David Stenlund

Hitting times in urn models and occupation times in one-dimensional diffusion models

Åbo Akademi University

David Stenlund

Born 1988

Previous studies and degrees

Master of Science, Åbo Akademi University, 2014

Bachelor of Science, Åbo Akademi University, 2012

Hitting times in urn models and occupation times in one-dimensional diffusion models

David Stenlund

Mathematics
Faculty of Science and Engineering
Åbo Akademi University
Åbo, Finland 2020

ISBN 978-952-12-3963-3
ISBN 978-952-12-3954-0 (PDF)

Painosalama Oy
Åbo 2020

Mabinogion

I Mabinogion, en walesisk saga,
beskrivs en uppgift för matematiken:
två flockar får, en svart, en vit, ses draga
i bet längs floden på var sida viken.

De bräkningar som hörs är inte svaga,
för varje ljud får krafter av magiken
och skrämmer herden att ta får av daga.
Det får som bräkt tycks öka flodtrafiken:

ett djur ses lystra, byta färg och sida!
Till slut har alla får fått samma färg.
Hur länge tar det? Där är frågans märg.

Matematiken ger en strategi
som maximerar fårens mängd: såvida
en flock är större, ska det så förbli.

Erik Andersson (1948–2018)

This sonnet was written for a local newspaper (*Åbo Underrättelser*, 26 October 2016) after the author of the thesis won a Three Minute Thesis competition at Åbo Akademi University, during which the research project was presented in 180 seconds.

Preface

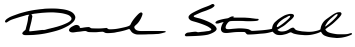
There are many persons who deserve recognition for their share in the work behind this thesis. First of all, I want to express my deepest gratitude to my supervisor Paavo Salminen for guiding me during my first steps as a researcher in mathematics. I appreciate his patience and support, as well as his valuable and detailed feedback during our discussions together.

I wish to thank Sören Christensen for reviewing my thesis and for agreeing to act as opponent during the public defense. I also thank Christophe Vignat for reviewing my thesis and for his many helpful comments. A special thanks goes to James Wan, with whom I had the pleasure of collaborating on one of the articles. The experience taught me much, and it was interesting to work together from very different parts of the world.

I appreciate the friendly and positive work environment at the department of Mathematics and Statistics, and I want to give my thanks to Göran Högnäs, Mikael Lindström and all my other colleagues there. The financial support that I have received from Åbo Akademi University and from the Magnus Ehrnrooth Foundation is gratefully acknowledged.

Finally, I want to thank my family and friends, including my parents who have encouraged my interest in mathematics throughout the years. Most of all I thank my wife Säde, who is ever by my side and whom I can always count on. Her support has been invaluable, especially during the final months of working on the thesis. Last of all I want to thank my dear children Hector and Victoria for all the joy they bring to my life.

Åbo, July 2020



David Stenlund

Sammanfattning på svenska

Den här avhandlingen handlar huvudsakligen om vissa funktionaler av Markovprocesser. Det är fråga om processer som varierar slumpmässigt med tiden, men med den speciella egenskapen att den framtida utvecklingen beror enbart av det nuvarande värdet hos processen och inte på tidigare värden. Denna "minneslösa" egenskap medför att Markovprocesser ofta är en användbar typ av stokastiska processer. De aspekter hos processerna som här undersöks har främst att göra med tiden. Dels undersöks så kallade träfftider, som innebär tiden det tar innan processen uppnår ett visst värde. Dels behandlas vistelsetider, som i sin tur anger hur lång tid som processen befinner sig inom ett visst område. I avhandlingen betraktas speciellt momenten av dessa tidsfunktionaler.

Avhandlingen kan sägas bestå av tre huvudavsnitt. Det första handlar om träfftider i urnmodeller, som är Markovprocesser i diskret tid. I synnerhet behandlas den så kallade Mabinogionmodellen. Den kan beskrivas som att man gör slumpmässiga dragningar ur en låda innehållande svarta och vita kulor, och varje gång lägger man till en kula av den dragna färgen men tar samtidigt bort en av motsatt färg. Antalet svarta respektive vita kulor kommer då att variera med antalet gjorda dragningar, men förr eller senare kommer samtliga kulor att vara av samma färg och då upphör processen. I avhandlingen ges uttryck för den förväntade tiden tills detta absorberande tillstånd uppnås, som en funktion av processens starttillstånd, både när det gäller den vanliga Mabinogionmodellen och när man tillämpar en strategi som löser ett visst optimeringsproblem. Resultaten jämförs också med den förväntade träfftiden i en annan urnmodell, den så kallade Ehrenfestmodellen.

Det andra huvudavsnittet berör vistelsetider för diffusioner, som är Markovprocesser i kontinuerlig tid. Ett typexempel på en diffusionsprocess är det som kallas Brownsk rörelse, som kan användas för att beskriva små partiklars rörelse i en vätska, men som även tillämpas inom många andra områden, exempelvis finansmatematiken. Här behandlas endimensionella diffusionsprocesser, och bland annat ges en rekursiv formel för momenten av vistelsetiden på de positiva reella talen i det fall att diffusionen uppfyller en viss skalningsegenskap, eller för Laplacetransformen av momenten när diffusionen är helt godtycklig. Rekursionsformeln baserar sig på den Greenska kärnan hos diffusionen. Resultaten tillämpas framför allt på skeva Besselprocesser, vilket i det fallet leder till en explicit och relativt enkel momentformel.

I det tredje och sista huvudavsnittet behandlas kombinatoriska summationsformler som anknyter till de övriga resultaten i avhandlingen. Å ena sidan handlar detta om

ändliga dubbelsummor där termerna består av kvoter av binomialkoefficienter, som med hjälp av hypergeometriska funktioner kan skrivas om och i vissa fall förenklas betydligt. Å andra sidan betraktas summor som innehåller Stirlingtal av såväl första som andra slaget, och här ges nya bevis för några formler med sådana uttryck. Bevisen är baserade på den erhållna rekursionsformeln för momenten av den positiva vistelsetiden hos skeva Besselprocesser.

Resultaten presenteras främst i de fyra artiklar som ingår i doktorsavhandlingen. Den inledande delen är ämnad att fungera dels som en introduktion till ämnet, med bakgrundsinformation som kan behövas för att ta del av innehållet i artiklarna, dels även som en sammanfattning och övergripande diskussion.

Abstract

The main subject of this thesis is certain functionals of Markov processes. The thesis can be said to consist of three parts. The first part concerns hitting times in urn models, which are Markov processes in discrete time. In particular, the expected time to absorption in the Mabinogion model is studied. For instance, we give formulas for the expected time to absorption as a function of the initial state of the process, both in the ordinary Mabinogion model and under a strategy that solves an optimal control problem. The second part of the thesis is about occupation times of one-dimensional diffusions, which are continuous Markov processes. We give a recursive formula for the moments of the occupation time on the positive real line, in the case that the diffusion has a self-similar property, or for the Laplace transform of the moments, in case of a general diffusion. The recurrence is based on the Green kernel of the diffusion. In the third part of the thesis, we give results on some combinatorial summation identities that are connected to the other presented results. These include double sums with ratios of binomial coefficients, as well as sums involving Stirling numbers of both first and second kind.

Contents

Preface **v**

Sammanfattning på svenska **vii**

Abstract **ix**

Contents **xi**

1 Introduction **1**
1.1 Structure of the thesis 2
1.2 List of original articles 3
1.3 Contributions by the author 3

2 Preliminaries on stochastic processes **5**
2.1 General Markov processes 5
2.2 Markov chains 8
2.3 Examples of urn models 10
2.4 One-dimensional diffusions 13
2.5 Examples of diffusions 16
2.6 Briefly on martingales 19

3 The Mabinogion urn model **21**
3.1 The Mabinogion sheep problem 21
3.2 Hitting times in the Mabinogion and Ehrenfest models 24

4 Occupation times **27**
4.1 Occupation times for one-dimensional diffusions 27
4.2 Occupation times for some discrete Markov processes 32

5 Combinatorial identities **35**
5.1 Computer-generated proofs of hypergeometric identities 35
5.2 Remarks on identities in the articles 39

Bibliography **43**

Chapter 1

Introduction

Future events always have a degree of uncertainty. Therefore, in mathematical models of real world phenomena, it is often appropriate to use random processes. One important aim in probability theory is to develop the mathematical theory behind such stochastic processes. Although a stochastic process is by definition random, so that future values of the process cannot in general be predicted with absolute certainty, we may still be able to determine the probabilities for different outcomes, and thereby know in advance how the process is expected to behave.

When studying stochastic processes there are many different aspects that might be of significance. Often the value and fluctuations of a stochastic process are relevant features. For instance, given that a certain stochastic process is used to model a financial asset, an investor may want to know what value the process is likely to have after a given amount of time, or the variance of the process, signifying the amount of risk.

However, sometimes the variable of interest rather concerns the time of the process. For instance, how long time will it take before the process reaches a determined level for the first time? This is a so called *hitting time*, and it is important in many financial problems, for instance when finding the optimal time to sell an asset. Another question might be: how long time will the process stay above a given threshold? This is an example of an *occupation time* of the process. Both these concepts are central here and will be explained in more detail later.

This thesis concerns different types of Markov processes. The defining feature of such processes is a so called memorylessness property, which means that future outcomes of the process only depends on the current value and not on how the process has moved in the past. This property makes Markov processes a very useful class of stochastic processes, with applications in many different fields.

The common theme of the thesis could be described as *moments of time functionals in Markov processes*. The title makes this somewhat more specific. Namely, one part of the thesis focuses on hitting times in some urn models, which are examples of discrete Markov processes, and especially the absorption time in the Mabinogion urn model. The other main part of the thesis concerns occupation times, primarily the occupation

time on the positive real numbers for one-dimensional diffusion processes. The main results are expressions for either the first moment or general moments of the hitting time and occupation time distributions.

Although not the main topic of the thesis, another recurring theme is combinatorial summation identities. In fact, two of the included articles arose initially from a few particular identities that are used to derive some of the moment formulas. A discussion of such identities is, therefore, given in a separate chapter.

Functionals of Markov processes is far too wide a topic for a comprehensive treatment of the whole subject. Also, the list of references are but a selected sample from the vast literature on Markov processes. Naturally, the contents of the included articles in the thesis are rather specific, but the following chapters contain a somewhat more general treatment of the subject.

1.1 Structure of the thesis

The thesis consists of an introduction part and four original research articles. Summaries of the articles are given as part of Chapters 3–5, and the full articles are enclosed at the end of the thesis. Chapter 2 provides some background on stochastic processes. First, Markov processes are treated in a general setting. Thereafter, Markov processes in discrete time, called Markov chains, and one-dimensional diffusion processes are discussed separately. In both cases, a number of supplementary examples are given. Finally, martingale processes are briefly introduced.

The presentation in Chapter 3 is focused on the so called Mabinogion urn model. This includes a summary of the results in Article I regarding the absorption time of the Mabinogion process. Chapter 4 is about occupation time functionals for Markov processes. The primary instance concerned is the occupation time on the positive real numbers for one-dimensional diffusions, which is the topic of Article III, but a couple of new results are also given for occupation times in the Mabinogion urn model. The introduction part of the thesis is concluded in Chapter 5 with a discussion on combinatorial identities, including summaries of Articles II and IV, as well as an outline of computerized proofs.

1.2 List of original articles

The following four articles are included in the thesis, and will be referred to using the numbers I–IV.

I Stenlund, D. (2018). On the Mabinogion urn model. *Adv. in Appl. Probab.* **50**, 327–346.

II Stenlund, D. and Wan, J. G. (2019). Some double sums involving ratios of binomial coefficients arising from urn models. *J. Integer Seq.* **22**, Art. 19.1.8, 1–17.

III Salminen, P. and Stenlund, D. (2020). On occupation times of one-dimensional diffusions. *J. Theoret. Probab.* DOI 10.1007/s10959-020-00993-3

IV Stenlund, D. (2020). Some observations on the connection between Stirling numbers and Bessel numbers. Submitted, preprint available at arXiv:2007.11557 [math.CO].

1.3 Contributions by the author

The following comments clarify the contributions of the author of the thesis (DS) to the two co-authored articles.

II The original idea is due to DS. The proof method and some extensions were provided by JW and later refined by DS. Both authors contributed to the development of the results and to the writing of the article.

III The basic idea behind the method is due to PS. Most of the calculations were carried out by DS. Otherwise, both authors contributed jointly to the article.

Chapter 2

Preliminaries on stochastic processes

A *stochastic process* is a mathematical object that describes the development over time of a random phenomenon. To be more precise, a stochastic process $X = (X_t)_{t \in T}$ is a collection of random variables X_t that are defined on the same probability space $(\Omega, \mathcal{F}, P)$ and take values in the same space. The variable t is usually interpreted as the time of the process, and thus the stochastic process is seen as moving randomly over time. If the time set T is countable, the stochastic process "jumps" at discrete time intervals, whereas if T is an interval in $\mathbb{R}$, the transitions happen continuously in time. The processes primarily treated here are Markov processes, either discrete urn models or continuous diffusion models. In this chapter, we define some important concepts regarding these processes.

2.1 General Markov processes

Let $X = (X_t)_{t \in T}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a measurable state space $(E, \mathcal{E})$. Here $\mathcal{E}$ is a σ-algebra on E. The index set T can be defined in a very general way, although in what follows we take T to be either $\mathbb{N} := \{0, 1, 2, \dots\}$ or $\mathbb{R}_+ := [0, \infty)$. Assuming that the process X is in a given state $x \in E$ at time s, there is a certain probability that X will have a value in a given subset $A \subseteq E$ at time t. If all such transition probabilities depend only on the time difference $t - s$, and not on the individual values of s and t, the process is said to be homogeneous in time. Here we will only consider time-homogeneous stochastic processes.

