defined by the formula

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

When working with Markov chains the events will often be expressed by random variables for example as $A = (X(2) = j)$ and $B = (X(1) = i)$. One may show that for three sets A, B, C with $P(B \cap C) > 0$ then

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C).$$

For Markov chains with three sets given as $A = (X(2) = k)$, $B = (X(1) = j)$, and $C = P(X(0) = i)$ this may be written out as

$$\begin{aligned} & P(X(2) = k, X(1) = j, X(0) = i) \\ = \; & P(X(2) = k|X(1) = j, X(0) = i) \cdot P(X(1) = j|X(0) = i) \cdot P(X(0) = i) \\ = \; & P(X(2) = k|X(1) = j)P(X(1) = j|X(0) = i)P(X(0) = i) \\ = \; & P_{j,k}P_{i,j}\phi(i), \end{aligned}$$

where $\overline{\phi} = (\phi(i))$ is the initial distribution and $P = (P_{i,j})$ the matrix of transition probabilities for the Markov chain. Note that only the second equality above explicitly makes use of the fact that $(X(n))_{n \in \mathbb{N}_0}$ is a Markov chain. Some probability distributions

The binomial distribution

The binomial distribution with integral parameter n and probability parameter p has support on the set $\{0, 1, \ldots, n\}$ and the density is given by

$$p_j = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, 1, \ldots, n.$$

The binomial distribution has mean np and variance $np(1-p)$.

The binomial distribution describes the distribution of the number of successes in n independent replications of an experiment with two possible outcomes (success/failure) with probability of success equal to p.

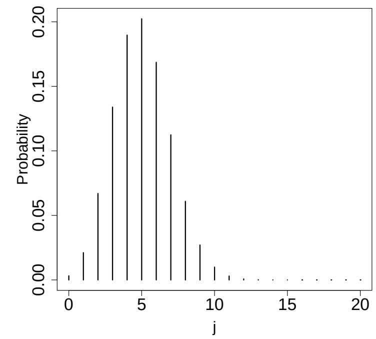


Figure B.1: Binomial distribution with integral parameter $n = 20$ and probability parameter $p = 1/4$.

The Poisson distribution

The Poisson distribution with parameter λ has support on the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and the density is given by

$$p_j = \frac{\lambda^j}{j!} e^{-\lambda}, \quad j \geq 0.$$

The Poisson distribution has mean λ and variance λ.

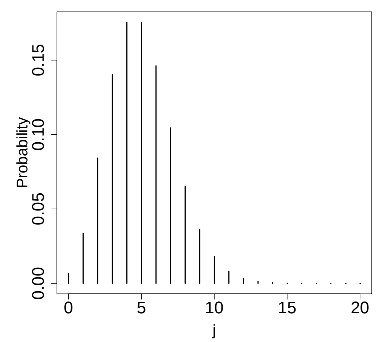


Figure B.2: Poisson distribution with parameter $\lambda = 5$.

The geometric distribution

The geometric distribution with probability parameter p has support on the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and the density is given by

$$p_j = (1-p)^j p, \quad j = 0, 1, 2, \ldots.$$

The geometric distribution has mean $\frac{1-p}{p}$ and variance $\frac{1-p}{p^2}$.

The geometric distribution describes the number of failures before the first success in a sequence of experiments with two possible outcomes (success/failure) with probability of success equal to p.

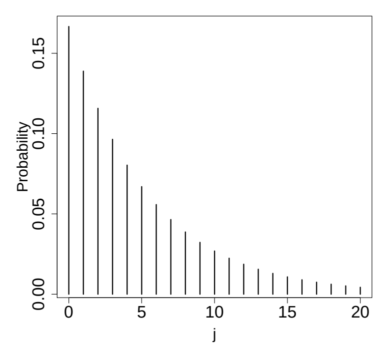


Figure B.3: Geometric distribution with probability parameter $p = 1/6$.

The negative binomial distribution

The negative binomial distribution with integral parameter r and probability parameter p has support on the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and the density is given by

$$p_j = \binom{r + j - 1}{j} p^r (1-p)^j, \quad j \geq 0.$$

The negative binomial distribution has mean $\frac{r(1-p)}{p}$ and variance $\frac{r(1-p)}{p^2}$.

The negative binomial distribution with probability parameter p and integer-valued integral parameter $r \in \mathbb{N}$ describes the distribution of the sum of r independent geometrically distributed random variables with probability parameter p.

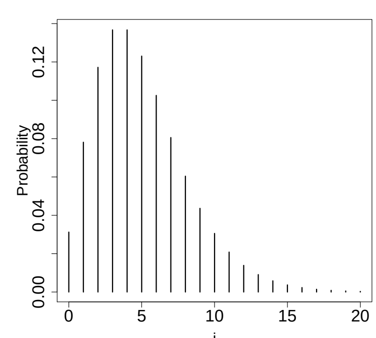


Figure B.4: Negative binomial distribution with integral parameter $r = 5$ and probability parameter $p = 1/2$.

The exponential distribution

The exponential distribution with rate parameter $\mu > 0$ is a continuous distribution on $[0, \infty)$ with density

$$f(t) = \mu \exp(-\mu t), \quad t > 0,$$

and cumulative distribution function

$$F(t) = \int_0^t f(s) ds = 1 - \exp(-\mu t), \quad t > 0.$$

The exponential distribution has mean $1/\mu$ and variance $1/\mu^2$.

For a continuous time Markov chain the distribution of the waiting time to the next jump follows an exponential distribution.

The gamma distribution

The gamma distribution with shape parameter λ and rate parameter $\mu > 0$ is a continuous distribution on $[0, \infty)$ with density

$$f(t) = \frac{t^{\lambda-1}\mu^{\lambda}}{\Gamma(\lambda)} \exp(-\mu t), \quad t > 0,$$

and cumulative distribution function

$$F(t) = \int_0^t f(s) ds, \quad t > 0.$$

The normalising constant in the density for the gamma distribution is given by the gamma integral

$$\Gamma(\lambda) = \int_0^{\infty} s^{\lambda-1} \exp(-s) ds$$

and for integer-valued shape parameter λ it holds that $\Gamma(\lambda) = (\lambda - 1)!$ The gamma distribution has mean λ/μ and variance λ/μ^2.

The gamma distribution with rate parameter μ and integer-valued shape parameter $\lambda \in \mathbb{N}$ is the distribution of the sum of λ independent exponentially distributed random variables with rate parameter μ.

