Universität Zürich[UZH]

Institute of Mathematics
Universität Zürich

Master Thesis

Martingales in random trees and their applications

Author:
Michael Kaufmann
19-914-415

Supervisor:
Prof. Dr. Jean Bertoin

A thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science in Mathematics
at the
Universität Zürich

January 2025

Abstract

Trees have various applications in computer science where they serve as a fundamental object for data structures and algorithms. They can be used to model population dynamics. Besides, they are studied in graph theory. Pólya urns are used in statistics and to study random trees.
This thesis introduces the reader to random trees, in particular to random binary search trees and random m-ary search trees, both in discrete time, and to their continuous time versions, the Yule tree and the multitype branching process with $m-1$ types.
Additionally, Pólya urns in discrete and continuous time and the random bisection problem are introduced.
We study the asymptotic behaviour of these processes by finding suitable martingales and analysing their asymptotic behaviour. We use analytical methods and results from probability theory, especially about Markov chains and martingales, as well as from linear algebra, in particular the theory of eigenvalues. Moreover, we make use of properties of trees, relations between the processes, connections between martingales and embedding discrete time processes in continuous time.
We elaborate a distributional equation of a limit of a random variable arising in one of these processes and characterize its distribution.
The Galton-Watson process and its relation to random trees are introduced. We analyse the extinction probability of this processes, find a martingale and elaborate the behaviour of its limit.
This thesis provides a comprehensive introduction to these topics, suited for readers with about one year of background in probability theory.

"Can you give me a hint how to solve this?"
-"Yeah, think harder!"

"Pass on what you have learned."

"Es macht Spass a sini Grenze zgo!"

Danksagung

Während der ersten Übungsstunde von Linearer Algebra II fragte ich den Übungsleiter nach einem Tipp für eine Aufgabe, über die ich bereits viel nachgedacht habe. Seine Antwort darauf: "Yeah, think harder!" Ich würde mal sagen, das beschreibt mein Mathematikstudium ziemlich gut. Danke Alberto, dass du mich bereits so früh mit dem Leben als Mathematikstudent vertraut gemacht hast.

Ich möchte mich bei meinem Betreuer, Prof. Dr. Jean Bertoin, dafür bedanken, dass er mir dieses äusserst spannende Thema für meine Masterarbeit ausgesucht hat, für seine Unterstützung und seine Zeit. Auch bin ich sehr froh, während zwei Semestern bei ihm Vorlesungen in Wahrscheinlichkeitstheorie besucht zu haben.
Des Weiteren möchte ich mich bei Dr. Thomas Lehéricy bedanken, dass er sich Zeit für mich genommen hat, um meine Fragen zu beantworten.

Ein grosses Dankeschön geht an all meine Kolleginnen und Kollegen, welche mich während des Studiums und während meiner Masterarbeit unterstützt haben und die mir gezeigt haben, dass es noch ein Leben neben dem Studium gibt.

Zu guter Letzt möchte ich mich bei meinen Eltern für ihre stetige Unterstützung bedanken. Sie ermöglichten es mir, die Ausbildung zu absolvieren, die ich wollte.

Contents

1 Introduction **1**

2 Preliminaries and notations **5**
2.1 Set-theoretical notations and definitions 5
2.2 Algebraic notations and preliminaries 5
2.3 Notations and preliminaries in probability theory 6
2.4 Analytical notations and preliminaries 10

3 Introduction to trees **11**
3.1 Words and alphabets 11
3.2 Ordered trees 11
3.3 Binary trees 12
3.4 Randomness on binary trees 16

4 A theorem about martingales **20**

5 The binary search tree martingale **21**
5.1 The level polynomial 21
5.2 The binary search tree martingale 25

6 The bisection martingale **27**
6.1 Point processes 27
6.2 A martingale in the discrete time branching random walk in $\mathbb{R}$ 29
6.3 The bisection problem and the associated martingale 32

7 The Yule tree martingale **34**
7.1 A martingale in the continuous time branching random walk in $\mathbb{R}$ 34
7.2 The Yule tree martingale 37

8 The asymptotic behaviour of the BST martingale **40**
8.1 First results 40
8.2 A relation between subtrees 41
8.3 The 0-1 law of the BST martingale limit 43
8.4 L^p-boundedness and L^p-convergence 43
8.5 When the martingale limit is zero 46

9 Connections between the binary search tree and Yule tree **48**
9.1 The embedding principle 48
9.2 The martingale connection 50
9.3 The profile of a binary search tree 52