Definition 2.1. *A family of mappings $P_t \colon E \times \mathcal{E} \to [0, 1]$, $t \in T$, is called a* time-homogeneous transition function *if*

(i) the mapping $x \mapsto P_t(x, A)$ is measurable for all $A \in \mathcal{E}$ and $t \in T$,

(ii) the mapping $A \mapsto P_t(x, A)$ is a probability measure on $\mathcal{E}$ for all $x \in E$ and $t \in T$,

(iii) for all $x \in E$ and $A \in \mathcal{E}$,

$$P_0(x, A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

(iv) the Chapman–Kolmogorov *equation*

$$P_{t+s}(x, A) = \int_E P_s(x, \mathrm{d}y) P_t(y, A)$$

holds for all $x \in E$, $A \in \mathcal{E}$ *and* $s, t \in T$.

For a stochastic process, the interpretation of the transition function $P_t(x, A)$ is that it represents the probability that the process moves from state x to a state in the set A in the time t,

$$P_t(x, A) = \mathbf{P}(X_t \in A | X_0 = x).$$

The linear operator P_t generated by the transition function is defined as

$$P_t f(x) := \int_E f(y) P_t(x, \mathrm{d}y),$$

where f is a bounded and measurable function on E. It can be seen that the family $\{P_t : t \in T\}$ is a semi-group, since

$$P_{t+s} = P_t P_s.$$

Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration, that is, a non-decreasing ordering of sub-σ-algebras in $\mathcal{F}$.

Definition 2.2. *The process* X *is said to be* adapted *to the filtration* $(\mathcal{F}_t)_{t \in T}$ *if* X_t *is* $\mathcal{F}_t$*-measurable for every* $t \in T$.

Loosely speaking, a filtration represents the information about the process that is available at time t, and if a process X is adapted to the filtration then we can determine X_t based only on this information. Every stochastic process X is adapted to its natural filtration $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$, which is the smallest σ-algebra generated by $(X_s)_{s \leq t}$. It is a standard assumption that the filtration $(\mathcal{F}_t)_{t \in T}$ of a stochastic process is right-continuous,

$$\mathcal{F}_t = \mathcal{F}_t^+ := \bigcap_{s > t} \mathcal{F}_s,$$

meaning that any information about the process known immediately after t is also known at time t.

Definition 2.3. *An adapted process* X *is called a* time-homogeneous Markov process *with respect to the filtration* $(\mathcal{F}_t)_{t \in T}$ *and with transition function* P_t *if, for any bounded and measurable function* f *and* $s, t \in T$,

$$\mathbf{E}(f(X_{t+s}) | \mathcal{F}_s) = P_t f(X_s), \quad \mathbf{P}\text{-}a.s. \tag{1}$$

The defining property of a Markov process is perhaps best illustrated by considering the function $f = \mathbb{1}_A$ for some $A \in \mathcal{E}$, in which case the identity in (1) becomes

$$\mathbf{P}(X_{t+s} \in A | \mathcal{F}_s) = P_t(X_s, A), \quad \mathbf{P}\text{-a.s.} \tag{2}$$

In other words, at any time s, the probability that the process takes a value in the set A at any fixed point in the future depends only on the current value X_s of the process, and not on any of the past values. All Markov processes are therefore said to be memoryless. This fundamental property makes Markov processes a very useful class of stochastic processes. There is also a stronger version of the Markov property, but for that we first need the notion of stopping times.

Definition 2.4. *A random variable τ that takes values in $T \cup \{\infty\}$ is called a* stopping time *with respect to the filtration $(\mathcal{F}_t)_{t\in T}$ if, for all $t \in T$,*

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

This definition means that τ is a stopping time if we at any time t can determine whether τ has occurred or not, if we know everything that has happened up to t. For instance, a typical example of a stopping time is the first hitting time of a process X on a given value a, defined as

$$H_a := \inf\{t \in T : X_t = a\}. \tag{3}$$

Under very general assumptions on X this is a stopping time, since at any point we can check if the process X has already attained the value a. On the other hand, the time at which the process X attains its maximum value is *not* a stopping time, since at any t we cannot know if the maximum value has already been attained or if X will eventually reach a greater value than the highest so far.

In the following, let $\mathbf{P}_x$ and $\mathbf{E}_x$ denote the probability measure and expectation, respectively, when the process starts at $x \in E$. Let τ be a stopping time with respect to a right-continuous filtration $(\mathcal{F}_t)_{t\in T}$, and define the stopping time σ-algebra

$$\mathcal{F}_\tau := \Big\{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t,\ \forall t \geq 0\Big\},$$

which describes the information of the process known up to and at the stopping time τ.

Definition 2.5 (Strong Markov property)**.** *The stochastic process X is said to have the strong Markov property at a finite stopping time τ if, for any $t \geq 0$,*

$$\mathbf{E}_x\big(f(X_{\tau+t})|\mathcal{F}_\tau\big) = \mathbf{E}_{X_\tau}\big(f(X_t)\big), \quad \mathbf{P}_x\text{-a.s.}$$

The process X is called a strong Markov process if it has the strong Markov property at any finite stopping time τ.

The strong Markov property means that the Markov property holds not only at deterministic times but also at stopping times. As its name suggests, the strong Markov property implies the Markov property, since any deterministic time is clearly a stopping time. However, the reverse is not necessarily true. The strong Markov property is of major importance in the treatment of occupation times of diffusions in Article III, where a diffusion process starting from an arbitrary positive value is restarted at the first hitting time of zero.

2.2 Markov chains

In this section we focus on Markov processes in discrete time.

Definition 2.6. *A Markov process $(X_t)_{t\in T}$ with a countable index set T is called a* Markov chain.

By the term "Markov chain" we hereafter refer to Markov processes where both the state space E and the time index set $T = \mathbb{N}$ are countable, and $\mathcal{E}$ is a discrete σ-algebra on E. Note, however, that the definition also includes Markov processes in discrete time taking values in a general measurable space [36]. Sometimes the term is also used in a wider sense, for instance "continuous-time Markov chain" referring to a Markov process with countable state space E but $T = [0, \infty)$, although such cases are not treated here.

For a Markov chain, the Markov property in (2) can be expressed as the identity

$$\mathbf{P}(X_{n+1} = a_{n+1}|X_0 = a_0, \dots, X_n = a_n) = \mathbf{P}(X_{n+1} = a_{n+1}|X_n = a_n)$$

holding for every $n \in \mathbb{N}$ and all states $\{a_0, a_1 \dots, a_{n+1}\} \subseteq E$. It may be noted that with a discrete state space, the Markov property is equivalent with the strong Markov property, which means that every Markov chain also has the strong Markov property.

For a Markov chain that is time-homogeneous, the transition probabilities can be written as a transition probability matrix P with entries

$$P_{i,j} \coloneqq \mathbf{P}(X_{n+1} = a_j|X_n = a_i).$$

The probabilities of transitioning from one state to another in precisely k steps are subsequently given by the entries in the kth power of the matrix P, that is,

$$(P^k)_{i,j} = \mathbf{P}(X_{n+k} = a_j|X_n = a_i).$$

Especially if the state space is small, a Markov chain can also be described using a graph with directed edges, where every node is a state, and an edge from node i to node j represents a positive transition probability from state i to state j. The values of the transition probabilities are usually indicated as weights on the edges.

Definition 2.7. *A state a_i is called an* absorbing state *if*

$$P_{i,i} = 1, \quad P_{i,j} = 0, \ i \neq j.$$

A Markov chain is called an absorbing Markov chain *if it has at least one absorbing state and it is possible to move from any state to an absorbing state in a finite number of steps.*

Later we will see some examples of absorbing Markov chains. When such a process reaches an absorbing state, it cannot move to any other state after that. The following theorem says that this will eventually happen with probability 1.

Theorem 2.8. *Let X be an absorbing Markov chain with finite state space E. Then, for any initial state, the time until X reaches an absorbing state is almost surely finite.*

Proof. Take any state $a_i \in E$. Since X is an absorbing Markov chain, there exists a set $A \subseteq E$ of absorbing states, at least one of which can be reached from a_i. Let $m_i = \min\{n : \mathbf{P}(X_n \in A | X_0 = a_i) > 0\}$ be the least amount of steps needed to reach one of the absorbing states, and let m be the largest of these numbers for all states $a_i \in E$. Then A can be reached from any state in m steps. Since the number of states in E is finite, it follows that m is also finite. Hence, for any initial state, the probability of absorption in at most m steps is at least

$$p := \min_{a_i \in E} \mathbf{P}(X_m \in A | X_0 = a_i) > 0.$$

Conversely, the probability of X not being absorbed after m steps is less than or equal to $1 - p$, and, by iteration and the Markov property, the corresponding probability after nm steps is less than or equal to $(1 - p)^n$. Since $|1 - p| < 1$, this tends to zero when $n \to \infty$, and therefore the probability of no absorption at all also tends to zero, which means that X reaches an absorbing state almost surely. □

Corollary 2.9. *Let X be a Markov chain with finite state space E and no absorbing states. If a given state $\alpha \in E$ can be reached from every other state in E, then, for any initial state, the first hitting time of α is almost surely finite.*

Proof. If α is the initial state, then the hitting time is 0. If the process X starts in another state we can consider α as being an absorbing state without it affecting the first hitting time from any other state. This new Markov chain is absorbing, with α being the only absorbing state, and the first hitting time of α is then equivalent with the time to absorption. By Theorem 2.8, it follows that the hitting time of α is almost surely finite. □

Next we will have a look at some stochastic processes that are Markov chains. A random walk is a very typical example of such a process.

Definition 2.10. *A stochastic process $(S_n)_{n \in \mathbb{N}}$ is called a* random walk *if*

$$S_n = S_0 + \sum_{i=1}^{n} X_i,$$

where $\{X_n : n \geq 1\}$ are independent and identically distributed random variables. If $S_0 \in \mathbb{Z}$ and the random variables X_i are integer valued, then the process S is a random walk on the integers $\mathbb{Z}$.

From the definition it is clear that any random walk is a Markov chain, since at any point in the process, the next step is independent of all past steps. When $S_0 = 0$ and $\mathbf{P}(X_i = +1) = \mathbf{P}(X_i = -1) = \frac{1}{2}$, the random walk is said to be *simple and symmetric*. This is a standard example of a random walk, and one that has been widely studied. The process is easy to picture – you can think of repeatedly flipping a coin and moving either up or down depending on the coin flip – but it nevertheless has extensive results, not the least since it can be seen as a discrete approximation of Brownian motion (see

Section 2.5). For more background and some famous results on random walks, see Révész [35].

A famous application of a Markov chain model is the PageRank algorithm, which is likely the most famous sorting algorithm used by Google to rank webpages for search engine results [16]. Although not the only ranking algorithm used, PageRank has been a major reason behind the success of the Google search engine. In the PageRank model, the world wide web is seen as an enormous Markov chain, where the states are web pages and the transitions are determined by links to or from other pages. The process is like a person randomly surfing on the web by clicking on links to new pages (but never using the "back"-button – that would violate the memoryless Markov property). In addition, there is a certain probability that the random surfer gets bored and jumps to a completely random page at any point in the process. Then, Markov chain theory ensures that this system has a unique stationary distribution, which describes the long term probability that the random surfer will visit any given web page. This is a measure of the relative importance of the web pages, and the stationary distribution of the Markov chain thus corresponds to the web page ranking.

Other applications of Markov chains include various Markov Chain Monte Carlo (MCMC) methods. These are used for sampling from a probability distribution (usually in high dimensions) by constructing a Markov chain with the target distribution as its stationary limit. After a large number of steps, the relative amount of time that the process spends in different states will then correspond to the target distribution. A number of applications of the MCMC approach, including an interesting example from cryptography, are given by Diaconis [8].

2.3 Examples of urn models

In this section, we will see a few more examples of Markov chains, in the form of *urn models*. The concept of an urn model in probability theory is very old. Urn models are first mentioned in *Ars Conjectandi* (1713) by J. Bernoulli, although some problems that can be expressed in terms of urn models were studied earlier by Huygens. The image of drawing balls from an urn is simply a representation of a discrete uniform distribution on a number of objects. A large amount of results in probability theory can be derived from this rather simple model, as demonstrated by Johnson and Kotz [19]. They show, for instance, that many probability distributions can be thought of in terms of urn models, and list several areas of applications, including models for genetics, population growth and learning curves. See also Kotz and Balakrishnan [26] and Mahmoud [30] for more applications and theory of urn models in general.

In what follows, we let the term "urn model" refer explicitly to a general Pólya urn model. The basic idea of such a model is the following. Imagine an urn containing balls of a number of different colours. At every time step in the process, one ball is drawn uniformly at random from the urn, and is replaced after its colour has been observed. Depending on the colour of the drawn ball, the state space then changes according to some predetermined system, i.e., some balls of certain colours may be

removed from or added to the urn. This change is assumed to depend only on the colour of the drawn ball. Thereafter, the procedure is repeated, meaning that a new ball is drawn from the urn, and so on. The resulting process is a (time-homogeneous) Markov chain, since at any point in the process the transition probabilities correspond to the probabilities of certain colours being drawn in the next step, and clearly these probabilities only depend on the current state of the urn.

Assuming that there are n colours, the system determining the change in composition at each step can be described by an $n \times n$ replacement matrix M, where each entry $M_{i,j}$ is an integer corresponding to the number of balls of colour j that are added to the urn when a ball of colour i is drawn. A negative number naturally means that some balls are removed instead. In this case we take it for granted that the number of removed balls of a certain colour can be at most the number of remaining balls of that colour, so that there are never any negative numbers of balls in the urn.

Here follows a brief presentation of some basic Pólya-type urn models. In all of the following examples, we consider urns with only two colours, so the replacement matrix M can be written

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

For consistency with results presented later, we call the colours white and black. If $a = d = -1$ and $b = c = 0$, then the model simply corresponds to drawing without replacement from the urn.

Example 2.11. The classical example is often called a *Pólya-Eggenberger urn* after the authors of an influential paper on the subject [9]. This is the case when $a = d > 0$ and $b = c = 0$, whereby the replacement matrix becomes

$$M_P = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

In other words, every time a white ball is drawn, that ball and an additional number a of white balls are added to the urn, and similarly when a black ball is drawn, the same number a of black balls are added. Thus, the total number of balls in the urn increases with every step of the process. A typical question regarding this model would be to determine how the proportion of white balls in the urn is distributed after a certain number of steps. The Pólya-Eggenberger urn model has been widely studied and also generalized in various ways. For further reading we refer to Mahmoud [30].