B.2 *Some formulaes for sums and series*

In many of the exercises you are asked to compute the mean of the invariant distribution for Markov chains on a finite or countable state space, S. If the invariant probability vector is given as $\overline{\pi} = (\pi_i)$ then the mean is given as

$$\mu = \sum_{i \in S} i \pi_i.$$

For other exercises you have an unnormalized version $\overline{\nu} = (\nu_i)$ of an invariant vector and you need to find out if $\sum_{i \in S} \nu_i < \infty$ such that you can define the invariant probability as

$$\pi_j = \frac{\nu_j}{\sum_{i \in S} \nu_i}.$$

Some of the frequently occuring sums or series in this connection are

$$\begin{aligned}
\sum_{i=0}^{N} \alpha\beta^i &= \alpha\tfrac{1-\beta^{N+1}}{1-\beta}, \quad \alpha \in \mathbb{R}, \beta \neq 1 \\
\sum_{i=0}^{\infty} \alpha\beta^i &= \alpha\tfrac{1}{1-\beta}, \quad \alpha \in \mathbb{R}, |\beta| < 1 \\
\sum_{i=0}^{\infty} \alpha i\beta^i &= \alpha\tfrac{\beta}{(1-\beta)^2}, \quad \alpha \in \mathbb{R}, |\beta| < 1 \\
\sum_{i=0}^{\infty} \alpha\tfrac{\beta^i}{i!} &= \alpha\exp(\beta), \quad \alpha, \beta \in \mathbb{R} \\
\sum_{i=0}^{\infty} \alpha i\tfrac{\beta^i}{i!} &= \alpha\beta\exp(\beta), \quad \alpha, \beta \in \mathbb{R}.
\end{aligned}$$

B.3 Some results for matrices

Determinants of a square matrix

For a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

the **determinant** is defined as $\det(A) = a_{11}a_{22} - a_{12}a_{21}$. For a higher order square matrix A of dimension k the determinant may be defined recursively as

$$\det A = \sum_{j=1}^{k} (-1)^{1+j} a_{1j} \det A^*_{1j}, \quad \leftarrow \text{expansion by first row}$$

where

$$A^*_{1j} = \begin{pmatrix} a_{21} & a_{22} & \dots & a_{2j-1} & a_{2j+1} & \dots & a_{2k-1} & a_{2k} \\ a_{31} & a_{32} & \dots & a_{3j-1} & a_{3j+1} & \dots & a_{3k-1} & a_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kj-1} & a_{kj+1} & \dots & a_{kk-1} & a_{kk} \end{pmatrix}$$

is the $(k-1) \times (k-1)$ matrix obtained by removing from A all entries from row 1 or column j.

For a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

the definition leads to the following formula for the determinant

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{12}a_{21}.$$

Diagonalisation of matrices

Let A be a $k \times k$ matrix. An **eigenvalue** for A is a (real or complex) number λ such that there exists a nonzero **eigenvector** v with

$$Av = \lambda v.$$

The eigenvalues of A are exactly the zeroes of the **characteristic polynomial**

$$g(\lambda) = \det(A - \lambda I).$$

If $\lambda_1, \ldots, \lambda_k$ are the k roots of the characteristic polynomial P and $v_1, \ldots, v_k$ are corresponding eigenvectors then

$$\underbrace{A\begin{pmatrix} v_1 & \ldots & v_k \end{pmatrix}}_{:=O} = \underbrace{\begin{pmatrix} v_1 & \ldots v_k \end{pmatrix}}_{:=O} \underbrace{\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_k \end{pmatrix}}_{:=D}.$$

If O is invertible then we get the useful identity

$$A = ODO^{-1}.$$

Note that above we consider socalled right eigenvectors. Similarly one may consider left eigenvectors defined as row vectors $v \neq 0$ solving the equation

$$vA = \lambda v.$$

For some of the exercises in Chapters 2 and 3 we consider right eigenvectors.

Exponential matrices

For any $k \times k$ matrix A consider the matrices obtained by raising A to higher powers A^n. It turns out that the finite sums

$$\sum_{n=0}^{N} \frac{A^n}{n!} = I + \frac{A}{1!} + \frac{A^2}{2!} + \ldots + \frac{A^N}{N!}$$

converge as $N \to \infty$ (entry-by-entry). This allows us to define the exponential matrix $\exp(A)$ as the limit

$$\exp(A) := \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

Note that for convenience we use the notation A^0 for the identity matrix.

It is very important to note that the exponential matrix does not satisfy the same rules as the usual exponential function. In particular, except for very special cases it holds that

$$\exp(A+B) \neq \exp(A) \cdot \exp(B).$$

Closed form expressions for exponential matrices are rarely available. One important exception is the case where we can find an invertible matrix O such that

$$O^{-1}AO = D = \begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_k \end{pmatrix}$$

is a diagonal matrix. Using that $A = ODO^{-1}$ direct computations show that

$$\exp(A) = O^{-1} \begin{pmatrix} \exp(d_1) & 0 & 0 & \dots & 0 \\ 0 & \exp(d_2) & 0 & \dots & 0 \\ 0 & 0 & \exp(d_3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \exp(d_k) \end{pmatrix} O.$$

B.4 First order differential equations

For a continuous-time Markov chain on S with transition intensities $Q = (q_{i,j})_{i,j \in S}$, the transition probabilties

$$P_{i,j}(t) = P(X(s+t) = j | X(s) = i), i, j \in S,$$

always satisfy the backward differential equations

$$P'_{i,j}(t) = \sum_{l \in S} q_{i,l} \cdot P_{l,j}(t), j \in S$$

with the boundary condition that $P_{i,j}(0) = 0, i \neq j$, and $P_{i,i}(0) = 1$.

The solution has an explicit solution given as

$$P(t) = \exp(Qt), t \geq 0,$$

when the state space, S, is finite but computation of the exponential matrix may be infeasible. For Markov chains on countable state spaces no closed form formula for the transition probabilities exist.

Sometimes we can get nice explicit formulas for some of the transition probabilities, $P_{i,j}(t)$, by solving some of the backward or forward differential equations. Remember that the forward differential equations take the form

$$P'_{i,j}(t) = \sum_{l \in S} P_{i,l}(t) \cdot q_{l,j}, j \in S.$$

To solve the differential equations you might find it useful to know that

$$\begin{aligned} f'(t) = \alpha \exp(\beta t) &\Rightarrow f(t) = \tfrac{\alpha}{\beta} \exp(\beta t) + c \\ f'(t) = \beta f(t) &\Rightarrow f(t) = c \exp(\beta t) \\ f'(t) = \alpha f(t) + \beta \exp(\gamma t) + \delta &\Rightarrow f(t) = \tfrac{\beta}{\gamma - \alpha} \exp(\gamma t) - \tfrac{\delta}{\alpha} + c \exp(\alpha t) \end{aligned}$$

where c is a constant. Note that the last expression is only valid for $\gamma \neq \alpha$.

B.5 Second order linear recurrence equations

Many of the deeper results on Markov chains in Chapters 2 and 3 are stated in terms of the solution to a system of equations. For Markov chains allowing only jumps of size one the system of equations will occasionally take the following form

$$az_{i-1} + bz_{i+1} \quad = \quad z_i, \quad l < i < u, \tag{B.1}$$

where l or u can be $-\infty$ or $+\infty$. [1] It is clear that if we know z_j, z_{j+1} for some time index j (and if $a, b \neq 0$) then we may recursively determine the values of z_i for the remaining indicies i. In mathematical terms one can formally show that the solution to (B.1) is a vector space of dimension 2 and we shall below describe two linearly independent solutions.