10 Limit connection between the BST and bisection martingales **56**
10.1 The bisection martingale of a binary search tree 56

10.2 Martingale limit connection 58

11 A discrete time martingale in Pólya urn processes 61
11.1 Definition of the process and a martingale 61
11.2 Eigenvectors and eigenvalues 63

12 A discrete time martingale in m-ary search tree processes 64
12.1 Definition of the process and a martingale 64
12.2 Some linear algebra in m-ary search tree processes 68
12.3 The associated Pólya urn process 69

13 Asymptotic behaviour of Pólya urns and m-ary search trees in discrete time part I 70
13.1 The Pólya process 70
13.2 A Jordan basis of linear forms of a Pólya process 73
13.3 Semisimplicity, large and small projections 74
13.4 Asymptotics of a Pólya process 75
13.5 The asymptotics of a large Pólya urn process in discrete time 79
13.6 The asymptotics of a large m-ary search tree process in discrete time . . 81

14 Pólya urns in continuous time 84
14.1 Definition of the process 84
14.2 The embedding principle 85

15 A continuous time multitype branching process with $m-1$ types . . 86
15.1 Definition of the process in continuous and discrete time 86
15.2 The embedding principle 87

16 Martingales in the Pólya urn process and the multitype branching process with $m-1$ types in continuous time 89

17 Asymptotic behaviour of the continuous time Pólya urn process . . 94
17.1 Asymptotics of a large Pólya urn process in continuous time 94
17.2 A martingale connection 96
17.3 Dislocation equations for continuous time Pólya urn processes 97

18 Asymptotic behaviour of the continuous time multitype branching process with $m-1$ types 99
18.1 Asymptotics of a large multitype branching process with $m-1$ types in continuous time 99
18.2 A martingale connection 102

19 Asymptotic behaviour of Pólya urns and m-ary search trees in discrete time part II 103
19.1 Notations 103

19.2 First results . . . 104
19.3 Asymptotics of small discrete time Pólya urn and m-ary search tree processes . . . 108

20 Smoothing transformations . . . 110
20.1 A distributional equation . . . 110
20.2 The smoothing transformation . . . 113
20.3 Support and density . . . 117
20.4 Exponential moments . . . 120

21 Galton-Watson processes . . . 124
21.1 Definition of the process and associated trees . . . 124
21.2 The Galton-Watson martingale . . . 125
21.3 The extinction probability . . . 125

22 Conclusion . . . 132

Appendix . . . 134
I.1 A graph-theoretical introduction to trees . . . 134
I.2 Some properties of the exponential distribution . . . 139
I.3 A result about conditional expectation and equality in law . . . 140
I.4 Some convergence results in probability theory . . . 141
I.5 A property of the continuous time Markov chain . . . 142
I.6 Kolmogorov's existence theorem . . . 143
I.7 Some analytical results . . . 144

List of Figures . . . 146

References . . . 147

1 Introduction

Trees serve as a fundamental object in data science, where they have various applications as data structures and for algorithms. Certain kinds of trees are well-suited to model the dynamics of populations. In this case, randomness in trees comes into play. In probability theory, random trees are particular cases of branching processes and random walks. Moreover, they are standard objects in graph theory and combinatorics.
Pólya urns are branching processes as well as random walks and have applications in statistics and studying random trees.

A frequently used concept in random trees, Pólya urns and, more generally, branching processes is that of a martingale. The definitions of these different processes allow us to find or construct martingales, which are well-known stochastic processes and studied in a variety of ways in probability theory. The idea is then to apply common results in the theory of martingales to these more specific processes.
To name but a few authors and some of their work in this field, who also had an influence to this thesis: Asmussen and Hering ([4]), Athreya and Ney ([6]), Bertoin ([8]), Biggins ([9], [11], [10]), Chauvin ([14], [15], [16], [17], [18], [19], [20], [21]), Jabbour-Hattab ([30]), Janson ([31]), Kingman ([33]), Liu ([17]), Pouyanne ([17], [18], [19], [44]) and Rouault ([8], [20], [21]).

In 2011, Prof. Dr. Brigitte Chauvin taught in a summer school about random trees, Pólya urns and their connections with martingales. This thesis takes the theory of her lectures (see [15]) as an inspiration for the topics. First, the reader is introduced to the theory of trees, in particular to random trees, as well as Pólya urns and branching random walks. Afterwards, we study these random processes. We discover the results stated in [15] and extend the discussed topics. The main focus lies on completeness of the proofs as well as comprehensibility. We attain this by a different structure of the topics, adding more details and explanations as well as mentioning related results to expand the study of these subjects. Additionally, Galton-Watson processes are introduced due to their significance in the theory of random trees, branching processes and random walks.