Example 2.12. Another famous urn model is the *Ehrenfest urn*, for which

$$M_E = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

In this model, every time a ball is drawn from the urn it is replaced by a ball of the opposite colour. This urn model differs significantly from the Pólya-Eggenberger urn, since the total number of balls in the urn always stays the same throughout the process. Note that if, after any step, all balls are of the same colour, then one of these has to

change colour in the following step. Hence, the process has two reflecting boundaries represented by the states when all balls are either white or black. Also note that unless the process is stopped at some point it continues on indefinitely without converging to any single state. In fact the Markov chain is recurrent, meaning that every visited state will almost surely be visited again at some point.

The Ehrenfest model was intended to illustrate the second law of thermodynamics. The connection is perhaps clearer if we imagine, rather than two colours of balls in the same urn, that there are two urns and that every time a ball is drawn it is moved to the other urn instead. This can be seen as a crude model for the diffusion of moving particles in two adjacent containers. The particles tend to spread out, and the system will approach an equilibrium with an equal concentration of particles in both containers. In a similar way, the process in the Ehrenfest urn model will always tend towards the symmetric state, in which there is an equal number of white and black balls. In Section 3.2 we will return to the Ehrenfest urn and obtain a formula for the expected time it will take to reach this state if the process starts with balls of only a single colour.

Example 2.13. There are also urn models where the number of balls decreases at every step. The *OK Corral urn* is such a model [22, 46]. If a white ball is drawn, then a black ball is permanently removed from the urn, and vice versa. The name of this model obviously alludes to the famous gunfight at the O.K. Corral in 1881. The urn model is described by the matrix

$$M_{OK} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

and we see that, unlike in the Ehrenfest model, the same state can never be visited again. It is also clear that the process is an absorbing Markov chain, since no more transitions can take place when there are only balls of one colour left in the urn. See Kingman and Volkov [23, 24] for results on the number of remaining balls at absorption.

Example 2.14. The urn model that primarily is studied in this thesis is the *Mabinogion urn model* [45]. This model will be described in more detail in Chapter 3, and for now we only remark that the corresponding replacement matrix is

$$M_M = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

which gives the process a sort of dual nature compared to the Ehrenfest urn. The Mabinogion urn shares some similarities with all of the aforementioned models, i.e., the Pólya-Eggenberger model (the drawn colour increases in number), the Ehrenfest model (the total number of balls stays constant) and the OK Corral model (the opposite colour decreases in number), although it seems that the Mabinogion model has not been as much studied in the literature as its more famous counterparts. The Mabinogion process is an absorbing Markov chain, and the main results in Article I are on the expected number of steps until the process stops by reaching one of the absorbing states.

2.4 One-dimensional diffusions

We now move from discrete Markov chains to processes in continuous time, and give some basic properties of one-dimensional diffusion processes. These processes are complicated to describe in detail, and a rigorous treatment of diffusion theory inevitably demands rather lengthy explanations. Therefore, some simplifications are made in the presentation below, and a few technical details may be omitted for the sake of readability. The references [5, 12, 17, 37, 38] provide further results on diffusion processes in general.

Definition 2.15. *A stochastic process X that takes values on an interval $I \subseteq \mathbb{R}$ is called a* one-dimensional diffusion process *(also called linear diffusion) if it is a time-homogeneous strong Markov process with almost surely continuous sample paths.*

Definition 2.16. *A diffusion process X is called* regular *if*

$$\mathbf{P}_x(H_y < \infty) > 0,$$

for all $x, y \in I$.

Recall from (3) that H_a here denotes the first hitting time of $a \in I$. The regularity property means that from any starting point, the diffusion can reach any other point in the state space in a finite number of steps. We will assume that all diffusion processes treated here are regular diffusions.

Every diffusion process can be described through three characteristics known as the speed measure, scale function and killing measure. The *killing measure* k is related to the distribution of the location of X at the exit time $\zeta := \inf\{t > 0 : X_t \notin I\}$, when the process first leaves the interval I. We will not consider diffusions with killing, so hereafter we assume, for simplicity, that the killing measure $k \equiv 0$, meaning that the diffusion will never leave the interval I on which it lives.

Definition 2.17. *The* scale function *S of the diffusion X is a continuous and strictly increasing function from I to $\mathbb{R}$ such that, for any $x \in (a, b) \subset I$,*

$$\mathbf{P}_x(H_a < H_b) = 1 - \mathbf{P}_x(H_b < H_a) = \frac{S(b) - S(x)}{S(b) - S(a)}.$$

If $S(x) = x$ the diffusion is said to be in natural scale.

For a proof of the existence of S, see Revuz and Yor [37, p. 301]. The scale function describes the drift of the diffusion. For example, if X is in natural scale and starts at some point $X_0 = x$, then it is equally likely that X hits $x + 1$ before it hits $x - 1$ as it is for the other way around. On the other hand, a diffusion process with some other scale function may instead be more likely to go up than to go down. If the diffusion X has the scale function S, then it follows that $(S(X_t))_{t \geq 0}$ is a diffusion in natural scale.

Note that if S is a scale function of X then $\hat{S} = aS + b$, where $a > 0$ and b are constants, is also a scale function. Moreover, it holds that all scale functions of X must

be of the form $\hat{S}$ [37, p. 301], so up to this kind of increasing affine transformation, the scale function is uniquely determined (which explains that it is usually referred to as *the* scale function).

Definition 2.18. *Let $F\colon [a,b] \to \mathbb{R}$ be a strictly increasing function. A real-valued function u is F-*concave *on $[a,b]$ if, for all $a \le l < x < r \le b$,*

$$u(x) \ge \frac{F(r)-F(x)}{F(r)-F(l)}u(l) + \frac{F(x)-F(l)}{F(r)-F(l)}u(r).$$

*A function u is F-*convex *if $-u$ is F-concave. The right and left F-derivatives of u are defined as*

$$\frac{\mathrm{d}^+u}{\mathrm{d}F}(x) := \lim_{h\downarrow 0}\frac{u(x+h)-u(x)}{F(x+h)-F(x)} \quad \textit{and} \quad \frac{\mathrm{d}^-u}{\mathrm{d}F}(x) := \lim_{h\downarrow 0}\frac{u(x-h)-u(x)}{F(x-h)-F(x)},$$

respectively.

The right and left F-derivatives in the above definition exist on (a,b) and are non-increasing functions, and furthermore, if F is continuous, they are right and left continuous, respectively [37, pp. 544–547].

Proposition 2.19. *There exists a unique measure m on the interior of I such that, for any open subinterval (a,b) where $[a,b] \subset I$, it holds for all $x \in (a,b)$ that*

$$\mathbf{E}_x\big(\min\{H_a, H_b\}\big) = \int_a^b G_{a,b}(x,y)m(\mathrm{d}y), \tag{4}$$

where

$$G_{a,b}(x,y) = \begin{cases} \dfrac{(S(x)-S(a))(S(b)-S(y))}{S(b)-S(a)}, & a \le x \le y \le b, \\ \dfrac{(S(y)-S(a))(S(b)-S(x))}{S(b)-S(a)}, & a \le y \le x \le b. \end{cases}$$

Proof. See Revuz and Yor [37, p. 304], or Freedman [12, pp. 126–128] for a diffusion in natural scale. The proof is based on the fact that the left hand side of (4) is a continuous and strictly S-concave function in x on the interval $[a,b]$. □

Definition 2.20. *The measure m in* (4) *is called the* speed measure *of the diffusion X.*

As can be seen from (4), the speed measure is connected to the expected exit time of the process from a given interval. Thereby, the speed measure is an indicator of how fast the process moves, although if the speed measure is large, the diffusion actually moves slower. Rogers and Williams remark that "if the name 'speed' measure were not so well established, we would be tempted to call m the 'sloth' measure" [38, p. 277]. One could thus think of m as a measure of how much time it takes for the process to move through a certain area (or point). For instance, a *sticky point* is a point where the diffusion stays for a positive amount of time before moving on, and this is reflected in the speed measure by an additional term of the Dirac measure in this point. We could also allow m to be infinite, in which case $m(\{x\}) = \infty$ would mean that x is an absorbing point where the diffusion stops altogether. However, note that then the diffusion is no longer regular.

Definition 2.21. *Let $C_b(I)$ be the space of continuous and bounded functions from I to $\mathbb{R}$ equipped with the supremum norm, $\|f\| = \sup_{x\in I} |f(x)|$. The* infinitesimal generator *of the diffusion X is the operator $\mathcal{G}$ given by*

$$\mathcal{G}f \coloneqq \lim_{t\downarrow 0} \frac{P_t f - f}{t},$$

whose domain is all functions $f \in C_b(I)$ for which the limit above exists in $C_b(I)$ and

$$\sup_{t>0} \left\| \frac{P_t f - f}{t} \right\| < \infty.$$

One could say that the infinitesimal generator describes the movement of the diffusion process in an infinitesimal time interval, and this generator thus contains a lot of information about the process. The infinitesimal generator can also be described through

$$\mathcal{G}f(x) = \frac{\mathrm{d}}{\mathrm{d}m} \frac{\mathrm{d}^+}{\mathrm{d}S} f(x).$$

There exists two linearly independent functions ψ_λ and φ_λ that are continuous and positive solutions to the generalized differential equation

$$\mathcal{G}f(x) = \lambda f(x), \tag{5}$$

such that ψ_λ is increasing and φ_λ is decreasing, and both functions are uniquely determined up to a multiplicative constant under certain boundary conditions. These functions are called the *fundamental solutions* to (5). Their Wronskian determinant, with respect to the scale derivative, is given by

$$\begin{aligned} w_\lambda &\coloneqq \frac{\mathrm{d}^+\psi(x)}{\mathrm{d}S}\varphi(x) - \varphi(x)\frac{\mathrm{d}^+\varphi(x)}{\mathrm{d}S} \\ &= \frac{\mathrm{d}^-\psi(x)}{\mathrm{d}S}\varphi(x) - \varphi(x)\frac{\mathrm{d}^-\varphi(x)}{\mathrm{d}S}, \end{aligned} \tag{6}$$

and depends only on λ and not on x. The fundamental solutions determine the Laplace transform of the first hitting time H_a of a point $a \in I$, namely

$$\mathbf{E}_x\left(\mathrm{e}^{-\lambda H_a}\right) = \begin{cases} \dfrac{\psi_\lambda(x)}{\psi_\lambda(a)}, & x \le a, \\ \dfrac{\varphi_\lambda(x)}{\varphi_\lambda(a)}, & x \ge a, \end{cases}$$

see Itô and McKean [17, p. 128] or Jeanblanc, Yor and Chesney [18, p. 278].

For any $t \ge 0$, $x \in I$ and $A \in \mathcal{B}(I)$, it holds that

$$P_t(x, A) = \int_A p(t; x, y) m(\mathrm{d}y),$$

where the transition density p is symmetric, $p(t; x, y) = p(t; y, x)$, and jointly continuous in all variables (Itô and McKean [17, p. 149]).

Definition 2.22. *The* Green kernel *(or Green's function)* G_λ *is defined as*

$$G_\lambda(x, y) := \int_0^\infty \mathrm{e}^{-\lambda t} p(t; x, y)\, \mathrm{d}t, \tag{7}$$

where p *is the transition density with respect to the speed measure* m*.*

Thus, the Green kernel is the Laplace transform of the transition density $p(t; x, y)$ with respect to the time variable t. It can also be expressed as

$$G_\lambda(x, y) = \begin{cases} w_\lambda^{-1} \psi_\lambda(x) \varphi_\lambda(y), & x \le y, \\ w_\lambda^{-1} \psi_\lambda(y) \varphi_\lambda(x), & x \ge y, \end{cases}$$

where ψ_λ and φ_λ are the fundamental solutions to (5) and w_λ is the Wronskian constant as defined in (6), so the functions ψ_λ and φ_λ can be determined directly from the Green kernel when this is known.

For a list of characteristic features, including speed measure, scale function, infinitesimal generator and Green kernel, for several common one-dimensional diffusions, see Borodin and Salminen [5, Appendix 1].

2.5 Examples of diffusions

There are numerous applications of diffusion processes in a wide range of fields. In this section are a few examples of well-known one-dimensional diffusions.

Brownian motion

The origin of the term "diffusion" comes from physics and the tendency of particles in a gas or fluid to spread out (diffuse) from an area of higher concentration. In 1827, the botanist Robert Brown observed that microscopic particles in a fluid displayed tiny and seemingly erratic movements [6]. Brown himself could not give an explanation, but he examined particles from many different materials, both organic and inorganic (including a fragment from the Sphinx), and concluded that such a motion is found in all sufficiently small particles suspended in a fluid. This phenomenon is now known as Brownian motion. The explanation for the motion – that it is caused by the collisions of molecules in the fluid – was given by Albert Einstein in 1905, and his paper on the subject gave conclusive arguments for the existence of atoms.

A stochastic process corresponding to Brownian motion was first described by Louis Bachelier in 1900, but the first rigorous mathematical construction was done by Norbert Wiener, who proved the existence of a stochastic process that satisfies all the properties in the following definition.

Definition 2.23. *A stochastic process* $(W_t)_{t \ge 0}$ *is called a* standard Brownian motion *if the following properties hold.*

(i) $W_0 = 0$.

(ii) The process has independent increments; for all $0 \leq t_1 < t_2 < \cdots < t_n$*, the random variables*

$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$$

are independent.

(iii) The increments are stationary and normally distributed; for all $0 \leq t_1 < t_2$*,* $W_{t_2} - W_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$*.*

(iv) The function $t \mapsto W_t$ *is almost surely continuous.*

In order to distinguish the stochastic process from the physical phenomenon also called Brownian motion, the process defined above is sometimes called a Wiener process. Despite the slight risk of confusion, the term Brownian motion will hereafter be used for the stochastic process, since this notation is widely used. A tribute to Wiener is instead given through the notation $(W_t)_{t\geq 0}$ for a standard Brownian motion.