[1] We emphasize that in many cases we will get a system of equations like (B.1), but where the coefficients a and b depend on i. This case is not covered by the solution strategy described in this paragraph.

We express the solution in terms of the roots

$$\alpha_1 = \frac{1 + \sqrt{1 - 4ab}}{2b} \quad , \quad \alpha_2 = \frac{1 - \sqrt{1 - 4ab}}{2b}$$

to the characteristic equation for (B.1), which is given as

$$\alpha = a + b\alpha^2.$$

We give the solution for the two cases depending on whether there are two distinct roots.

($\alpha_1 \neq \alpha_2$) Any solution to (B.1) can be written on the form

$$z_i = c_1 \alpha_1^i + c_2 \alpha_2^i, \quad l \leq i \leq u.$$

($\alpha_1 = \alpha_2$) Here $\alpha_1 = \alpha_2 = \frac{1}{2b}$ and any solution to (B.1) can be written on the form

$$z_i = c_1 \left(\frac{1}{2b}\right)^i + c_2 i \left(\frac{1}{2b}\right)^i, \quad l \leq i \leq u.$$

The constants c_1, c_2 can be found from boundary conditions imposed by the relevant model.

B.6 The ratio test

Consider a series

$$\sum_{n=1}^{\infty} a_n$$

of real or complex numbers a_n. The simplest form of the ratio test suggests that we try to compute the limit

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

and

- if $L < 1$ then the series converges absolutely
- if $L > 1$ then the series do not converge.

If the limit ($= L$) does not exist or equals 1, then no conclusion may be drawn from this formulation of the ratio test.

B.7 Integral test for convergence

For a non-negative (Lebesgue measurable), monotonely decreasing function f on $[1, \infty)$ then we have the upper and lower bounds

$$\int_1^{\infty} f(x)dx \leq \sum_{n=1}^{\infty} f(n) \leq f(1) + \int_1^{\infty} f(x)dx.$$

Given a series $\sum_{n=1}^{\infty} a_n$ of non-negative terms a_n this implies that if we can find a non-negative, monotonely descresing function f on $[1, \infty)$ with $f(n) = a_n$ then

- the series converge if $\int_1^{\infty} f(x)dx < +\infty$
- the series diverge if $\int_1^{\infty} f(x)dx = +\infty$.

The integral test is often used in slightly modified setups. For example, to demonstrate convergence of a series, we only need $f(n) \geq a_n$ for n sufficiently large with f being integrable towards $+\infty$.

B.8 *How to do certain computations in R*

Matrices and systems of linear equations

The following code defines a 4×4 matrix P of transition probabilities for a discrete time Markov chain.

```
P<-matrix(nrow=4, ncol=4)
P[1, ]<-c(0, 4/6 ,1/6 ,1/6)
P[2, ]<-c(0, 0 ,5/7 ,2/7)
P[3, ]<-c(0, 3/5 ,0 ,2/5)
P[4, ]<-c(0, 1/2 ,1/2 ,0)
P

##      [,1]      [,2]      [,3]      [,4]
## [1,]    0 0.6666667 0.1666667 0.1666667
## [2,]    0 0.0000000 0.7142857 0.2857143
## [3,]    0 0.6000000 0.0000000 0.4000000
## [4,]    0 0.5000000 0.5000000 0.0000000
```

To compute the n step probabilities given by P^n you need to know how to do matrix multiplication. Below we demonstrate how to compute P^2, P^4, and P^8.

```
P2 <- P %*% P
P2

##      [,1]      [,2]      [,3]      [,4]
## [1,]    0 0.1833333 0.5595238 0.2571429
## [2,]    0 0.5714286 0.1428571 0.2857143
## [3,]    0 0.2000000 0.6285714 0.1714286
## [4,]    0 0.3000000 0.3571429 0.3428571

P4 <- P2 %*% P2
P4

##      [,1]      [,2]      [,3]      [,4]
## [1,]    0 0.2938095 0.4697279 0.2364626
## [2,]    0 0.4408163 0.2734694 0.2857143
## [3,]    0 0.2914286 0.4848980 0.2236735
## [4,]    0 0.3457143 0.3897959 0.2644898

P8 <- P4 %*% P4
P8

##      [,1]      [,2]      [,3]      [,4]
## [1,]    0 0.3481567 0.4002902 0.2515532
```

```
## [2,]    0 0.3727913 0.3645248 0.2626839
## [3,]    0 0.3471067 0.4020098 0.2508835
## [4,]    0 0.3574321 0.3866506 0.2559174
```

The invariant distribution $\overline{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$ for the Markov chain must solve the equation $\overline{\pi}P = \overline{\pi}$. In other words the invariant probability is a normalized left eigenvector of P associated with eigenvalue 1. Below we demonstrate how to find left eigenvectors and extract a normalized version of the eigenvector with eigenvalue 1.

```
lefteigen <- eigen(t(P))
lefteigen

## $values
## [1]  1.000000 -0.646385 -0.353615  0.000000
##
## $vectors
##           [,1]       [,2]       [,3]       [,4]
## [1,] 0.0000000  0.0000000  0.0000000  0.7467709
## [2,] 0.6133511 -0.5457060 -0.3294501  0.1633561
## [3,] 0.6571619  0.7988318 -0.4822656 -0.4278375
## [4,] 0.4381080 -0.2531258  0.8117158 -0.4822895

normInv <- lefteigen$vectors[, 1]/sum(lefteigen$vectors[, 1])
normInv

## [1] 0.0000000 0.3589744 0.3846154 0.2564103
```

Note that (in accordance with theory) the colums of P^n approaches the invariant probability vector computed above as $n \to \infty$.

In Section B.8 below we define a transition matrix Q of a continuous time Markov chain on four states. An invariant distribution $\overline{\pi}$ for this chain must satisfy the matrix equation $\overline{\pi}Q = 0$ as well as the condition $\sum_{i \in S} \pi_i = 1$. One way to compute the invariant distribution in R is to define the matrix $\overline{Q}$ obtained by adding to Q a column of ones and then solve the equation $\pi\overline{Q} = (0, 0, 0, 0, 1)$. The code below works to find the invariant distribution in any case where only one recurrent class of states exist such that $\overline{\pi}$ is unique.

```
Q <- matrix(nrow = 4, ncol = 4)
Q[1, ] <- c(-6, 4, 1, 1)
Q[2, ] <- c(0, -7, 5, 2)
Q[3, ] <- c(0, 3, -5, 2)
Q[4, ] <- c(0, 0.5, 0.5, -1)
```

```
Q1 <- cbind(Q, 1)
Q1

##      [,1] [,2] [,3] [,4] [,5]
## [1,]   -6  4.0  1.0    1    1
## [2,]    0 -7.0  5.0    2    1
## [3,]    0  3.0 -5.0    2    1
## [4,]    0  0.5  0.5   -1    1

lm.fit(t(Q1), c(0, 0, 0, 0, 1))$coefficients

##            x1            x2            x3            x4
## -1.497024e-17  1.333333e-01  2.000000e-01  6.666667e-01

round(lm.fit(t(Q1), c(0 ,0 ,0 ,0 ,1))$coefficients, digits=6)

##       x1       x2       x3       x4
## 0.000000 0.133333 0.200000 0.666667
```

Note that the last line of code rounds the solution down to 6 significant digits showing that the invariant probability is given by $\overline{\pi} = (0, 2/15, 1/5, 2/3)$.