The main purpose of this thesis is to give a comprehensive introduction to the discussed topics, suited for readers with about one year of background in probability theory.

Outline of the thesis

The thesis is organized as follows.

In section 2, the notations as well as some definitions in mathematics which we require for this thesis are recalled.

Section 3 introduces the reader to trees in a set-theoretical way which is well-suited for the discussions of the stochastic processes afterwards. We see binary search trees and how randomness comes into play.

An essential theorem involving martingales, which simplifies several proofs in the

upcoming topics is stated in section 4.

We then extend in section 5 the theory of random binary search trees by defining the profile of a tree and a specific polynomial suited for studying trees. Several results concerning this polynomial lead to a martingale containing all the informations about the form of the corresponding tree.

After an introduction to point processes in section 6, this notion is used to define the discrete time supercritical branching random walk in $\mathbb{R}$ and find a martingale involved in this process. We describe the random bisection problem and, since it is a particular case of the discrete time branching random walk, construct the bisection martingale.

The notion of point processes is used again in section 7 to define the continuous time supercritical branching random walk in $\mathbb{R}$. Similar to the discrete time case, we find a martingale. We introduce the Yule tree, which is a specific kind of continuous time supercritical branching random walk. Moreover, it can be seen as a continuous time version of a random binary search tree. By choosing the correct parameters, we receive the corresponding Yule tree martingale.

Section 8 deals with the limit of the binary search tree martingale. Properties of subtrees and some analytical results lead to a zero-one law for the martingale limit. Afterwards, we determine where the limit is zero, where it is positive and how the martingale converges.

The results from section 8 are then applied in section 9. First of all, we show how the random binary search tree and the Yule tree are related and that we can consider both processes in the same probability space. This is called the embedding principle and it allows us to connect the two corresponding martingales. As a consequence, we receive a connection between the martingale limits in a second step. The section ends by proving a convergence result about the profile of the binary search tree using the martingale limit connection.

Some properties of subtrees in a Yule tree as well as a splitting formula for its martingale are derived in section 10. By the embedding principle, we receive similar results for random binary search trees. As a consequence, we construct a bisection branching random walk and its corresponding martingale from a random binary search tree. Moreover, we prove that the limit of the bisection and the binary search tree martingale are equal on a specific interval.

In section 11 (resp. 12), we introduce the Pólya urn process (resp. the random m-ary search tree process) in discrete time, with its replacement (resp. composition) matrix and corresponding martingale. Furthermore, we mention the eigenvalues of the matrices as well as suitable eigenvectors for the upcoming results. The definition of a small and large process is given. Additionally, section 12 describes the relation between Pólya urns and random m-ary search trees and the construction of a Pólya urn from a random m-ary search tree.

The relation between Pólya urns and random m-ary search trees allows us to consider both processes at the same time in section 13. We define Pólya processes as a

generalization of Pólya urn processes and elaborate the asymptotic behaviour of large processes. The method relies on the theory of eigenvalues, decomposition into eigenspaces and the Jordan normal form as well as some special kind of polynomials that we call reduced polynomials. The same method to find martingales in Pólya urn processes applies to Pólya processes. With help of another similar theorem, we then prove a statement about the asymptotics of large Pólya processes and about the limits of the corresponding martingales. Constructing Pólya processes from a Pólya urn process resp. a random m-ary search tree process finally leads to the results about the asymptotic behaviour of these processes and the limits of their martingales.

The continuous time Pólya urn process is defined in section 14 and a relation to the discrete time Pólya urn process is elaborated in order to consider both processes in the same probability space (the embedding principle).

In section 15, we introduce the continuous time multitype branching process with $m-1$ types and its discrete time version, which is a slight generalization of a random m-ary search tree process. We establish a relation between the continuous and discrete time process allowing us to consider both processes in the same probability space (the embedding principle).

Due to similar behaviour, we consider in section 16 the two continuous time processes we introduced in the last two sections simultaneously and compute their derivatives. After solving a differential equation, we are able to find martingales in these processes.