As described in the previous section, a diffusion process is determined by its scale function S and speed measure m. A standard Brownian motion is a regular diffusion on $\mathbb{R}$ with the speed measure and scale function given by

$$m(\mathrm{d}x) = 2\,\mathrm{d}x \quad \text{and} \quad S(x) = x,$$

respectively. Recall that this choice is unique only up to a certain affine transformation. Some authors instead let the speed measure be equal to the Lebesgue measure, but here we prefer, as is a common choice, to have the Brownian motion in natural scale.

Brownian motion can be considered to be the most important stochastic process, due to both its purely mathematical properties and its widespread impact on models in other fields [40, p. 29]. It lies in the intersection of several significant classes of stochastic processes, being both a Markov process and a martingale, as well as a Lévy process and a Gaussian process. Furthermore, Brownian motion is the quintessential diffusion process. In fact, other diffusions can be constructed via a time transformation on a Brownian motion [38, Ch. V, Sect. 7]. Standard Brownian motion is also central in stochastic integrals (Itô calculus) and stochastic differential equations.

Diffusions as solutions to stochastic differential equations

Some diffusions can be expressed as a solution to a stochastic differential equation (SDE) of the form

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

where W is a standard Brownian motion. Here μ is called the drift function and σ the volatility function of the diffusion satisfying some regularity assumptions. This gives an alternative representation of the diffusion process compared to the characteristics of scale function and speed measure. Note that the equation above is for time-homogeneous diffusions, which we focus on here, but it could be written more generally by letting the functions μ and σ depend also on t.

Geometric Brownian motion

Some commonly used applications of one-dimensional diffusions are found in financial mathematics. In particular, the famous Black–Scholes–Merton market model assumes that the stock price of a risky asset follows a diffusion process known as geometric Brownian motion. It was the first widely used model for fair pricing of option contracts, and the model has had a significant influence on the financial option markets. In 1997, Scholes and Merton were awarded the Nobel Prize in Economics for their work (Black had passed away by then).

A geometric Brownian motion $(S_t)_{t\geq 0}$ satisfies the SDE

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t,$$

for constants μ and σ. This is one of few SDEs that can be solved analytically, and the solution is given by

$$S_t = S_0 \exp\Big(\big(\mu - \tfrac{1}{2}\sigma^2\big)t + \sigma W_t\Big).$$

Letting $\nu = \frac{\mu}{\sigma^2} - \frac{1}{2}$, the speed measure of the geometric Brownian motion is

$$m(\mathrm{d}x) = \frac{2}{\sigma^2} x^{2\nu-1}\,\mathrm{d}x,$$

and the scale function is

$$S(x) = \begin{cases} -\frac{x^{-2\nu}}{2\nu}, & \nu \neq 0, \\ \ln(x), & \nu = 0. \end{cases}$$

Geometric Brownian motion is a very widely used model for stock price movements. The assumption that volatility is constant over time is, however, seldom realistic, and therefore the model is often modified so that σ is a function of time and stock price, or of another Brownian motion to account for randomness in the volatility. An advantage of the ordinary geometric Brownian motion is that it allows for relatively easy calculations compared to other more intricate processes.

Bessel processes

Let W be a Brownian motion in n dimensions, $W_t = (W_t^{(1)}, \ldots, W_t^{(n)})$, where the components $W^{(i)}$ are independent Brownian motions in $\mathbb{R}$. Then the distance from the origin, given by the Euclidean norm $\|W_t\|$ in $\mathbb{R}^n$, is a *Bessel process*, which is a one-dimensional diffusion. The Bessel process X satisfies the SDE

$$\mathrm{d}X_t = \frac{n-1}{2X_t}\,\mathrm{d}t + \mathrm{d}W_t.$$

This SDE is reasonable also if n is not an integer, so the dimension parameter n of a Bessel process can be any real number. An alternative and commonly used parameter is $\nu = n/2 - 1$, which is used in Article III. With this parametrization, the speed measure is given by

$$m(\mathrm{d}x) = 2x^{2\nu+1}\,\mathrm{d}x,$$

and the scale function by

$$S(x) = \begin{cases} -\frac{x^{-2\nu}}{2\nu}, & \nu \neq 0, \\ \ln(x), & \nu = 0. \end{cases}$$

The state space of a Bessel process is either $[0, \infty)$ or $(0, \infty)$, depending on ν, but the process can be made two-sided by allowing excursions from 0 to be either on the positive side or the negative side. Furthermore, if the probabilities for the next excursion being positive or negative are not necessarily equal, but some number $\beta \in (0, 1)$, then the process is called a *skew Bessel process*. Note that this only makes sense for $\nu \in (-1, 0)$, since for positive ν (dimension ≥ 2) the Bessel process never hits 0 and is therefore strictly positive at all times, whereas for $\nu \leq -1$ the drift towards zero is so strong that the Bessel process is absorbed immediately when it hits 0.

The occupation time on the positive real line $(0, \infty)$ of skew Bessel processes is studied in Article III. These processes include, as special cases, skew Brownian motion (when $\nu = -1/2$) and standard Brownian motion (when also $\beta = 1/2$).

2.6 Briefly on martingales

Most of this presentation concerns different types of Markov processes, but a few words need to be said about another useful class of stochastic processes, namely that of martingales. The theory of martingales, primarily pioneered by Doob based on some earlier work by Lévy and Ville, has become a very important part of modern probability theory.

The defining feature of a martingale process is that at any given time during the process, the conditional expected value at a future time is always equal to the present value. Therefore, the notion of a martingale can be thought of as a fair game — if the process represents the payoff from some form of betting game, then the martingale property ensures that the player is neither expected to make a profit nor expected to lose, and hence the game is fair. This defining property is given a precise meaning in the definition below.

Definition 2.24. *A stochastic process X on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a* martingale *with respect to the filtration $(\mathcal{F}_t)_{t\in T}$ if*

(i) *X is adapted to $(\mathcal{F}_t)_{t\in T}$,*

(ii) $\mathbf{E}(|X_t|) < \infty, \quad \forall t \in T,$

(iii) $\mathbf{E}(X_t|\mathcal{F}_s) = X_s, \quad a.s. \quad \forall s, t \in T, s \leq t.$

If the last identity holds with inequality $\leq$ ($\geq$) instead, then X is called a supermartingale (submartingale) with respect to $(\mathcal{F}_t)_{t\in T}$.

Note that a martingale is always defined with respect to a given filtration. The process may not be a martingale when another filtration is considered. The martingale property is not connected to the Markov property described in Definition 2.3, so there

are some martingales that are also Markov processes, and there are some martingales that are not. Naturally, it is the former group that are of interest in this thesis.

A few examples of martingale processes include the simple, symmetric random walk (in discrete time) and standard Brownian motion (in continuous time). In addition, the concept of martingales will be used in Section 3.1 when we consider an optimal control strategy in the Mabinogion urn model.

Theorem 2.25. *Let ξ be any random variable with $\mathbf{E}(|\xi|) < \infty$ and let $(\mathcal{F}_t)_{t\in T}$ be a filtration. Then the stochastic process $(Z_t)_{t\in T}$ defined by*

$$Z_t = \mathbf{E}(\xi|\mathcal{F}_t)$$

is a martingale with respect to $(\mathcal{F}_t)_{t\in T}$.

Proof. (i) Clearly, Z_t is adapted to $(\mathcal{F}_t)_{t\in T}$.
(ii) By the law of total expectation,

$$\mathbf{E}(|Z_t|) = \mathbf{E}(|\mathbf{E}(\xi|\mathcal{F}_t)|) \leq \mathbf{E}(\mathbf{E}(|\xi||\mathcal{F}_t)) = \mathbf{E}(|\xi|) < \infty.$$

(iii) For $s, t \in T$ with $s \leq t$ it follows directly from the tower property of conditional expectation that

$$\mathbf{E}(Z_t|\mathcal{F}_s) = \mathbf{E}(\mathbf{E}(\xi|\mathcal{F}_t)|\mathcal{F}_s) = \mathbf{E}(\xi|\mathcal{F}_s) = Z_s,$$

almost surely. □

A martingale defined as in Theorem 2.25 is sometimes called a "Doob martingale", and is a useful construction of a martingale process. The following theorem on the convergence of supermartingales (or submartingales) was also first given by Doob, and is a central result in martingale theory.

Theorem 2.26 (Martingale convergence theorem)**.** *Let X be a supermartingale, either right-continuous or in discrete time, which is bounded in $\mathcal{L}^1$, i.e., $\sup_t \mathbf{E}(|X_t|) < \infty$. Then the limit $X_\infty := \lim_{t\to\infty} X_t$ exists almost surely and is a.s. finite.*

Proof. We refer to Williams [45, Ch. 11] for a proof of the martingale convergence theorem in discrete time. The same reasoning can be extended to the (right-)continuous case; cf. Theorems (2.2) and (2.10) in Revuz and Yor [37, Ch. II]. □

The next chapter contains an example of how the martingale convergence theorem can be used to show the optimality of a strategy in an optimal control problem.

Chapter 3

The Mabinogion urn model

3.1 The Mabinogion sheep problem

We now take a closer look at the Mabinogion urn model, which is the topic of Article I. Recall from the description in Section 2.3 that this urn model is defined in the following way:

- There is an urn containing balls of two colours (white and black).
- At times $1, 2, 3, \ldots$ one ball is drawn uniformly at random from the urn. The colour of the ball is observed, after which the drawn ball is returned to the urn.
- If all remaining balls are of the same colour, nothing happens (the process has been absorbed).
- Otherwise, if the drawn ball was white, one of the black balls is replaced by a white ball. Conversely, if the drawn ball was black, one of the white balls is replaced by a black ball.

Let W_n and B_n be the number of white balls and black balls, respectively, at time n in the Mabinogion urn model. Then the stochastic process $\{(W_n, B_n) : n = 0, 1, \ldots\}$ is a Markov chain. Moreover, it is an absorbing Markov chain with two absorbing states, namely when all balls are of the same colour, either black or white. Since the state space is finite (the number of states being one more than the total number of balls in the urn), it follows from Theorem 2.8 that the first hitting time of either of the absorbing states,

$$H \coloneqq \min\{n : W_n = 0 \text{ or } B_n = 0\},$$

is almost surely finite. We refer to this as the absorption time of the process.

The Mabinogion urn model has been suggested as a discrete model for transmission of radiation damage [3, pp. 227–230] and as a model for a presidential election campaign between two candidates [11]. The model was first connected to the name "Mabinogion" by David Williams in his book *Probability with Martingales* [45, pp. 159–163], where he considered a certain optimal control problem. The rest of this section is largely based

on his account. Although Williams did not explicitly call it an urn model, it is clear that the system can be described as such.

The Mabinogion is a collection of eleven medieval tales from Wales that were written down in the 12th–13th centuries. These stories thus belong to the oldest prose literature of Britain. Some of these tales are Arthurian legends, and for instance there are many similarities between the story of Peredur and the more well-known tales about knight Percival, one of the knights of the round table. It is within the story of Peredur, son of Evrawc, that the following passage is found [15, pp. 144–145]:

> And he came towards a valley, through which ran a river; and the borders of the valley were wooded, and on each side of the river were level meadows. And on one side of the river he saw a flock of white sheep, and on the other a flock of black sheep. And whenever one of the white sheep bleated, one of the black sheep would cross over and become white; and when one of the black sheep bleated, one of the white sheep would cross over and become black.

This magical herd of sheep is never mentioned again in *The Mabinogion*. They leave the story as abruptly as they had entered it. Nevertheless, this short passage seems to have served as inspiration for the optimal control problem that will be referred to as the Mabinogion sheep problem. The problem goes as follows. Suppose that the process can be adjusted by permanently removing any number of balls from the urn just after time 0 and just after any time $1, 2, \ldots$ when a ball is drawn from the urn. What would be the optimal strategy if the objective is to *maximize the expected final number of black balls*?

It is obvious that no black balls should be removed during the process, so we can assume that only white balls are removed. Consider the following control strategy. At time 0 and after each time a ball has been drawn from the urn:

- if the number of white balls is more than or equal to the number of black balls, immediately reduce the number of white balls to one less than the number of black balls;
- otherwise, do nothing.

This is an optimal strategy, as proved by Williams [45] using a fine example of applied martingale theory. In agreement with his notation, we will henceforth refer to this particular strategy as Policy A. Here is included a sketch of the proof, without going into all the details.

For $w, b \in \mathbb{N}$, let $V(w, b)$ be the expected number of black balls at absorption, given that the process starts with w white and b black balls in the urn; that is,

$$V(w, b) := \mathbf{E}(B_H | W_0 = w, B_0 = b),$$

where H is the absorption time. Correspondingly, let $V^A(w, b)$ be the expected number of black balls at absorption when the control strategy described above is applied

throughout the process. From the description of Policy A it is clear that for any $w, b \in \mathbb{N}$ we have the following identities defining the value function V^A:

(a) $V^A(0,b) = b, \quad V^A(w,0) = 0,$

(b) $V^A(w,b) = V^A(w-1,b), \quad w \geq b > 0,$

(c) $V^A(w,b) = \frac{w}{w+b}V^A(w+1,b-1) + \frac{b}{w+b}V^A(w-1,b+1), \quad 0 < w < b.$

From the recursion in (c) together with the boundary conditions in (a) we can find formulas for the value of $V^A(w,b)$ for any $0 \leq w < b$, and using (b) this is then extended to any $w, b \in \mathbb{N}$. These formulas are given in Proposition 3 in Article I, and are therefore not repeated here. However, it can also be shown that equations (b) and (c) combine into the following result.

Lemma 3.1. *For any $w, b \in \mathbb{Z}^+$,*

$$V^A(w,b) \geq V^A(w-1,b), \tag{8}$$

$$V^A(w,b) \geq \frac{w}{w+b}V^A(w+1,b-1) + \frac{b}{w+b}V^A(w-1,b+1). \tag{9}$$

The proof of the lemma is divided into a few separate cases depending on whether $w < b$ or $w \geq b$ and whether the total number of balls is even or odd. Note that there is always an equality in either (8) or (9) due to equations (b) and (c). The rest is shown based on the explicit formulas for $V^A(w,b)$ and some algebraic manipulations. The details of the proof are omitted here, since the calculations are tedious [41] and an outline of the proof is already given by Williams [45, Section 15.4]. The result in Lemma 3.1 is nonetheless essential for the optimality proof of Policy A, which is discussed below.