Computing exponential matrices

For a continuous time Markov chain on a finite state space the transition probabilities

$$P_{ij}(t) = P(X(t+s) = j | X(s) = i)$$

is given as the entries of the matrix $\exp(Qt)$, where Q is the intensity matrix of the Markov chain. The `MatrixExp` function of the `msm`-package may be used to compute exponential matrices in R. Below we demonstrate how to compute the transition probabilities of the four state Markov chain with transition intensity matrix

$$Q = \begin{pmatrix} -6 & 4 & 1 & 1 \\ 0 & -7 & 5 & 2 \\ 0 & 3 & -5 & 2 \\ 0 & 0.5 & 0.5 & -1 \end{pmatrix}.$$

Note that before running the following code on your computer you must install the `msm` package. Initially we compute the matrix of transition probabilities at time $t = 0.1$.

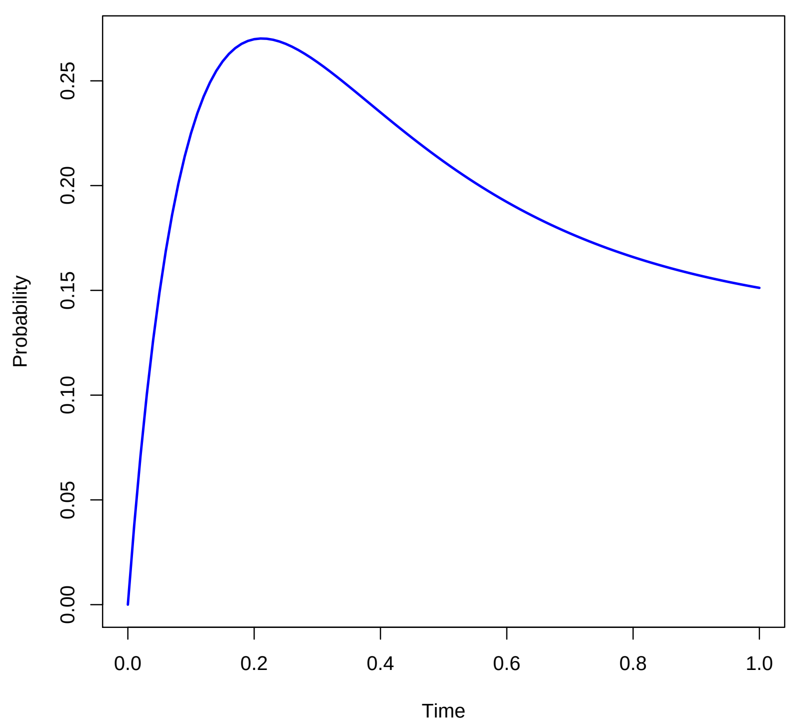
```
library(msm)
P_1 <- MatrixExp(Q, 0.1)
P_1
```

```
##           [,1]       [,2]       [,3]      [,4]
## [1,] 0.5488116 0.22501400 0.11738870 0.1087857
## [2,] 0.0000000 0.54095713 0.28625502 0.1727879
## [3,] 0.0000000 0.17307769 0.65413446 0.1727879
## [4,] 0.0000000 0.03988527 0.04650866 0.9136061
```

You will often need to find the transition probabilities at several values of the time argument for instance if you want to plot the transtion probabilities as function of time. Below we compute the transition probabilities at all time points between 0 and 1 in steps of 0.01. The result is stored as a three dimensional array and we demonstrate how to plot the function $t \to P_{12}(t)$.

```
timearg <- seq(0,1,by=0.01)
res <- lapply(timearg, function(t){MatrixExp(Q, t)})
trprob <- array(unlist(res), dim=c(dim(res[[1]]), length(res)))
plot(timearg, trprob[1,2,], lwd=2, col="blue", type='l'
,xlab="Time", ylab="Probability")
```

Figure B.5: Transition probability $P_{1,2}(t)$ for a continuous-time Markov chain.



Simulation of Markov chains

The sample path of a Markov chain may easily be simulation using the dynamical description summarised in the transition diagram. The following code defines a function that can simulate sample paths for both discrete- and continuous-time Markov chains.

```
simMC<-function(tr, nJump = 10, phi0 = NULL){
nStates<-dim(tr)[1]
cont <- (sum(tr) == 0)
if(cont){
  pJump <- matrix(data = 0, nrow = nStates, ncol = nStates)
  for(i in 1:nStates){pJump[i, -i] <- (-tr[i, -i]/tr[i, i])}}
else{pJump<-tr}

if(is.null(phi0)){phi0 <- c(1, rep(0, nStates-1))}
states<-rep(0,nStates+1)
jumps<-rep(0,nStates+1)
states[1] <- sample(nStates, 1, prob = phi0)
tmax <- 0
for(i in 1:nJump){
if(cont){
  jumps[i+1] <- (-rexp(1)/tr[states[i], states[i]]) + jumps[i]
  }
else{jumps[i+1] <- i}
states[i+1] <- sample(nStates, 1, prob = pJump[states[i], ])
}
return(list(y=states,t=jumps))
}
```

The following code simulates and displays the sample path up to step 25 for the discrete time Markov chain with transition matrix P of Section B.8 and initial distribution $\overline{\phi} = (1/4, 1/4, 1/4, 1/4)$.

```
mcDisc <- simMC(P, nJump = 25, phi0 = rep(0.25, 4))
mcDisc$y

##  [1] 1 2 3 4 2 3 2 4 2 4 2 3 2 3 2 3 2 4 2 4 2 3 4 3 2 3
```

Below we simulate and plot the sample path for the first 10 jumps of the continuous-time Markov chain with intensity matrix Q of Section B.8 for initial distribution given by $\overline{\phi} = (1, 0, 0, 0)$.