The derived martingale of the continuous time Pólya urn process (resp. of the continuous time multitype branching process with $m-1$ types) is then used in section 17 (resp. 18) to elaborate the asymptotic behaviour of the process when it is large. The reduced polynomials from section 13 play a central role as well as projections into eigenspaces of the replacement (resp. composition) matrix and the limit of the martingale. Afterwards, a martingale limit connection between the martingales of the continuous and discrete time processes is established. Additionally, we state and prove a result involving dislocation equations and the limit of the continuous time Pólya urn martingale in section 17.

To elaborate the asymptotic behaviour of the small discrete time Pólya urn and random m-ary search tree process in section 19, we use the developed asymptotic results about the continuous time versions from the last two sections. The main result is a convergence in distribution of these processes to a Gaussian limit. The main methods are again the decomposition into eigenspaces and the embedding principle.

The continuous time multitype branching process with $m-1$ types satisfies certain distributional equations that can be summed up to a fixed point equation when considering the limit of a specific projection of the corresponding martingale. This distributional equation is studied in section 20. Defining two distances on a space of probability distributions leads to two complete metric spaces. The Banach fixed-point theorem allows us to prove in two ways that the distributional equation has a unique solution in this space. Furthermore, we characterize the solution of this distributional equation by showing that it attains a density and other properties. This involves a new

martingale.

In section 21, we introduce Galton-Watson processes and the corresponding Galton-Watson trees. After finding a convergent martingale, we elaborate the extinction probability of these processes using the probability generating function. We state and prove a theorem concerning the limit of the Galton-Watson martingale.

The conclusion in section 22 summarizes what we have seen in this thesis and gives an outlook about further related topics.

An interested reader finds in the appendix an introduction to trees in terms of graph theory and the relation to the set-theoretical definition of trees we used in the thesis. Besides, several well-known results in probability theory and other related topics are stated, which are required for the completeness of the thesis.

2 Preliminaries and notations

We state the notations we use in this thesis and recall some important definitions.

2.1 Set-theoretical notations and definitions

Notation 2.1. The natural numbers without zero are written as $\mathbb{N} = \{1, 2, 3, \dots\}$, whereas the natural numbers including zero are $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$. The integers and rational numbers are denoted by $\mathbb{Z}$ and $\mathbb{Q}$ respectively. We write $\mathbb{R}$ for the real and $\mathbb{C}$ for the complex numbers respectively. We denote by $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$ the set of non-negative real numbers and by $\mathbb{R}_{>0} = \{x \in \mathbb{R} \mid x > 0\}$ the set of positive real numbers.

Notation 2.2. Let S be a set. We denote its cardinality by $\text{card}(S)$ or $|S|$.

Notation 2.3. Let $z = a + bi \in \mathbb{C}$ be a complex number, where $a, b \in \mathbb{R}$ and i is the imaginary unit. We denote by $\Re(z) = a$ the real part of z and by $\Im(z) = b$ its imaginary part. The conjugate of z is $\overline{z} = a - bi$ and we denote by $|z|$ the absolute value of z.

Definition 2.4 ([40])**.** The pair $(S, \leq)$ is called a totally ordered set if S is a set and if $\leq$ is an order relation on S such that for all elements $x, y \in S$, we have either $x \leq y$ or $y \leq x$.

Notation 2.5 ([15])**.** Let $x_1, x_2, \dots, x_n$ be n distinct elements in a totally ordered set $(S, \leq)$. We write $x_{(n)}$ for the largest of these elements according to the order relation, $x_{(n-1)}$ for the second largest element etc. up to the smallest element denoted by $x_{(1)}$, such that we have the following sequence of inequalities:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}.$$

So $x_{(i)}$ denotes the i-th smallest element of the $x_1, \dots, x_n$ according to the order relation.

Notation 2.6 ([16])**.** Let X be a set and let $A \subseteq X$ be a subset. The indicator function is denoted by $\mathbb{1}_A \colon X \to \{0, 1\}$ and defined by

$$\mathbb{1}_A(x) \coloneqq \begin{cases} 1 \text{ if } x \in A; \\ 0 \text{ if } x \notin A. \end{cases}$$

Notation 2.7. Let f be a map, its image is denoted by $\text{Im}(f)$.

2.2 Algebraic notations and preliminaries

Notation 2.8. We denote by $(e_i)_{1 \leq i \leq n}$ the canonical basis of $\mathbb{R}^n$, $n \in \mathbb{N}$, where

$$e_i = (e_{ij})_{1 \leq j \leq n} \in \mathbb{R}^{n \times 1}, \; e_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j; \end{cases} \; i = 1, \dots, n.$$

Notation 2.9. The identity matrix is denoted by I.