Since w and b in the definition of $V^A(w,b)$ are the initial numbers of white and black balls, respectively, the process $V^A(W_n, B_n)$ gives the expected final number of black balls when Policy A is applied throughout the process starting *after* step number n. Assume first that Policy A is applied already from the very beginning of the process, and let $(\mathcal{F}_n^A)_{n\in\mathbb{N}}$ be the natural filtration generated by $(W_n, B_n)_{n\in\mathbb{N}}$ in this case. Then, since the number of black balls at absorption is always finite, $B_H \leq W_0 + B_0 < \infty$, it follows from Theorem 2.25 that $V^A(W_n, B_n)$ is in fact a martingale with respect to $(\mathcal{F}_n^A)_{n\geq 0}$.

On the other hand, the implication of Lemma 3.1 is that at any given point the value function for applying Policy A is never less than if Policy A were applied after first removing one white ball (or, by iteration, any number of white balls) or letting the process run for one step more. Thus, if $(\mathcal{F}_n^S)_{n\in\mathbb{N}}$ is the natural filtration generated by $(W_n, B_n)_{n\in\mathbb{N}}$ when an *arbitrary* strategy S is used throughout the process, then it follows that $V^A(W_n, B_n)$ is instead a *supermartingale* with respect to $(\mathcal{F}_n^S)_{n\in\mathbb{N}}$. Since it is also non-negative, it follows from Theorem 2.26 that the process converges almost surely to the limit $V^A(W_\infty, B_\infty)$, which by definition is the number of black balls at absorption. Thus, for any strategy S,

$$\mathbf{E}(B_H) = \mathbf{E}\big(V^A(W_\infty, B_\infty)\big) \leq V^A(W_0, B_0),$$

since $V^A(W_n, B_n)$ is a supermartingale. This means that the expected final number of black balls using any possible strategy is never more than if Policy A is used throughout, which proves the following desired result.

Theorem 3.2. *Policy A is the optimal strategy for maximizing the expected final number of black balls in the Mabinogion sheep problem.*

Article I includes a section on some other control strategies in the Mabinogion urn model. More precisely, the discussion is focused on control strategies where the proportion of black balls in the urn is always kept above a given threshold q. Note that the special case $q = 1/2$ corresponds to Policy A. Simulations show that Policy A is no longer optimal if a certain discount factor $\mu > 0$ is introduced, so that the value function is $e^{-\mu H} B_H$ and, thus, does not depend only on the final number of black balls B_H but also on the absorption time H. When μ grows, there are strategies with other values of q that are better than Policy A, since the time to absorption is significantly lower. Based on the simulations, the conjecture is that for any q there is a certain value of μ for which that particular strategy is optimal. However, further research is needed to confirm whether this is correct or not.

3.2 Hitting times in the Mabinogion and Ehrenfest models

The focus in Article I is on the Mabinogion urn model, but not primarily on the final number of black balls, but rather the expected time until an absorbing state is reached. This expected time to absorption, when starting the process with w white and b black balls, is defined as

$$T(w, b) := \mathbf{E}(H | W_0 = w, B_0 = b).$$

By solving the recursion

$$T(w, b) = 1 + \frac{w}{w+b} T(w+1, b-1) + \frac{b}{w+b} T(w-1, b+1), \quad w, b > 0,$$

with boundary conditions $T(0, b) = T(w, 0) = 0$ we obtain the explicit formula

$$T(w, b) = \frac{w+b}{2(w+b-1)} \sum_{i=0}^{\min\{w,b\}-1} \sum_{j=i}^{w+b-2-i} \frac{\binom{w+b-1}{i}}{\binom{w+b-2}{j}}, \tag{10}$$

and the special case

$$T(k, k) = k \sum_{i=0}^{k-1} \frac{1}{2i+1}, \tag{11}$$

as given in Proposition 2 in Article I. One of the main objectives in the article is to compare the expected absorption times in the ordinary Mabinogion model and in the controlled version when Policy A is applied. Formulas for the expected absorption time in the latter case are given in Proposition 4.

A precise asymptotic limit is obtained for both the uncontrolled and the controlled model (see Corollary 1 and Theorem 1, respectively), when the process starts with an

equal number of white and black balls and this number tends to infinity. A comparison of the results in this symmetric case leads to the conclusion that although the expected time to absorption is greater in the uncontrolled process, that is, $T(k,k) > T^A(k,k)$, the asymptotic growth is of the same order of magnitude,

$$\lim_{k\to\infty} T(k,k)/T^A(k,k) = 1.$$

If the initial proportion of black balls is strictly greater than $1/2$, then the expected time to absorption is also of the same order of magnitude in both the controlled and uncontrolled model, but the asymptotic growth is radically different. This result, given in Proposition 5, is in part based on Flajolet and Huillet [11]. Some remarks are also made on the uncontrolled process conditioned on the event that there are only black balls left at the end; see Propositions 6 and 7.

The expected time to absorption in the (uncontrolled) Mabinogion process can be compared to similar results for first hitting times in the related Ehrenfest urn model. In fact, the distribution of the absorption time was obtained by Flajolet and Huillet [11, Theorem 1] by applying a time-reversal transformation on the Ehrenfest urn model. Here we concentrate on the first moment and derive a formula for the expected first hitting time of a given state in the Ehrenfest urn, which is similar to (10).

The Ehrenfest urn has a reverse mechanism compared to the Mabinogion urn. The process has no absorbing states, but instead of an absorption time we can find the first hitting time of a given state, say, when there are α white balls in the urn. Let $H_\alpha := \min\{n : W_n = \alpha\}$ and let the expected value of this hitting time be

$$H_\alpha(w,b) := \mathbf{E}(H_\alpha | W_0 = w, B_0 = b).$$

Assuming that $0 \le \alpha \le w+b$ so the state is attainable, this hitting time is almost surely finite, according to Corollary 2.9, and can be found by solving the recursion

$$H_\alpha(w,b) = 1 + \frac{b}{w+b} H_\alpha(w+1,b-1) + \frac{w}{w+b} H_\alpha(w-1,b+1),$$

for $0 < w < \alpha$ with boundary value $H_\alpha(\alpha,b) = 0$ and the reflecting boundary condition $H_\alpha(0,b) = 1 + H_\alpha(1,b-1)$. The following lemma gives a general solution to such a recursion (cf. Lemma 1 in Article I for recursions with two known boundary values).

Lemma 3.3. *Assume that $X(k), k \in \mathbb{Z}$ satisfies the following recursion for all $a \le k < b$:*

$$X(k) = p(k)X(k-1) + (1-p(k))X(k+1) + r(k), \tag{12}$$

where $p(a) = 0$ and $p(k) \in (0,1)$ for all other k. Then, for any $a \le n \le b$,

$$X(n) = X(b) + \sum_{i=n}^{b-1} \sum_{j=0}^{i-a} \frac{r(a+j)}{1-p(a+j)} \prod_{m=j+1}^{i-a} \frac{p(a+m)}{1-p(a+m)}. \tag{13}$$

The proof of the lemma is omitted here, but follows a similar method as in the proof of Proposition 6 in Article I. Using this result it is easy to confirm that the expected time until there are α white balls in the urn is given by the following formula.

Proposition 3.4. *For any $w, b, \alpha \in \mathbb{N}$ with $w \leq \alpha \leq w + b$,*

$$H_\alpha(w, b) = \sum_{j=w}^{\alpha-1} \sum_{i=0}^{j} \frac{\binom{w+b}{i}}{\binom{w+b-1}{j}}. \tag{14}$$

In particular, for any $k \in \mathbb{N}$,

$$H_k(0, 2k) = k \sum_{i=0}^{k-1} \frac{1}{2i+1}. \tag{15}$$

Proof. The result follows from Lemma 3.3. Let $X(k) = H_\alpha(k, N-k)$ where $N = w + b$. Then (12) holds for all $0 \leq k < \alpha$ with $p(k) = k/N$ and $r(k) = 1$. Inserting this and $X(\alpha) = 0$ into (13) gives (14) after simplification. The special case when $\alpha = k$, $w = 0$ and $b = 2k$ follows directly using the first identity in Lemma 2 in Article 1. □

Remark 3.5. *The formula in* (14) *holds when $w \leq \alpha$, but if $w > \alpha$ we can use the symmetry of the Ehrenfest urn to conclude that $H_\alpha(w, b) = H_{w+b-\alpha}(b, w)$, for which* (14) *applies.*

The result in (14) was also found by Palacios [32]. The symmetric case was shown by Blom [4] using an integral representation of $H_\alpha(0, n)$. These results are now directly connected through Lemma 2 in Article 1, or the more general identities presented in Article II. Letting $\alpha = w + 1$ in Proposition 3.4 immediately gives the expected transition time to the state with one more white ball than initially,

$$H_{w+1}(w, b) = \frac{1}{\binom{w+b-1}{w}} \sum_{i=0}^{w} \binom{w+b}{i},$$

which has been derived by Lathrop, Goldstein and Chen [28]. Note that they consider a version of the Ehrenfest model where the colour change at each step is not definite but has a fixed probability p, but this is easily included in (14) by multiplying the right hand side by the factor $1/p$.

It is a very interesting fact that $H_k(0, 2k) = T(k, k)$, as can be seen from equations (11) and (15). In other words, the expected time to absorption in the Mabinogion urn with $2k$ balls, when starting with an equal number of white and black balls, is equal to the expected time to move from 0 to k white balls in the Ehrenfest urn with $2k$ balls. An approximate similarity between these values could be expected, due to the inverse nature of the urn models, but it is not immediately clear why these values are precisely equal. For instance, consider that in this case the Mabinogion process can fluctuate around the symmetric starting point before going off to either of the absorbing states, whereas the Ehrenfest process is unlikely to fluctuate near the very asymmetric starting point and is stopped as soon as it reaches the symmetric state. Note also that the equality holds only in this special case; the expected time to absorption in the Mabinogion urn starting with m white balls is namely not in general equal to the expected time to move from 0 to m white balls in the Ehrenfest urn, except when this m is precisely half the total number of balls.

Chapter 4

Occupation times

The occupation time on a set S is the amount of time that a stochastic process X spends in S up to a given time t, which may be either a random or a deterministic time. Applications are found in financial mathematics, e.g., in the form of occupation time derivatives. These are contingent claims where the payoff depends both on the value of the underlying asset and on an occupation time; see Jeanblanc et al. [18, Sect. 2.5.4]. In this chapter we focus on occupation times of some Markov processes.

4.1 Occupation times for one-dimensional diffusions

In this section, we will restrict ourselves to the topic of Article III, namely positive occupation times of one-dimensional diffusions. For a diffusion X, the positive occupation time A_t^X is defined as

$$A_t^X := \int_0^t \mathbb{1}_{[0,\infty)}(X_s)\,\mathrm{d}s,$$

and, to ease the notation, we call this simply A_t if the diffusion process X is implicit. This is the time that the diffusion spends on the positive real numbers up to time t. Note that the set S is sometimes taken to be $(0,\infty)$ instead, it being a matter of definition whether the point 0 is included or not. However, there is a distinction only in the case when the process has a so called sticky point at 0, which means that the time spent at 0 has a positive value.

A famous example of an occupation time formula is Lévy's arcsine law, which states that the proportion of time that a standard Brownian motion W is positive follows an arcsine distribution. In other words, it holds that

$$\mathbf{P}_0\left(\frac{A_t^W}{t} \le x\right) = \frac{2}{\pi}\arcsin(\sqrt{x}), \quad x \in [0,1]. \tag{16}$$

This rather surprising result was proved by Lévy [29] and has since then been much studied and also generalized for other processes. A particular example is skew Bessel processes, that were first studied by Barlow, Pitman and Yor [2]. The probability

density for the occupation time on $[0,1]$ for a skew Bessel process initiated at 0 was given by Watanabe [43], based on Lamperti [27], and equals, for $x \in (0,1)$,

$$f(x) = \frac{\frac{1}{\pi}\sin(-\nu\pi)\beta(1-\beta)(x(1-x))^{-\nu-1}}{\beta^2(1-x)^{-2\nu} + (1-\beta)^2 x^{-2\nu} + 2\beta(1-\beta)(x(1-x))^{-\nu}\cos(-\nu\pi)}. \tag{17}$$

Additionally, the Laplace transform of the moment generating function of A_t is [43]

$$\int_0^\infty \mathrm{e}^{-\lambda t}\mathbf{E}(\mathrm{e}^{-rA_t})\,\mathrm{d}t = \frac{\beta(\lambda+r)^{-\nu-1} + (1-\beta)\lambda^{-\nu-1}}{\beta(\lambda+r)^{-\nu} + (1-\beta)\lambda^{-\nu}}.$$

It is easily confirmed that when $\nu = -1/2$ and $\beta = 1/2$ the expressions above are equal to those for an arcsine distribution. The distribution in (17) is called a Lamperti-type distribution. The moments of this distribution are difficult to compute from the density, but they are derived recursively in Article III, which results in a surprisingly neat formula in the form of a finite polynomial in the parameters β and ν. As shown in Theorem 4 therein, for any $n \geq 1$,

$$\mathbf{E}_0(A_1^n) = \sum_{k=0}^{n-1}\sum_{j=0}^{k} \frac{(-1)^j j!}{(n-1)!}\begin{bmatrix} n \\ k+1 \end{bmatrix}\begin{Bmatrix} k+1 \\ j+1 \end{Bmatrix}\nu^k\beta^{j+1}, \tag{18}$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ and $\begin{Bmatrix} n \\ k \end{Bmatrix}$ are Stirling numbers of the first and the second kind, respectively.

When considering path functionals for diffusions, including the positive occupation time, the usual method is to use the Feynman–Kac formula. This remarkable theorem provides a link between stochastic differential equations (SDE) on one hand and ordinary partial differential equations (PDE) on the other hand.