The function `simMC` does not apply for simulation of Markov chains on countable state spaces. However, for the most common examples discussed in these lecture notes it should be easy (or at

```
set.seed(2013)
mcCont<-simMC(Q, nJump = 10, phi0 = c(1, 0, 0, 0))
plot(mcCont$t, mcCont$y, type = 's', xlab = "Time"
     , ylab = "State", axes = F, col = "blue", lwd = 2)
axis(side = 1)
axis(side = 2, at = 1:4)
```

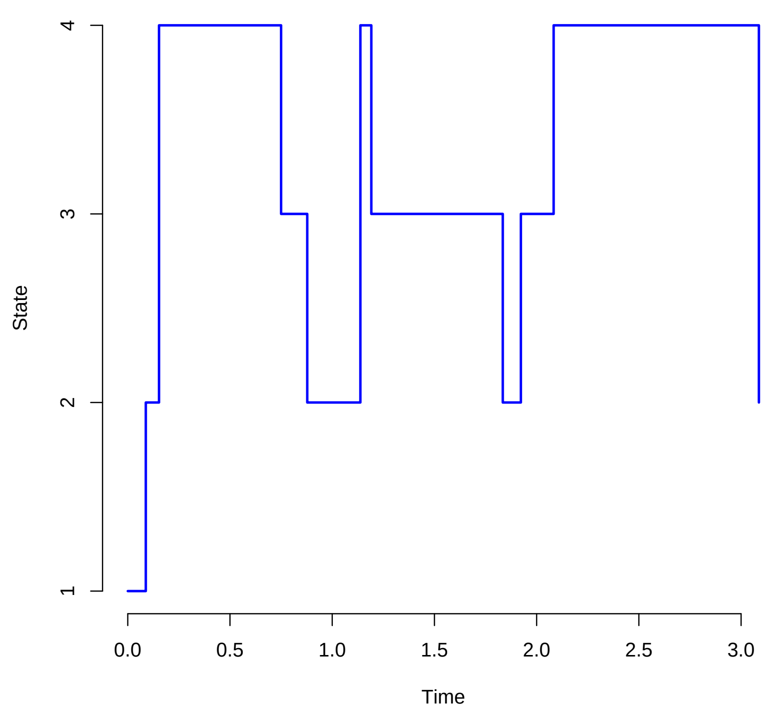


Figure B.6: Simulated sample path of a continuous-time Markov chain.

least possible) to write simple functions for simulation of the sample paths based on the transition diagram of the Markov chain.

The Poisson process is a continuous time Markov chain on $\mathbb{N}_0$ moving only in jumps of size 1 hence everything simplifies as we only need to simulate the jump times. The waiting times between jumps are independent and identically distributed exponential variables with rate parameter λ (-the intensity of the Poisson process). Below we show how to simulate the first 50 jump times of a Poisson process with intensity 1 and plot the resulting sample path.

```
set.seed(2013)
wait <- rexp(50, rate = 1)
t <- cumsum(c(0, wait))
plot(0:50, t, type = 's', lwd = 2, col = "blue"
     , xlab = "Time", ylab = "Number of jumps")
```

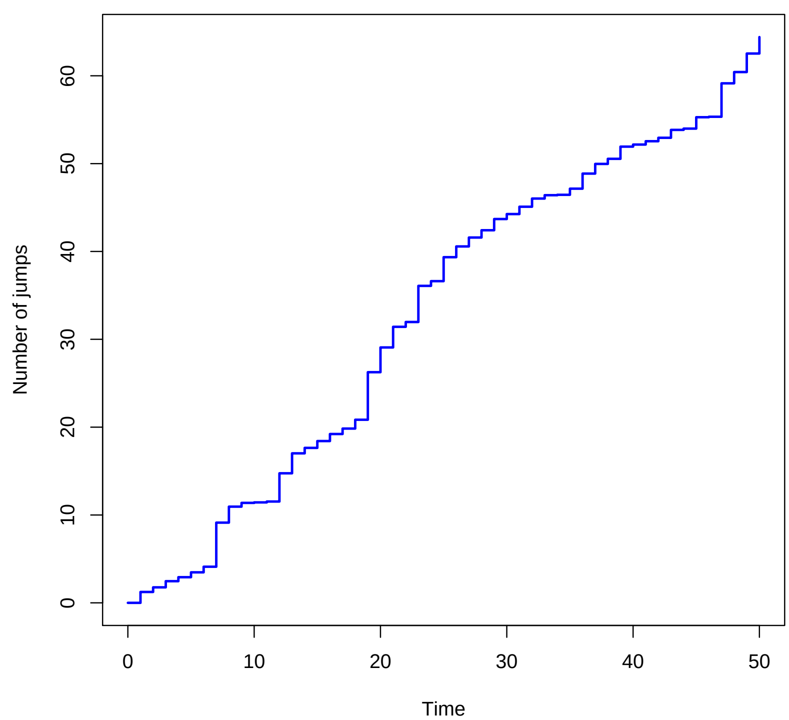


Figure B.7: Simulated sample path of a Poisson process with intensity 1.

C
Proofs of selected results

C.1 Recurrence criterion 1

In this exercise we outline a strategy for proving the recurrence criterion given in Theorem 10 on page 22 of Chapter 2. The complete proof may be found on page 35-36 in Jacobsen and Keiding [1985]. We consider a discrete-time Markov chain $(X(n))_{n\geq 0}$ on S and define the time of the first visit to state j

$$T_j = \inf\{n > 0 | X(n) = j\}.$$

We further introduce the probability

$$f_{ij}^{(n)} = P(X(n) = j, X(n-1), \ldots, X(1) \neq j | X(0) = i)$$

that the first visit to state j happens at time n assuming that $P(X(0) = i) = 1$. The quantity

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

then describes the probability of ever reaching state j if $P(X(0) = i) = 1$ and you are reminded that state i is recurrent if and only if $f_{ii} = 1$.

1. By splitting the event $(X(n) = j)$ according to the time of the *first* visit to state j show that

$$(P^n)_{i,j} = \sum_{m=1}^{n} (P^{n-m})_{j,j} f_{ij}^{(m)}.$$

2. Summing the expression of question 1. over $n = 1, \ldots, N$ verify the following upper bound

$$\sum_{n=1}^{N} (P^n)_{i,j} \leq \sum_{m=1}^{N} f_{ij}^{(m)} \cdot \sum_{k=0}^{N} (P^k)_{j,j}$$

for $N > 0$.

3. Show that for any $M < N$ then we have a lower bound

$$\sum_{n=1}^{N} (P^n)_{i,j} \geq \sum_{m=1}^{M} f_{ij}^{(m)} \cdot \sum_{k=0}^{N-M} (P^k)_{j,j}.$$

4. Keeping $M < N$ fixed divide the inequalities of 2. and 3. by $\sum_{n=0}^{N}(P^{nn})_{jj}$ to show that

$$\sum_{m=1}^{M} f_{ij}^{(m)} \cdot \left(1 - \frac{\sum_{n=N-M+1}^{N}(P^n)_{j,j}}{\sum_{n=0}^{N}(P^n)_{j,j}}\right) \leq \frac{\sum_{n=1}^{N}(P^n)_{i,j}}{1+\sum_{n=1}^{N}(P^n)_{j,j}} \leq \sum_{m=1}^{N} f_{ij}^{(m)}.$$

5. Let $N \to \infty$ in the expression of question 4. to get that for any $M > 0$ then

$$\begin{aligned} \sum_{m=1}^{M} f_{ij}^{(m)} &\leq \liminf_{N\to\infty} \frac{\sum_{n=1}^{N}(P^n)_{i,j}}{1+\sum_{n=1}^{N}(P^n)_{j,j}} \\ &\leq \limsup_{N\to\infty} \frac{\sum_{n=1}^{N}(P^n)_{i,j}}{1+\sum_{n=1}^{N}(P^n)_{j,j}} \leq \sum_{m=1}^{\infty} f_{ij}^{(k)} = f_{ij}. \end{aligned}$$

6. Finally, let $M \to \infty$ to get that

$$f_{ij} = \frac{\sum_{n=1}^{\infty}(P^n)_{i,j}}{1+\sum_{n=1}^{\infty}(P^n)_{j,j}}.$$

7. Consider the case $i = j$ to get the result in Theorem 10 of Chapter 2.

C.2 *Number of visits to state j*

In this exercise we give a proof of Theorem 12 from Chapter 2. We study the total number of visits to state j defined as

$$N_j = \sum_{n=1}^{\infty} 1(X(n) = j).$$

More precisely, we show that

$$P(N_j = n|X(0) = j) = f_{jj}^{n}(1 - f_{jj}), \quad n \geq 0, \tag{C.1}$$

where we refer to Exercise C.1 for an explanation of the notation.