Notation 2.10. Let $A \in \mathbb{C}^{n\times n}$ be a matrix. We write A^T for the transpose of A. Let $v \in \mathbb{C}^n$ be an n-dimensional column vector, the corresponding row vector is v^T.

Definition 2.11 ([44])**.** Let $A \in \mathbb{C}^{n\times n}$ be a matrix. The spectrum of A, denoted by $\mathrm{Sp}(A)$, is the set of eigenvalues of A. An eigenvector v of A with eigenvalue λ satisfies $Av = \lambda v$. In the corresponding dual space, there exists a linear form u such that $u(v) = 1$.
If a linear form u satisfies $u \circ A = \lambda u$, we call it an eigenform of A to the eigenvalue λ. We can view u as a row vector, then the composition $u \circ A$ can also be seen as a matrix multiplication uA.

Definition 2.12 ([17])**.** The canonical scalar product for $x, y \in \mathbb{R}^n$ is $\langle x, y\rangle = x^T y$. The canonical scalar product for $x, y \in \mathbb{C}$ is $\langle x, y\rangle = \Re(\overline{x}y) = \Re(x)\Re(y) + \Im(x)\Im(y)$. It follows that $\langle x, y\rangle = \langle x\overline{y}, 1\rangle$.

Definition 2.13 ([35])**.** The symmetric group of degree $n \in \mathbb{N}$ is the group S_n of bijections on the set $\{1, 2, \ldots, n\}$. The elements σ_k, $1 \le k \le n$, of S_n are called the permutations.

2.3 Notations and preliminaries in probability theory

Notation 2.14 ([37])**.** Let X be a random variable in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The probability of an event A is written $\mathbb{P}(A)$. We denote by $\mathbb{E}(X)$ the expectation of X and $\mathrm{Var}(X)$ is its variance. If an event A is said to be holding almost surely, i.e. $\mathbb{P}(A) = 1$, it is denoted by a.s. The characteristic function of a random variable X is $\phi_X(t) = \mathbb{E}(e^{i\langle t,X\rangle})$, where $\langle\cdot,\cdot\rangle$ is the canonical scalar product in $\mathbb{R}$ or $\mathbb{C}$ respectively, depending on whether the random variable X is real- or complex-valued. If several random variables are independent and identically distributed, we denote it by i.i.d.

Notation 2.15. Let X and Y be two random variables. We write $X \overset{\mathcal{L}}{=} Y$ if they have the same distribution. Furthermore, we write $\mathcal{L}(X)$ for the distribution of X.

Definition 2.16 ([37])**.** Let $a, b \in \mathbb{R}$ with $a < b$. A random variable X is uniformly distributed on the integral $[a, b]$, denoted by $X \overset{\mathcal{L}}{=} \mathrm{Uni}([a, b])$, if its density function is

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

Its expectation and variance are given by

$$E(X) = \frac{a+b}{2} \text{ and } \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

respectively.

Definition 2.17 ([16])**.** Let $\lambda \in \mathbb{R}_{>0}$. A random variable X is exponentially distributed with parameter λ, denoted by $X \stackrel{\mathcal{L}}{=} \text{Exp}(\lambda)$, if its density function is

$$f(x) = \lambda \cdot e^{-\lambda x} \cdot \mathbb{1}_{\mathbb{R}_{\geq 0}}(x),\ x \in \mathbb{R}.$$

It can also be characterized by $\mathbb{P}(X > x) = e^{-\lambda x}$ for all $x \in \mathbb{R}_{\geq x}$.
Its mean and variance are given by

$$\mathbb{E}(X) = \frac{1}{\lambda} \text{ and } \text{Var}(X) = \frac{1}{\lambda^2}$$

respectively.

Definition 2.18 ([16])**.** Let $\lambda \in \mathbb{R}_{>0}$. A random variable X is Poisson distributed with parameter λ, denoted by $X \stackrel{\mathcal{L}}{=} \text{Poisson}(\lambda)$, if it takes values in $\mathbb{N}$ and for every $k \in \mathbb{N}$ we have

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Its mean and variance are given by

$$\mathbb{E}(X) = \text{Var}(X) = \lambda.$$

Definition 2.19 ([16])**.** Let $m \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_{\geq 0}$. A random variable X is normally distributed with parameters m and σ^2 if its density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2},\ x \in \mathbb{R}.$$

This distribution is also called Gaussian distribution. We write $X \stackrel{\mathcal{L}}{=} \mathcal{N}(m, \sigma^2)$ if a random variable X is normally distributed with parameters m and σ^2. Its mean and variance are given by

$$E(X) = m \text{ and } \text{Var}(X) = \sigma^2$$

respectively.