Theorem 4.1 (Feynman–Kac)**.** *Assume that a diffusion process $(X_t)_{t\geq 0}$ started at $X_0 = x$ obeys the time-homogeneous stochastic differential equation*

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

where W_t denotes a standard Brownian motion. Let $u\colon \mathbb{R}\times[0,\infty) \to \mathbb{R}$ be a function that satisfies the partial differential equation

$$\mu(x)\frac{\partial u}{\partial x}(x,t) + \frac{1}{2}\sigma(x)^2\frac{\partial^2 u}{\partial x^2}(x,t) - \frac{\partial u}{\partial t}(x,t) - V(x,t)u(x,t) = 0, \tag{19}$$

for $t > 0$ and $x \in \mathbb{R}$, with initial condition

$$u(x,0) = g(x),$$

for some bounded functions $V\colon \mathbb{R}\times[0,\infty) \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$. Then

$$u(x,t) = \mathbf{E}_x\left(g(X_t)\exp\left(-\int_0^t V(X_s, t-s)\,\mathrm{d}s\right)\right). \tag{20}$$

Note that there are several different formulations of the Feynman–Kac formula. In another commonly used version, the initial condition is replaced by the terminal condition

$$u(x, T) = g(x),$$

with $t \in (0, T)$ in the solution. Also, the formula can be stated in a more general form with an additional function $f(x, t)$ on the right hand side of (19) and with an SDE that is not time-homogeneous, so that μ and σ are also functions of t. For a proof of the Feynman–Kac formula, see Karatzas and Shreve [21, p. 366] or Klebaner [25, p. 157].

As an illustration of how the Feynman–Kac formula can be applied, we give a proof for the arcsine law for Brownian motion. Variations of the same proof are found in, e.g., Mörters and Peres [31], Schilling [39] and Steele [40].

Example 4.2. Let $(W_t)_{t\geq 0}$ be a standard Brownian motion with $W_0 = 0$. To prove the arcsine law in (16), we equivalently show that the density of A_t is equal to

$$f_{A_t}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{2}{\pi}\arcsin\sqrt{\frac{x}{t}}\right) = \frac{1}{\pi\sqrt{x(t-x)}}, \quad x \in [0, t].$$

We use the Feynman–Kac formula to find the Laplace transform of A_t,

$$u(x, t) \coloneqq \mathbf{E}_x(\mathrm{e}^{-\lambda A_t}) = \mathbf{E}_x\left(\exp\left(-\lambda\int_0^t \mathbb{1}_{[0,\infty)}(W_s)\,\mathrm{d}s\right)\right).$$

Comparing this expression to (20), we identify the functions $g(x) \equiv 1$ and $V(x, t) = \lambda\mathbb{1}_{[0,\infty)}(x)$. Note that the function $u(x, t)$ is bounded in t. For a standard Brownian motion the drift function is $\mu(x) \equiv 0$ and the volatility function is $\sigma(x) \equiv 1$, so the PDE in (19) becomes

$$\begin{aligned} &\tfrac{1}{2}\tfrac{\partial^2}{\partial x^2}u(x, t) - \tfrac{\partial}{\partial t}u(x, t) - \lambda u(x, t) = 0, && x > 0,\\ &\tfrac{1}{2}\tfrac{\partial^2}{\partial x^2}u(x, t) - \tfrac{\partial}{\partial t}u(x, t) = 0, && x < 0. \end{aligned}$$

Each of these PDEs can be converted to an ordinary differential equation (ODE) by taking the Laplace transform with respect to t on both sides. Writing

$$w(x) \coloneqq \int_0^\infty \mathrm{e}^{-\gamma t}u(x, t)\,\mathrm{d}t,$$

and noting that

$$\int_0^\infty \mathrm{e}^{-\gamma t}\tfrac{\partial}{\partial t}u(x, t)\,\mathrm{d}t = \gamma w(x) - 1,$$

the resulting ODEs are

$$\begin{aligned} &\tfrac{1}{2}w''(x) - (\gamma + \lambda)w(x) = -1, && x > 0,\\ &\tfrac{1}{2}w''(x) - \gamma w(x) = -1, && x < 0. \end{aligned}$$

Solving these two linear second order differential equations, we find the bounded solution

$$w(x) = \begin{cases} A\mathrm{e}^{-\sqrt{2(\gamma+\lambda)}x} + \frac{1}{\gamma+\lambda}, & x > 0, \\ B\mathrm{e}^{\sqrt{2\lambda}x} + \frac{1}{\gamma}, & x < 0. \end{cases}$$

Since $u(x,t)$ is a solution to (19), it has to be twice differentiable in x, from which follows that both $w(x)$ and $w'(x)$ have to be continuous at the point $x = 0$. Equating the left and right limits at 0 leads to

$$A = \frac{1}{\sqrt{\gamma(\gamma+\lambda)}} - \frac{1}{\gamma+\lambda}, \quad B = \frac{1}{\sqrt{\gamma(\gamma+\lambda)}} - \frac{1}{\gamma}.$$

Thus, we have found the general solution for $w(x)$. However, since we are looking at the positive occupation time when the Brownian motion starts at 0, we are really only interested in knowing $w(0)$, which is

$$w(0) = \frac{1}{\sqrt{\gamma(\gamma+\lambda)}}.$$

Recall that this function is actually a double Laplace transform of A_t. By the uniqueness of the Laplace transform, we can invert the transform to obtain the desired result. From a table of inverse Laplace transforms [10, Eq. (5.3.34)] we find that

$$u(0,t) = \mathrm{e}^{-\lambda t/2} J_0(i\lambda t/2),$$

where $J_\nu(z)$ is a Bessel function of the first kind. Using an integral representation for $J_0(z)$ [1, Eq. (9.1.21)], this becomes

$$u(0,t) = \frac{2}{\pi}\int_0^{\pi/2} \mathrm{e}^{-\lambda t \sin^2\theta}\,\mathrm{d}\theta = \int_0^t \mathrm{e}^{-\lambda y}\frac{1}{\pi\sqrt{y(t-y)}}\,\mathrm{d}y,$$

after the variable change $y = t\sin^2\theta$. Since

$$u(0,t) = \mathbf{E}_0(\mathrm{e}^{-\lambda A_t}) = \int_0^t \mathrm{e}^{-\lambda y} f_{A_t}(y)\,\mathrm{d}y,$$

we see that the density of A_t is precisely that of an arcsine distribution, and the proof is complete.

The above example illustrates the method of using the Feynman–Kac formula to find the distribution of the occupation time for a diffusion. Namely, this is done by solving a parabolic differential equation and thereafter taking the inverse Laplace transform of the result. The approach presented in Article III is instead to recursively find the moments of the distribution of A_t for a general diffusion. Rather than using the Feynman–Kac formula, the central tool is Kac's moment formula for additive functionals [20].

One of the main results in Article III, given in Theorem 1, is a formula for the moment generating function of A_T, where T is an exponentially distributed time

independent of the diffusion X. The formula corresponds to known results in the literature [34, 42, 43] but is of a different form, containing integrals of the Green kernel rather than (scale) derivatives of the fundamental solutions.

The other main result, in Theorem 2, is a recursive formula for the Laplace transforms of the moments $\mathbf{E}_0(A_t^n)$, in case of a general diffusion. If the diffusion is self-similar, meaning that for any $a \geq 0$ there exists $b \geq 0$ such that

$$(X_{at})_{t\geq 0} \stackrel{d}{=} (bX_t)_{t\geq 0}, \tag{21}$$

the result simplifies into a recurrence directly for the moments of A_1. The coefficients in the recurrence are obtained as an integral expression, where the integrand is determined by the Green kernel defined in (7). Let

$$U_n(\lambda) := \frac{\lambda^{n+1}}{n!} \int_0^\infty \mathrm{e}^{-\lambda t} \mathbf{E}_0(A_t^n)\, \mathrm{d}t$$

be the Laplace transform with respect to t of the nth moment of A_t (with a suitable factor). Theorem 2 states that the functions U_n are given recursively by

$$\begin{aligned} U_1(\lambda) &= \lambda \int_{I^+} G_\lambda(0, y) m(\mathrm{d}y), \\ U_n(\lambda) &= U_1(\lambda) + \sum_{k=1}^{n-1} (1 - U_{n-k}(\lambda)))\, D_k(\lambda), \quad n = 2, 3, \ldots, \end{aligned} \tag{22}$$

where the coefficients are

$$D_k(\lambda) := \frac{(-\lambda)^k}{(k-1)!} \int_{I^+} G_\lambda(0, y) \widehat{f}_\lambda^{(k-1)}(y; \lambda) m(\mathrm{d}y).$$

Here $I^+ = I \cap [0, \infty)$, $G_\lambda(x, y)$ is the Green kernel as in (7) and

$$\widehat{f}_\lambda^{(k)}(x; \lambda) = (-1)^k \mathbf{E}_x(H_0^k \mathrm{e}^{-\lambda H_0})$$

is the kth derivative of the Laplace transform of the first hitting time of 0 density. In case the diffusion X is self-similar, so that (21) holds, it follows, for any $\lambda > 0$, that

$$U_n(\lambda) = \mathbf{E}_0(A_1^n),$$

and the result simplifies into a recurrence for the moments of A_1.

The method used to prove the result in Theorem 2 is to apply the strong Markov property to restart the diffusion at the first hitting time of 0. Any transformed moment of A_t for a general (positive) starting point can then be expressed as a sum involving transformed lower moments of A_t with starting point 0, as well as a function describing the first hitting time of 0. When this is inserted into Kac's moment formula, the resulting expression can be solved for the nth moment, which leads to the result.

The result in Theorem 2 is applied on a few one-dimensional diffusions. In particular, this is done for skew Bessel processes, which are self-similar diffusions that

contain (skew) Brownian motion as a special case. Since the Green kernel is known, a recurrence for the moments of A_1 can be obtained using (22). Furthermore, the recurrence is solved, yielding the explicit formula for the moments of A_1 given in (18). The formula takes a yet simpler form for skew Brownian motion, when $\nu = -1/2$, namely

$$\mathbf{E}_0(A_1^n) = \sum_{k=0}^{n-1} \binom{n-1+k}{k} \frac{\beta^{n-k}}{2^{n+k-1}}. \tag{23}$$

A corresponding result holds for oscillating Brownian motion and the process called Brownian spider. Sticky Brownian motion is also discussed in Article III, being an example of a diffusion that does not have self-similarity.

4.2 Occupation times for some discrete Markov processes

Simple random walk

Since standard Brownian motion can be seen as the weak convergence limit of a symmetric random walk, it is perhaps not very surprising that the same result holds also in this discrete case, when the number of steps tends to infinity. To be precise, let S be a random walk starting from 0 where each step has zero mean and finite variance, and define

$$A_n^+ := \big|\{k \in \{1, \dots, n\} : S_k > 0\}\big|$$

as the number of steps up to time n that the random walk has a positive value. Then, for any $x \in [0, 1]$,

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{A_n^+}{n} \le x\right) = \frac{2}{\pi} \arcsin\big(\sqrt{x}\big).$$

For a proof of this fact, see Mörters and Peres [31, Remark 5.30].

Lamperti distribution

The arcsine law for a symmetric random walk was considerably generalized by Lamperti [27], who found the corresponding limiting density of the proportional occupation time for a large class of processes, including all recurrent random walks. The assumption on the process (not necessarily a Markov process) is that the state space is divided into two sets that share precisely one state, which is recurrent and also is the initial state of the process. Then, with A_n being the occupation time on one of the two sets, under certain conditions the limit distribution

$$F(x) = \lim_{n\to\infty} \mathbf{P}(A_n \le xn)$$

exists and has the Lamperti-type density in (17).

Mabinogion urn model

Let us for a moment go back to the Mabinogion urn model and consider the occupation time of the Mabinogion process in a given set. Since the states are not recurrent, the theorem by Lamperti does not apply in this case (although, for an infinite number of steps, the result is reminiscent of some special cases mentioned in the theorem). As the process is almost surely absorbed in a finite number of steps, we rather define the limiting occupation time in this case as the number of steps spent in the set before the process is absorbed.

Here we only consider the first moment of the distribution, i.e., the expected occupation time. In this case the method used in Article I for the expected time to absorption can readily be applied, leading to the new formulas presented in the rest of this section. For matter of simplicity, we let the given set be $\{0, 1, \dots, a\}$ and define the corresponding occupation time up to step n as the number of steps when there are at most a black balls, that is,

$$A_n(a) := \big|\{k \in \{0, \dots, n\} : B_k \leq a\}\big|.$$

For instance, if $a = \lfloor (w+b)/2 \rfloor$ this corresponds to the number of steps up to time n when $B_k \leq W_k$, meaning that the process is not in the optimal range regarding the Mabinogion sheep problem, and thus should be controlled by removing some of the white balls.

The expected value of the occupation time (before absorption) is the function

$$A_a(w, b) := \mathbf{E}(A_{H-1}(a) | W_0 = w, B_0 = b),$$

where H is the absorption time. This can be found in the same way as the expected time to absorption, and we have the following new result.

Proposition 4.3. *For any $w, b, a \in \mathbb{N}$ with $0 < a < w + b$,*

$$A_a(w, b) = \frac{N}{N-1} \left(\frac{\sum_{k=0}^{w-1} \binom{N-1}{k}}{2^{N-1}} \cdot \sum_{j=0}^{a-1} \sum_{i=0}^{j} \frac{\binom{N-1}{i}}{\binom{N-2}{j}} - \sum_{j=b}^{a-1} \sum_{i=0}^{j} \frac{\binom{N-1}{i}}{\binom{N-2}{j}} \right), \tag{24}$$

where $N = w + b$.