1. Split the event $(N_j \geq n+1)$ into the time of the first visit to state j and use the Markov property to obtain

$$\begin{aligned} &P(N_j \geq n+1|X(0) = i) \\ = &\sum_{k=1}^{\infty} P(N_j \geq n|X(0) = j)P(T_j = k|X(0) = i) \\ = &P(N_j \geq n|X(0) = j)f_{ij}, \quad n \geq 1. \end{aligned}$$

2. Use question 1. combined with the identity

$$P(N_j = n|X(0) = j) = P(N_j \geq n|X(0) = j) - P(N_j \geq n+1|X(0) = j),$$

and the initial condition $P(N_j \geq 1|X(0) = j) = f_{jj}$ to verify that

$$P(N_j = n|X(0) = j) = f_{jj}^n(1 - f_{jj}), \quad n \geq 1.$$

If the following questions we discuss the implication of (C.1) a little further. Remember that (by definition!) state j is recurrent if and only if $f_{jj} = 1$.

3. Argue that $P(N_j = 0|X(0) = j) = 1 - f_{jj}$.

4. Argue that if state j is transient then the numbers of visits to state j follows a geometric distribution. Write down an expression for the expected number, $E[N_j|X(0) = j]$, of visits to state j.

5. Argue that if state j is recurrent then

$$P(N_j = +\infty|X(0) = j) = 1.$$

6. Give a heuristic argument that for any initial state $i \neq j$ it holds that

$$\begin{aligned} P(N_j = 0|X(0) = i) &= 1 - f_{ij} \\ P(N_j = n|X(0) = i) &= f_{jj}^{n-1}(1 - f_{jj})f_{ij} \quad , n \in \mathbb{N} \\ P(N_j = +\infty|X(0) = i) &= f_{ij} \cdot 1(f_{jj} = 1). \end{aligned}$$

Comment From questions 6. and 7. of Exercise C.1 it follows easily that for a transient state j then

$$\sum_{n=0}^{\infty}(P^n)_{jj} = \frac{1}{1 - f_{jj}} \text{ and } \sum_{n=0}^{\infty}(P^n)_{ij} = \frac{f_{ij}}{1 - f_{jj}}.$$

Since it is trivial to see that

$$E[N_j|X(0) = i] = \sum_{n=0}^{\infty}(P^n)_{ij}$$

this gives us an expression for the expected number of visits to state j. The present exercise, however, gives a complete description of the distribution of N_j providing us with an expression for the density

$$P(N_j = n|X(0) = i).$$

C.3 Invariant distribution

In this exercise we sketch the proof of Theorem 22 in Chapter 2. We consider an irreducible, recurrent Markov chain on S and we shall discuss the existence of solutions to the system of equations

$$\nu(j) = \sum_{i \in S} \nu(i) P_{i,j}, \quad j \in S. \tag{C.2}$$

For any $i \in S$ we introduce the return time to any fixed state i

$$T_i = \inf\{n > 0 | X(n) = i\}$$

and the expected number of visits to state j before first visit to state i

$$\nu^{(i)}(j) = E\left[\sum_{n=0}^{T_i - 1} 1(X(n) = j) | X(0) = i\right].$$

1. Argue that $\nu^{(i)}(i) = 1$ and by recurrence of state i then that

$$\begin{aligned}\nu^{(i)}(j) &= E\left[\sum_{n=0}^{T_i - 1} 1(X(n) = j) | X(0) = i\right] \\ &= E\left[\sum_{n=1}^{T_i} 1(X(n) = j) | X(0) = i\right].\end{aligned}$$

2. Verify that

$$\begin{aligned}& E\left[\sum_{n=1}^{T_i} 1(X(n) = j) | X(0) = i\right] \\ = \; & E\left[\sum_{n=1}^{\infty} 1(X(n) = j, X(n-1), \ldots, X(1) \neq i | X(0) = i\right].\end{aligned}$$

3. Use the Markov property to show that for $n \geq 2$ then

$$\begin{aligned}& P(X(n) = j, X(n-1), \ldots, X(1) \neq i | X(0) = i) \\ = \; & \sum_{l \neq i} P(X(n) = j, X(n-1) = l, X(n-2), \ldots, X(1) \neq i | X(0) = i) \\ = \; & \sum_{l \neq i} P_{l,j} \cdot P(X(n-1) = l, X(n-2), \ldots, X(1) \neq i | X(0) = i) \\ = \; & \sum_{l \neq i} P_{l,j} \cdot E[1(X(n-1) = l, X(n-2), \ldots, X(1) \neq i) | X(0) = i].\end{aligned}$$

4. Use questions 1.-3. to show that

$$\begin{aligned}
& \nu^{(i)}(j) \\
= \; & P(X(1) = j | X(0) = i) \\
+ \; & \sum_{n=2}^{\infty} P(X(n) = j, X(n-1), \ldots, X(1) \neq i | X(0) = i) \\
= \; & P_{i,j} + \sum_{l \neq i} P_{l,j} \cdot \sum_{n=2}^{\infty} P(X(n-1) = l, X(n-2), \ldots, X(1) \neq i | X(0) = i) \\
= \; & P_{i,j} + \sum_{l \neq i} P_{l,j} \cdot \sum_{n=2}^{\infty} E[1(X(n-1) = l, X(n-2), \ldots, X(1) \neq i) | X(0) = i] \\
= \; & P_{i,j} + \sum_{l \neq i} P_{l,j} \cdot E[\sum_{n=1}^{T_i} 1(X(n) = l) | X(0) = i] \\
= \; & \sum_{l \in S} P_{l,j} \cdot E[\sum_{n=1}^{T_i} 1(X(n) = l) | X(0) = i] \\
= \; & \sum_{l \in S} \nu^{(i)}(l) P_{l,j}.
\end{aligned}$$

In questions 1.-4. we have demonstrated that for any $i \in S$ then the vector $(\nu^{(i)}(j))_{j \in S}$ solves (C.2). Trivially, $\nu^{(i)}(j) \geq 0$ and we shall discuss when the total mass $\sum_{j \in S} \nu^{(i)}(j)$ is finite.

5. Show that $\sum_{j \in S} \nu^{(i)}(j) = E[T_i | X(0) = i]$ and argue that $(\nu^i(j))_{j \in S}$ may be normalized into an invariant distribution (=probability) exactly if state i is positive recurrent.

We have now showed the existence of an invariant distribution for any irreducible, positive recurrent Markov chain in discrete time. An almost complete proof of the uniqueness part of Theorem 22 in Chapter 2 may be constructed along the lines given in the following questions 6.-10.

6. Show that for any solution to (C.2) it holds for $m \geq 1$ and any $l \in S$ that
$$\nu(j) = \sum_{i \in S} \nu(i)(P^m)_{i,j} \geq \nu(l)(P^m)_{l,j}.$$
Deduce that for any non-negative solution (different from zero!) we have that $\nu(j) > 0$ for all $j \in S$.

7. Let $\overline{\nu} = (\nu(j))_{j \in S}$ be any non-zero solution to (C.2). Argue from question 6. that we may assume that $\nu(i) = 1$ where i is any fixed state $i \in S$.

8. Use (without proof!) that for any solution to (C.2) with $\nu(i) = 1$ it holds that for all $j \in S$
$$\nu(j) \geq \nu^{(i)}(j).$$

9. Use question 1., 7. and 8. to argue that the vector $\overline{\mu} = (\mu(j))_{j\in S}$ defined by $\mu(j) = \nu(j) - \nu^{(i)}(j)$ is a non-negative solution to (C.2)

10. Use question 6. and 9. to deduce that for all $j \in S$ then

$$\mu(j) = 0$$

and conclude that $(\nu^{(i)}(j))_{j\in S}$ is the unique solution to (C.2) with i-th coordinate equal to 1.

Note that once we have showed the uniqueness (up to multiplication!) of solutions to (C.2) then it follows from question 5. that if for some $i_0 \in S$ $E[T_{i_0}|X(0) = i_0] < +\infty$ for some $i_0 \in S$ then for *any* $i \in S$ it holds that $E[T_i|X(0) = i] < +\infty$. In particular, the states in a recurrent class are either all positive recurrent or all null-recurrent. This was postulated in Definiton 21.

Since the solution to (C.2) is unique (up to multiplication) we further conclude that there is a unique solution, $\overline{\pi} = (\pi(j))_{j\in S}$, to (C.2)) with $\sum_{j\in S} \pi(j) = 1$ and that the solution may be represented as

$$\pi(j) = \frac{\nu^{(i)}(j)}{E[T_i|X(0) = i]} = \frac{E[\sum_{n=0}^{T_i-1} 1(X(n) = j)|X(0) = i]}{E[T_i|X(0) = i]}$$

for any fixed $i \in S$. Choosing $j = i$ above we conclude that

$$\pi(j) = \frac{1}{E[T_j|X(0) = j]}$$

hence the invariant probability mass in state j equals the inverse mean return time to state j.

C.4 Uniqueness of invariant distribution

In this exercise we consider the proof of Theorem 48 on page 80 of Chapter 3. Let $(X(t))_{t\geq 0}$ be an irreducible continuous-time Markov chain on S and let $\overline{\pi} = (\pi(i))_{i\in S}$ be a probability. Assume that for some $t_0 > 0$ then

$$\forall j \in S \quad : \quad \pi(j) = \sum_{i\in S} \pi(i) P_{i,j}(t_0). \tag{C.3}$$

1. Use that $\pi(i_0) > 0$ and that for irreducible Markov chain then $P_{i_0,j}(t_0) > 0$ to verify from (C.3) that

$$\forall j \in S \quad : \quad \pi(j) > 0.$$

2. Sum (C.3) over $j \in S$ to obtain

$$\sum_{j\in S} \pi(j) = \sum_{j\in S}\sum_{i\in S} \pi(i) P_{i,j}(t_0) = \sum_{i\in S} \pi(i) \sum_{j\in S} P_{i,j}(t_0) \leq \sum_{i\in S} \pi(i)$$

and conclude that

$$\forall i \in S \quad : \quad \sum_{j \in S} P_{i,j}(t_0) = 1.$$

3. Argue that $\overline{\pi}$ is the unique invariant distribution for the discrete-time Markov chain on S with transition probabilities $P(t_0) = (P_{i,j}(t_0))_{i,j \in S}$.

4. For an arbitrary $t > 0$ find n such that $t < nt_0$. Use that $P(nt_0) = (P(t_0))^n$ to deduce that $P(t)$ is a transition matrix, i.e. that

$$\forall i \in S \quad : \quad \sum_{j \in S} P_{i,j}(t) = 1.$$

5. Verify that

$$\overline{\pi}P(t)P(t_0) = \overline{\pi}P(t_0)P(t) = \overline{\pi}P(t)$$

and show that $\overline{\pi}P(t)$ is an invariant distribution for $P(t_0)$. Use question 3. to conclude that $\overline{\pi} = \overline{\pi}P(t)$.

C.5 *On the ergodic theorem for discrete-time Markov chains*

In this section we discuss a classical proof of the ergodic theorem for discrete-time Markov chains (Theorem 20 on page 32). The proof is based on the **renewal theorem** in discrete-time and follows the exposition in Asmussen [2003, Chapter 1.2].

Example 74 *The purpose of the present example is to motivate the formulation of the main result (Theorem 20). We consider a very simple board game where the player moves along spaces that are labelled* $0, 1, 2, \dots$.

In each round the player rolls a six sided dice and moves forward along the board according to the number of eyes of the dice. We assume that the game starts at field 0 *and the goal is to compute the probability,* u_n*, that the player will hit a trap that is hidden on the field labeled* n *for some* $n \in \mathbb{N}$.

Let us try to compute $u_1, u_2, \dots$ *by brute force. Clearly, we have that* $u_1 = 1/6$ *since we only hit field* 1 *if the first roll shows the face with exactly one eye. Otherwise we will just pass field* 1*. We may hit field* 2 *either by moving 2 steps ahead in first roll with the dice or by rolling one eye twice. In particular, we get that*

$$u_2 = 1/6 + 1/6 \cdot 1/6 = 7/36.$$

Another way to decompose the computation is to say that field 2 *may be reached either by rolling two eyes or by rolling one eye and then hitting field*

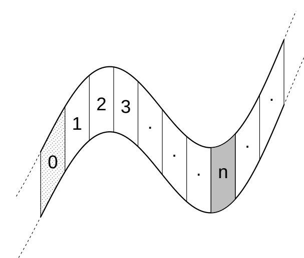


Figure C.1: Figure for illustration of the simple game discussed in Example 74.