Proof. The result is obtained by applying Lemma 1 in Article I. The difference in the recursion compared to that of $T(w, b)$ is simply that rather than having $r(k) = 1$ everywhere, the time is now increased only when k is within the suitable range, so

$$r(k) = \begin{cases} 1, & k \leq a, \\ 0, & k > a. \end{cases}$$

With this change, but otherwise following the proof of Proposition 2 in Article I, the result follows. □

Note that the rightmost double sum in (24) disappears when $b \geq a$. These double sum expressions with ratios of binomial coefficients are of a form given in (17) in Theorem 2, Article II, and can thus be evaluated using a ${}_3F_2$-hypergeometric function. However, as we have seen earlier, the expression simplifies considerably when we have a completely symmetric situation, in this case $a = w = b = k$ for some $k \in \mathbb{Z}_+$. In other words, we start with an equal number of white and black balls and look for the number of steps up to absorption when there are at least as many white balls as black ones.

Corollary 4.4. *For any $k \in \mathbb{Z}_+$,*

$$A_k(k,k) = \frac{2^{2k-2}}{\binom{2k}{k}} + \frac{k}{2}\sum_{i=0}^{k-1}\frac{1}{2i+1}. \tag{25}$$

Proof. The result follows from (24) and the second identity in Lemma 2 in Article I (which corresponds to (25) in Corollary 6 in Article II). □

Comparing the result in (25) to (11), we see that except for the first term the result is precisely half the expected time to absorption in this case. The first term can, in turn, be connected to the time that is spent in the symmetric starting point. This can be seen from the symmetry of the situation, but we can also find a more general expression for the local time at a given level a, defined through $L_n(a) \coloneqq \big|\{k \in \{0,\dots,n\} : B_k = a\}\big|$. The expected local time at a before absorption is

$$L_a(w,b) \coloneqq \mathbf{E}(L_{H-1}(a) | W_0 = w, B_0 = b),$$

for which the following result holds.

Proposition 4.5. *For any $w, b, a \in \mathbb{N}$ with $0 < a < w + b$,*

$$L_a(w,b) = \frac{N}{N-1}\cdot\frac{1}{2^{N-1}\binom{N-2}{a-1}}\sum_{i=\max\{a,b\}}^{N-1}\sum_{j=0}^{\min\{a,b\}-1}\binom{N-1}{i}\binom{N-1}{j}, \tag{26}$$

where $N = w + b$. Furthermore, for any $k \in \mathbb{Z}_+$,

$$L_k(k,k) = \frac{2^{2k-1}}{\binom{2k}{k}}. \tag{27}$$

Proof. Applying Lemma 1 in Article I once more, this time with $r(a) = 1$ and $r(k) = 0$ elsewhere, the result in (26) follows after some algebra. It is then easy to confirm that when $a = w = b = k$ this simplifies into (27). □

Occupation times in the Mabinogion model, or other urn models, could be investigated in more detail, and perhaps with more sophisticated methods. For instance, a future consideration is to determine the distribution of the occupation time in the Mabinogion model, rather than only the first moment. In particular, it would be of interest to compare the occupation time with the absorption time, and see if the ratio between them approaches something similar to a Lamperti distribution.

Chapter 5

Combinatorial identities

The main subject of the thesis is analysis of functionals of Markov processes in discrete or continuous time, as has been described in Chapters 3 and 4. However, there is another recurring theme that—although not intended to be of primary interest—also is central here; namely that of combinatorial identities. Such identities, in particular summations of binomial coefficients, has been a much studied subject, and there are works that list a large number of known identities [7, 13, 14]. In this concluding chapter, we first have a look at methods for proving identities using computer, and then consider specifically the identities that are examined in Articles II and IV.

5.1 Computer-generated proofs of hypergeometric identities

An interesting development in the last decades is the use of computers in mathematics. The computers steadily grow more powerful, and at the same time there has been a tremendous improvement in software and algorithms available. Today, computers are used not only for massive numerical calculations, but also for symbolic calculations, and even for mathematical proofs.

One area that allows for computer-generated proofs is that of combinatorial identities, or, to be more precise, identities with hypergeometric series. These include summations of binomial coefficients, such as the ones considered in Article II. There are rather powerful algorithms that essentially turn the proofs of hypergeometric identities into mechanical calculations, which is precisely where computers shine. Since these methods provide an alternative way of proving identities like the ones that are presented in Article II, a brief description may be appropriate here. More details and background information are given in the book by Petkovšek, Wilf and Zeilberger [33].

Hypergeometric functions

Definition 5.1. *A series* $\sum_{n=0}^{\infty} c_n$ *is called* hypergeometric *if the first term is* $c_0 = 1$ *and the ratio of consecutive terms is a rational function with respect to the summation index* n*, that is,*

$$\frac{c_{n+1}}{c_n} = \frac{P(n)}{Q(n)},$$

for all $n \geq 0$*, where* $P(n)$ *and* $Q(n)$ *are polynomials.*

The ordinary generating function for the sequence $c_0, c_1, \ldots$ of terms in a hypergeometric series is called a hypergeometric function. Writing the polynomials P and Q in completely factored form, the ratio of two consecutive terms in the series is

$$\frac{c_{n+1}}{c_n} = \frac{(n + a_1)(n + a_2) \cdots (n + a_p)}{(n + b_1)(n + b_2) \cdots (n + b_q)(n + 1)} d,$$

where p and q are non-negative integers and d is a constant given by the ratio of the leading term coefficients of P and Q. The factor $(n + 1)$ in the denominator is purely a historical convention. If it is not already a factor in $Q(n)$ a corresponding factor can simply be added to the numerator, without loss of generality. Since the series is, by definition, scaled so that $c_0 = 1$, it follows that the individual terms are of the form

$$c_n = \frac{(a_1)_n (a_2)_n \cdots (a_p)_n}{(b_1)_n (b_2)_n \cdots (b_q)_n n!} d^n,$$

where the Pochhammer symbol $(x)_n$ here denotes the rising factorial,

$$(x)_n = \begin{cases} x(x + 1)(x + 2) \cdots (x + n - 1), & n \geq 1, \\ 1, & n = 0. \end{cases}$$

When forming the generating function, the factor d^n can be absorbed into x^n via the rescaling $z = xd$, so the resulting hypergeometric function can be defined as follows.

Definition 5.2. *A* generalized hypergeometric function ${}_pF_q$ *is a function*

$${}_pF_q(a_1, \ldots, a_p\,;\, b_1, \ldots, b_q\,;\, z) = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \cdots (a_p)_n}{(b_1)_n (b_2)_n \cdots (b_q)_n} \frac{z^n}{n!},$$

where $a_1, \ldots, a_p$*,* $b_1, \ldots, b_q$ *are complex numbers and* z *is a complex variable.*

The function ${}_pF_q$ is well-defined when none of the lower parameters $b_1, \ldots, b_q$ is a negative integer or zero. If any of the parameters $a_1, \ldots, a_p$ is a negative integer or zero, then the series has a finite number of terms and the function reduces to a polynomial. In other cases, the series may or may not be convergent. If $p < q + 1$ then it converges for all z. If $p = q + 1$ the series converges for $|z| < 1$ and diverges for $|z| > 1$. If $p > q + 1$ the series is divergent for all $z \neq 0$.

The term "generalized" distinguishes the function ${}_pF_q$ defined above from the special case ${}_2F_1$, which is especially prevalent and therefore often referred to simply as

the hypergeometric function (or Gaussian hypergeometric function, which removes the ambiguity).

Many well-known functions can be written in terms of a (generalized) hypergeometric function. To name a few examples,

$$\begin{aligned}
\mathrm{e}^x &= {}_1F_1(1;1;x),\\
\ln(1+x) &= x\,{}_2F_1(1,1;2;-x),\\
(1+x)^a &= {}_2F_1(-a,1;1;x),\\
\arcsin(x) &= x\,{}_2F_1\Big(\frac{1}{2},\frac{1}{2};\frac{3}{2};x^2\Big),\\
\operatorname{erf}(x) &= \frac{2x}{\sqrt{\pi}}\,{}_1F_1\Big(\frac{1}{2};\frac{3}{2};-x^2\Big),\\
I_\nu(x) &= \frac{(x/2)^\nu}{\Gamma(\nu+1)}{}_1F_2(1;1,\nu+1;x^2/4),\\
y_n(x) &= {}_3F_1(1,-n,n+1;1;-x/2),
\end{aligned}$$

where $\operatorname{erf}(x)$ is the error function, $I_\nu(x)$ is a modified Bessel function and $y_n(x)$ is a Bessel polynomial. Another example, which is at the core of Article II, is that the sum of the first k elements in a row of Pascal's triangle can be written as a hypergeometric function, namely

$$\sum_{i=0}^{k}\binom{n}{i} = \frac{1}{2}\binom{n}{k}{}_2F_1\big(1,n+1;n+1-k;\tfrac{1}{2}\big).$$

The Wilf-Zeilberger method

Here follows a brief outline of the Wilf-Zeilberger (WZ) method, which can be used to verify a hypergeometric identity of the form

$$\sum_{k=-\infty}^{\infty} f(n,k) = r(n).$$

If $r(n) \neq 0$, the identity is first divided by the right hand side, so that it becomes

$$\sum_{k=-\infty}^{\infty} F(n,k) = 1,$$

where $F(n,k) \coloneqq f(n,k)/r(n)$. Otherwise, if r is zero, $F(n,k) \coloneqq f(n,k)$. Next, a method known as Gosper's algorithm is used to find, if possible, a hypergeometric function $G(n,k)$ such that

$$F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k). \tag{28}$$

Such a pair of functions F and G is called a *WZ pair*. Under the assumption

$$\lim_{k\to\pm\infty} G(n,k) = 0,$$

which holds for instance if the function $G(n,k)$ has compact support in k for each n, it follows, by summing both sides in (28) over all values of k, that $\sum_k F(n,k)$ is a constant in n. Hence, in order to verify the original identity, we only need to check it for one value of n (typically $n = 0$).

The rational function

$$R(n,k) := \frac{G(n,k)}{F(n,k)}$$

is called the *proof certificate* of the original hypergeometric identity. If Gosper's algorithm is able to find $R(n,k)$, the identity can easily be verified by constructing $G(n,k)$ and thereafter checking that (28) holds. This means that the proof can be summed up in the single rational function $R(n,k)$, and hence it is named a certificate of the proof.

Example 5.3. To illustrate the WZ method we consider one of the binomial identities that is found in this thesis, namely

$$\sum_{k=a}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k}\binom{k}{a} = 2^{n-2a-1}\binom{n-a}{a}\frac{n}{n-a}, \tag{29}$$

which is used in the proof of Theorem 2 in Article IV. We assume here that $n \geq 2a > 0$. Note that the summand vanishes for any $k < a$ or $k > \lfloor \frac{n}{2} \rfloor$, so we may take the sum as being over all integers. First we divide the summand by the right hand side and thereby get the function

$$F(n,k) = 2^{-(n-2a-1)}\frac{n-a}{n}\binom{n}{2k}\binom{k}{a}\Big/\binom{n-a}{a}.$$

This is given as input to the WZ program in a suitable computer algebra system, e.g., Mathematica [33, Sec. 7.5]. The program immediately returns the proof certificate,

$$R(n,k) = -\frac{(k-a)(2k-1)}{(n-a)(n-2k+1)},$$

which tells us that $\sum_{k=-\infty}^{\infty} F(n,k)$ is constant in n. If we want to verify this, we can define the function $G(n,k) := R(n,k)F(n,k)$ and check with a straightforward calculation that (28) holds. The only thing that remains of the proof is to show that $F(n,k) = 1$ for a single value of n in the range where F does not vanish, that is, $a \leq k \leq \lfloor \frac{n}{2} \rfloor$. If $n = 2a$, the only possible value for k is $k = a$, and it is trivial to verify that indeed $F(2a,a) = 1$. This proves the identity in (29).

Creative telescoping

The WZ method is useful for verifying hypergeometric identities where the right hand side of the equation is known (or, at least, conjectured). If this is not the case, then the more general Zeilberger's algorithm, also called the method of *creative telescoping*, may provide the answer.

Definition 5.4. *A function $F(n,k)$ is called a* proper hypergeometric term *if it can be written in the form*

$$F(n,k) = P(n,k)\frac{\prod_{i=1}^{N}(a_i n + b_i k + u_i)!}{\prod_{j=1}^{M}(c_j n + d_j k + v_j)!}x^n y^k,$$

where $N, M \in \mathbb{N}$ and $a_i, b_i, c_j, d_j \in \mathbb{Z}$ are constants, $u_i, v_j, x, y \in \mathbb{C}$ and $P(n,k)$ is a polynomial.

Note that, for instance, every term in a ${}_pF_q$-hypergeometric function is a proper hypergeometric term. Zeilberger's algorithm is based on the following result [44].

Theorem 5.5. *Let $F(n,k)$ be a proper hypergeometric term. Then F satisfies a recurrence*

$$\sum_{j=0}^{m} p_j(n)F(n+j,k) = G(n,k+1) - G(n,k), \tag{30}$$

in which $G(n,k)/F(n,k)$ is a rational function of n and k and $p_j(n)$ is a polynomial in n for all $j \in \{0,\ldots,m\}$.

Assuming again that $G(n,k)$ vanishes for sufficiently small and large k, when both sides in (30) are summed over all k, the result is that the sum $S(n) \coloneqq \sum_k F(n,k)$ satisfies the recurrence

$$\sum_{j=0}^{m} p_j(n)S(n+j) = 0. \tag{31}$$

Hence, solving this recurrence gives an expression for $\sum_k F(n,k)$. A very special instance was seen in the WZ method, namely (28) corresponds to $m = 1$, $p_0(n) = -1$ and $p_1(n) = 1$ in (30). In some cases, the recurrence in (31) can be readily solved; for instance, when all $p_j(n)$ are constants we have a linear recurrence relation with constant coefficients. However, in some cases it might be that there is no closed form solution, if we take *closed form* to mean a linear combination of a fixed number (i.e., independent of n) of hypergeometric terms. The question whether this is the case or not is solved by another algorithm called Hyper, developed by Petkovšek [33, Ch. 8]. For any recurrence relation with polynomial coefficients, this algorithm is able to either find the solution in closed form, or else verify that such a solution does not exist. The combination of the creative telescoping method and the algorithm Hyper leads to the remarkable fact that any sum where the summand is proper hypergeometric can be determined in closed form – or else a closed form is proved not to exist – using algorithms that are successfully carried out by a computer program.

5.2 Remarks on identities in the articles

Binomial double sums

Articles II and IV both stem from results in the other two articles. In Chapter 3 we have described the Mabinogion urn model, which is the topic of Article I. In the solution to

the recurrence relation for the expected time to absorption, certain binomial double sums appear. Article II contains a number of identities for such double sums involving ratios of binomial coefficients. The identity at the core of the article is

$$\sum_{j=0}^{n}\sum_{i=0}^{j}\frac{\binom{2n+2}{i}}{\binom{2n+1}{j}} = (n+1)\sum_{k=0}^{n}\frac{1}{2k+1}. \tag{32}$$

A generalized version of the identity above is presented in Theorem 1 in Article II, and some special cases thereof are identified in Corollaries 5 and 6.

The computer algorithms outlined in the previous section provide helpful tools when working with identities of hypergeometric nature. This does not, however, mean that other proof methods are obsolete. For a mere verification, it suffices to run a suitable computer program, but another proof may still give better understanding of a certain identity. The deliberate choice in Article II is therefore to present a more traditional proof method for the identities in question, especially since the method led from (32) to more general identities.

The idea behind the proof is based on hypergeometric functions. The inner sum can be written as a ${}_2F_1$-function and, after a few modifications, the original double sum is obtained as a single (infinite) sum. The difference between the expressions for two consecutive arguments can then be simplified using some hypergeometric transformations, and forming a telescoping sum of the result yields the desired identity.

In Theorem 2 are formulas for when the upper index of the binomial coefficients have n rather than $2n$ and the summand contains an additional factor c^{i-j}. In these cases the double sum is given in terms of hypergeometric functions, but a couple of neat special cases are part of Corollary 5. The general form of the double sums treated in the article can be given as

$$\sum_{j=0}^{n}\sum_{i=0}^{j}\frac{\binom{f_1(n)}{i}}{\binom{f_2(n)}{j}}c^{i-j},$$

where $n \in \mathbb{N}$ and f_1, f_2 are functions of n. Provided that $c \in \mathbb{C}\setminus\{-1,0\}$ and $f_2(n) \geq n$, this sum can be expressed using hypergeometric functions; see Theorem 2 and Remark 3. When $c = -1$ the sum is alternating, and some such cases are considered at the end of Section 4, which contains a discussion of a few special cases of interest.

Stirling number identities

Stirling numbers are two kinds of special numbers that appear in various situations, particularly in combinatorics. The unsigned Stirling number of the first kind $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ is the number of permutations of n objects which contain exactly k cycles. These numbers can also be seen as coefficients of the rising factorial,

$$(x)_n = x(x+1)\cdots(x+n-1) = \sum_{k=0}^{n}\begin{bmatrix} n \\ k \end{bmatrix}x^k.$$

The Stirling number of the second kind $\left\{ {n \atop k} \right\}$ is the number of ways to partition a set of n elements into k nonempty subsets. Both kinds of Stirling numbers appear in the summation identities treated in Article IV.

The motivation behind the article comes directly from the moment formulas obtained in Article III for the positive occupation time of skew Bessel processes and skew Brownian motion. A comparison of the coefficients in the polynomial expressions in (18) and (23) leads to the identity

$$\sum_{i=k}^{n} \begin{bmatrix} n \\ i \end{bmatrix} \left\{ {i \atop k} \right\} (-2)^{n-i} = b(n,k), \tag{33}$$

where $b(n,k)$ are signed Bessel numbers of the first kind. Leaving the framework of occupation times for skew Bessel processes, where the parameter ν is restricted to values in $(-1,0)$, the same method can be used to prove similar identities for other values of ν as well.

A few identities involving sums of both kinds of Stirling numbers are proved in Article IV. Although the results are known from literature, the proof method is new, being based on a particular recurrence relation. In fact, this is the recurrence for moments of the positive occupation time for a skew Bessel process, as obtained in Article III. For some fixed values of one of the parameters, the recurrence equation is solved in two different ways, and equating the coefficients in the resulting polynomial expressions leads to an identity similar to (33). One identity proved this way connects a sum of Stirling numbers to Bessel numbers of the second kind, and a couple of other examples are also included in the article.

Some general remarks

The occurrence of (discrete) combinatorial identities is rather natural when considering the Mabinogion urn model. On the other hand, their presence is somewhat more remarkable when deriving a moment formula for the positive occupation time of skew Bessel processes. The most astonishing part is perhaps that an integral of the Green kernel and a higher-order derivative of the Laplace transform of the first hitting time of zero (see (30) in Article III), both containing modified Bessel functions, simplifies into a single binomial coefficient (see (45) in the same article). This fact also makes the resulting recurrence relation sufficiently simple to allow the solution to be first conjectured and then proved by induction.

Binomial expressions are frequently encountered in the work, and the extensive list of identities compiled by Gould [13] has become a familiar reference while working on the thesis. In particular, terms of the form

$$p_n := \frac{1}{2^{2n}} \binom{2n}{n}$$

show up on several occasions. This number can be interpreted as the probability that a symmetric random walk attains its starting value after $2n$ steps. This provides

a superficial connection to the Mabinogion and Ehrenfest urn models, since both these processes are similar to random walks, although the transition probabilities are spatially non-homogeneous, so the individual steps are not identically distributed. However, due to asymmetry in these processes, the probability of returning after $2n$ steps will almost always be different from p_n. In fact, in the Mabinogion model this probability is always less than p_n, regardless of the choice of starting point and the value of n.

In the Mabinogion process, terms of the form p_n instead occur in expressions for the expected final number of black balls and the expected time to absorption when using Policy A and starting from a symmetric initial state; see Propositions 3 and 4 in Article I. Also, the numbers p_n are present in reciprocal form in the symmetric cases of the formulas for occupation time and local time given in Corollary 4.4 and Proposition 4.5, respectively.

Curiously enough, the number p_n is also equal to the nth moment of an arcsine distribution, which explains why these numbers turn up also in the setting of positive occupation times for diffusions. From the results in Article III, the number p_n can be seen as a special value of the polynomial

$$\sum_{k=0}^{n-1} \binom{n-1+k}{k} \frac{x^{n-k}}{2^{n+k-1}},$$

which, in turn, is a special case of the nth moment of the Lamperti distribution, given by

$$\sum_{k=0}^{n-1} \sum_{j=0}^{k} \frac{(-1)^j j!}{(n-1)!} \begin{bmatrix} n \\ k+1 \end{bmatrix} \begin{Bmatrix} k+1 \\ j+1 \end{Bmatrix} x^{j+1} z^k.$$

The numbers p_n are not in themselves of particular significance here, but they serve as a small example of how certain elements may be found in seemingly very different areas of probability theory, illustrating that the areas are, in fact, rather intertwined.

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (1966). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. Dover Publications, Inc., New York.

[2] Barlow, M., Pitman, J., and Yor, M. (1989). Une extension multidimensionnelle de la loi de l'arc sinus. In: *Séminaire de Probabilités, XXIII*, vol. 1372 of *Lecture Notes in Math.*, pp. 294–314. Springer, Berlin.

[3] Bharucha-Reid, A. T. (1960). *Elements of the theory of Markov processes and their applications*. McGraw-Hill Book Co., Inc., New York-Toronto-London.

[4] Blom, G. (1989). Mean transition times for the Ehrenfest urn model. *Adv. in Appl. Probab.* **21**(2), 479–480.

[5] Borodin, A. N. and Salminen, P. (2015). *Handbook of Brownian motion—facts and formulae*, corr. reprint, second edn. Birkhäuser Verlag, Basel.

[6] Brown, R. (1828). XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine* **4**(21), 161–173.

[7] Comtet, L. (1974). *Advanced combinatorics*, enlarged edn. D. Reidel Publishing Co., Dordrecht.

[8] Diaconis, P. (2009). The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)* **46**(2), 179–205.

[9] Eggenberger, F. and Pólya, G. (1923). Über die statistik verketteter vorgänge. *ZAMM Z. Angew. Math. Mech.* **3**(4), 117–134.

[10] Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1954). *Tables of integral transforms. Vol. I*. McGraw-Hill Book Company, Inc., New York-Toronto-London.

[11] Flajolet, P. and Huillet, T. (2008). Analytic combinatorics of the Mabinogion urn. In: *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pp. 549–571. Assoc. Discrete Math. Theor. Comput. Sci., Nancy.

[12] Freedman, D. (1971). *Brownian motion and diffusion*. Holden-Day, San Francisco, Calif.-Cambridge-Amsterdam.

[13] Gould, H. W. (1972). *Combinatorial identities*. Henry W. Gould, Morgantown, W.Va.

[14] Graham, R. L., Knuth, D. E., and Patashnik, O. (1994). *Concrete mathematics*, second edn. Addison-Wesley Publishing Company, Reading, MA.

[15] Guest, C. E. (1997). *The Mabinogion*. Dover Publications, Mineola, NY.

[16] Ishii, H. and Tempo, R. (2013). Markov chains and ranking problems in web search. In: J. Baillieul and T. Samad (eds.), *Encyclopedia of Systems and Control*, pp. 1–8. Springer, London.

[17] Itô, K. and McKean, H. P., Jr. (1974). *Diffusion processes and their sample paths*. Springer-Verlag, Berlin.

[18] Jeanblanc, M., Yor, M., and Chesney, M. (2009). *Mathematical methods for financial markets*. Springer-Verlag London, Ltd., London.

[19] Johnson, N. L. and Kotz, S. (1977). *Urn models and their application*. John Wiley & Sons, New York-London-Sydney.

[20] Kac, M. (1951). On some connections between probability theory and differential and integral equations. In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pp. 189–215. University of California Press, Berkeley and Los Angeles.

[21] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, second edn. Springer-Verlag, New York.

[22] Kingman, J. F. C. (1999). Martingales in the OK Corral. *Bull. London Math. Soc.* **31**(5), 601–606.

[23] Kingman, J. F. C. and Volkov, S. E. (2003). Solution to the OK Corral model via decoupling of Friedman's urn. *J. Theoret. Probab.* **16**(1), 267–276.

[24] Kingman, J. F. C. and Volkov, S. E. (2019). Correction to: Solution to the OK Corral model via decoupling of Friedman's Urn. *J. Theoret. Probab.* **32**(3), 1614–1615.

[25] Klebaner, F. C. (2012). *Introduction to stochastic calculus with applications*, third edn. Imperial College Press, London.

[26] Kotz, S. and Balakrishnan, N. (1997). Advances in urn models during the past two decades. In: *Advances in combinatorial methods and applications to probability and statistics*, Stat. Ind. Technol., pp. 203–257. Birkhäuser Boston, Boston, MA.

[27] Lamperti, J. (1958). An occupation time theorem for a class of stochastic processes. *Trans. Amer. Math. Soc.* **88**, 380–387.

[28] Lathrop, E. D., Goldstein, I. H., and Chen, Y.-P. (2016). A note on a generalized Ehrenfest urn model: another look at the mean transition times. *J. Appl. Probab.* **53**(2), 630–632.

[29] Lévy, P. (1939). Sur certains processus stochastiques homogènes. *Compositio Math.* **7**, 283–339.

[30] Mahmoud, H. M. (2009). *Pólya urn models*. CRC Press, Boca Raton, FL.

[31] Mörters, P. and Peres, Y. (2010). *Brownian motion*. Cambridge University Press, Cambridge.

[32] Palacios, J. L. (1994). Another look at the Ehrenfest urn via electric networks. *Adv. in Appl. Probab.* **26**(3), 820–824.

[33] Petkovšek, M., Wilf, H. S., and Zeilberger, D. (1996). $A = B$. A K Peters, Ltd., Wellesley, MA.

[34] Pitman, J. and Yor, M. (2003). Hitting, occupation and inverse local times of one-dimensional diffusions: martingale and excursion approaches. *Bernoulli* **9**(1), 1–24.

[35] Révész, P. (2013). *Random walk in random and non-random environments*, third edn. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.

[36] Revuz, D. (1984). *Markov chains*, second edn. North-Holland Publishing Co., Amsterdam.

[37] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, third edn. Springer-Verlag, Berlin.

[38] Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge University Press, Cambridge.

[39] Schilling, R. L. and Partzsch, L. (2012). *Brownian motion*. De Gruyter, Berlin.

[40] Steele, J. M. (2001). *Stochastic calculus and financial applications*. Springer-Verlag, New York.

[41] Stenlund, D. (2016). The Mabinogion sheep problem. Unpublished manuscript.

[42] Truman, A. and Williams, D. (1990). A generalised arc-sine law and Nelson's stochastic mechanics of one-dimensional time-homogeneous diffusions. In: *Diffusion processes and related problems in analysis, Vol. I (Evanston, IL, 1989)*, vol. 22 of *Progr. Probab.*, pp. 117–135. Birkhäuser Boston, Boston, MA.

[43] Watanabe, S. (1995). Generalized arc-sine laws for one-dimensional diffusion processes and random walks. In: *Stochastic analysis (Ithaca, NY, 1993)*, vol. 57 of *Proc. Sympos. Pure Math.*, pp. 157–172. Amer. Math. Soc., Providence, RI.

[44] Wilf, H. S. and Zeilberger, D. (1992). An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities. *Invent. Math.* **108**(3), 575–633.

[45] Williams, D. (1991). *Probability with martingales*. Cambridge University Press, Cambridge.

[46] Williams, D. and McIlroy, P. (1998). The OK Corral and the power of the law (a curious Poisson-kernel formula for a parabolic equation). *Bull. London Math. Soc.* **30**(2), 166–170.

David Stenlund

Hitting times in urn models and occupation times in one-dimensional diffusion models

The main subject of this thesis is certain functionals of Markov processes. The first category is hitting times in discrete urn models, and the expected time to absorption in the Mabinogion model in particular. The other category is occupation times of continuous one-dimensional diffusions. One of the results is a recursive formula for the moments of the occupation time on the positive real line. In addition, there are some results on combinatorial summation identities that are related to the subject.

ISBN 978-952-12-3954-0