2 *given that we restart the game at state* 1. *Clearly, the latter probability is exactly* u_1 *and we may rewrite the computation of* u_2 *as*

$$u_2 = 1/6 + 1/6 \cdot u_1.$$

Exploiting a similar idea we may now write down

$$\begin{aligned}
u_3 &= 1/6 + 1/6u_1 + 1/6u_2 \\
u_4 &= 1/6 + 1/6u_1 + 1/6u_2 + 1/6u_3 \\
u_5 &= 1/6 + 1/6u_1 + 1/6u_2 + 1/6u_3 + 1/6u_4 \\
u_6 &= 1/6 + 1/6u_1 + 1/6u_2 + 1/6u_3 + 1/6u_4 + 1/6u_5 \\
u_7 &= 1/6u_1 + 1/6u_2 + 1/6u_3 + 1/6u_4 + 1/6u_5 + 1/6u_6 \\
&\vdots \\
u_n &= 1/6u_{n-6} + 1/6u_{n-5} + 1/6u_{n-4} + 1/6u_{n-3} + 1/6u_{n-2} + 1/6u_{n-1},
\end{aligned}$$

allowing us to compute any u_n *by recursion. We find that*

$$u_1 \approx 0.1667, u_5 \approx 0.3088, u_{10} \approx 0.2893, u_{20} \approx 0.2856, u_{100} \approx 0.2857.$$

It appears that the probability, u_n*, of hitting a trap on field* n *tends to some limit as* $n \to \infty$. *Is this a coincidence? And can we give an interpretation of the limit?*

To further analyse the problem we introduce the notation f_n *for the probability of moving* n *steps ahead at each round. We have that*

$$f_n = \begin{cases} 1/6 & , n = 1, 2, 3, 4, 5, 6 \\ 0 & , n \geq 7 \end{cases}$$

and the general expression for u_n *may be given a more compact form*

$$\begin{aligned}
u_n &= f_6 u_{n-6} + f_5 u_{n-5} + f_4 u_{n-4} + f_3 u_{n-3} + f_2 u_{n-2} + f_1 u_{n-1} \\
&= \sum_{k=1}^{n} f_k u_{n-k},
\end{aligned}$$

if we let $u_0 \equiv 1$. *It turns out that sequences satisfying the recursive equation above will have a limit under mild regularity conditions and that the limit equals* $1/\mu$ *where*

$$\mu = \sum_{k=1}^{\infty} k f_k.$$

For our board game we have $\mu = 1/6(1 + 2 + 3 + 4 + 5 + 6) = 21/6 = 7/2$ *implying that* $u_n \to 2/7$ *for* $n \to \infty$. *Note that* $\mu = 7/2$ *equals the average number of eyes obtained by rolling a dice.* □

The general framework covering the situation in Example 74 deals with renewal sequences defined as solutions to the equation

$$u_n = \sum_{k=1}^{n} f_k u_{n-k}$$

where $(f_k)_{k\in\mathbb{N}}$ is a probability vector. We define the period of $(f_k)_{k\in\mathbb{N}}$ as the greatest common divisor of

$$D_f = \{k \in \mathbb{N} | f_k > 0\}.$$

We say that (u_n) is an aperiodic renewal sequence if (f_k) has period 1.

Theorem 75 (Renewal theorem in discrete-time) *For an aperiodic renewal sequence (u_n) generated by (f_k) it holds that*

$$\lim_{n\to\infty} u_n = \frac{1}{\sum_{k=1}^{\infty} k f_k}.$$

The Proof may be found in Asmussen [2003, Chapter 1.2]. □

Theorem 76 (Limit for discrete-time Markov chains) *Let $(X(n))_{n\geq 0}$ be an irreducible, positive recurrent and aperiodic Markov chain on a finite or countable S. Then for any initial distribution $\overline{\phi}$ and $j \in S$ it holds that*

$$\lim_{n\to\infty} P(X(n) = j) = \frac{1}{E[T_j | X(0) = j]}$$

where

$$T_j = \inf\{n > 0 | X(0) = i\}.$$

Proof: For any fixed $j \in S$ let $f_k = P(T_j = k | X(0) = j)$. By splitting the event $(X(n) = j)$ according to the time of the first visit to state j on $\{0, 1, 2, \ldots, n\}$ we get from the Markov property that

$$\begin{aligned} u_n &= P(X(n) = j | X(0) = j) \\ &= \sum_{k=1}^{n} P(T_j = k | X(0) = i) P(X(n-k) = j | X(0) = j) \\ &= \sum_{k=1}^{n} f_k u_{n-k}. \end{aligned}$$

Since state j is aperiodic it follows that that

$$(P^n)_{j,j} = u_n \to \frac{1}{\sum_{k=1}^{\infty} k f_k}.$$

For arbitrary i we have

$$\begin{aligned} (P^n)_{i,j} &= P(X(n) = j | X(0) = i) \\ &= \sum_{k=1}^{n} P(T_j = k | X(0) = i) P(X(n-k) = j | X(0) = j) \\ &= \sum_{k=1}^{n} P(T_j = k | X(0) = i) u_{n-k}. \end{aligned}$$

Dominated convergence now yields that also $P^n_{i,j} \to \frac{1}{\sum_{k=1}^{\infty} k f_k}$ for arbitrary i. Yet another application of dominated convergence completes the proof. □

D
Bibliography

Søren Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. ISBN 0-387-00211-1. Stochastic Modelling and Applied Probability.

Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. ISBN 0-387-98509-3. Gibbs fields, Monte Carlo simulation, and queues.

Martin Jacobsen and Niels Keiding. *Markovkæder*. Lecture notes. University of Copenhagen, 1985. ISBN 87-87438-37-2.

Gregory F. Lawler. *Introduction to stochastic processes*. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2006. ISBN 978-1-58488-651-8; 1-58488-651-X.

E
Index

absorbing state, 39
absorping state, 70
accesible, 19
aperiodic, 31

backward differential equations, 74
backward integral equation, 74
birth-and death process, 90

Chapman-Kolmogorov equation, 72
characteristic polynomial, 135
closed class, 19
communicate, 19
communication class, 19, 78
conditional probability, 123

determinant, 134
diagonalisation, 135

eigenvalue, 135
eigenvector, 135
escape time, 86
event, 123
explosion, 70
exponential distribution, 124
exponential matrix, 76, 135

forward differential equation, 75

hitting time, 20

initial distribution, 16, 68
integral test, 95, 138
invariant distribution, 34, 80
invariant measure, 34
irreducible, 19, 78

Kolmogorov extension theorem, 127
Kolmogorov's differential equation, 74

linear recurrence equation, 137
loop trick, 52

Markov chain, 15
markov property, 15, 68
minimal construction, 68

null-recurrent, 33

period, 30
Poisson distribution, 123
positive recurrent, 33, 86
probability measure, 123
pure birth process, 71, 90

random variable, 15, 124
random walk, 26
ratio test, 95, 138
recurrence criterion, 22, 28
recurrence time, 20
recurrent, 20, 79
renewal theorem, 32, 153
return time, 20, 86

second order difference equation, 29
state space, 15
stationary distribution, 80
stochastic process, 15, 67
stochastic variable, 124

time-homogeneous, 16
transient, 20, 79
transition diagram, 16
transition intensity, 68, 72
transition probability, 16, 68