Definition 2.20 ([37])**.** Let $\alpha, \lambda \in \mathbb{R}_{>0}$. A random variable X is said to be Gamma(α, λ)-distributed, i.e. $X \stackrel{\mathcal{L}}{=} \text{Gamma}(\alpha, \lambda)$, if its density function is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}_{>0}}(x),\ x \in \mathbb{R},$$

where $\Gamma(\alpha)$ is the Gamma function (see Definition 2.35). If $\alpha = 1$, this is the exponential distribution with parameter λ. The mean and variance are given by

$$E(X) = \frac{\alpha}{\lambda} \text{ and } \text{Var}(X) = \frac{\alpha}{\lambda^2}$$

respectively.

Remark 2.21 ([37]). A computation shows that if X is Gamma(α, λ)-distributed, then its moments are given by

$$\mathbb{E}(X^p) = \frac{\lambda^p \cdot \Gamma(\alpha + p)}{\Gamma(\alpha)} \text{ for } p \in \mathbb{N}.$$

Definition 2.22 ([27]). Let $(\Omega, \mathcal{F})$ be a measurable space. Let $w \in \Omega$. The Dirac measure of w is the measure on $(\Omega, \mathcal{F})$ defined by

$$\delta_{\{w\}}(A) \coloneqq \mathbb{1}_A(w) = \begin{cases} 1 \text{ if } w \in A; \\ 0 \text{ if } w \notin A. \end{cases}$$

Definition 2.23 ([16]). The Laplace transform of a random variable X is defined as $L(X)(t) = \mathbb{E}\left(e^{-\langle t, X\rangle}\right)$, where $\langle \cdot, \cdot \rangle$ is the canonical scalar product in $\mathbb{R}$ or $\mathbb{C}$ respectively, depending on whether the random variable X is real- or complex-valued.

Remark 2.24 ([16]). It follows from the definition that if two random variables X and Y are independent, then the Laplace transform of $X + Y$ is the product of the Laplace transforms of X and Y, i.e. $L_{X+Y} = L_X + L_Y$.

Remark 2.25 ([16]). The Laplace transform of an Exp(λ)-distributed random variable X is given by $\mathbb{E}(e^{-tX}) = \frac{\lambda}{\lambda+1}$.

Definition 2.26 ([37]). Let μ and ν be two probability measures on $\mathbb{R}^d$. The convolution of μ and ν is the measure $\mu * \nu$ on $\mathbb{R}^d$, such that for every non-negative measurable function f on $\mathbb{R}^d$,

$$\int_{\mathbb{R}^d} f(z)\, \mu * \nu(dx) = \int_{R^d} \int_{R^d} f(x + y)\, \mu(dx)\, \nu(dy).$$

Notation 2.27 ([37]). Let X be a random variable in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. The conditional expectation of the random variable X given $\mathcal{G}$ is $\mathbb{E}(X \mid \mathcal{G})$. If Y is another random variable in the same probability space, we write $\mathbb{E}(X \mid Y) = \mathbb{E}(X \mid \sigma(Y))$ for the conditional expectation of X given Y, where we denote by $\sigma(Y)$ the smallest sub-σ-algebra in $\mathcal{F}$ such that Y is measurable.

Definition 2.28 ([37]). Let K be either $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}_{>0}$ or $\mathbb{R}_{\geq 0}$. A filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing sequence of sub-σ-algebras of $\mathcal{F}$. Said differently, $(\mathcal{F}_k)_{k\in K}$, such that $\mathcal{F}_l \subseteq \mathcal{F}_k \subseteq \mathcal{F}$ for all $l \leq k$, is a filtration.
When we consider a stochastic process $(X_k)_{k\in K}$, we define the natural filtration $(\mathcal{F}_k)_{k\in K}$ of this process by $\mathcal{F}_k = \sigma(X_l,\ l \leq k)$.
A stochastic process $(X_k)_{k\in K}$ is said to be adapted to the filtration $(\mathcal{F}_k)_{k\in K}$ if X_k is $\mathcal{F}_k$-measurable for all $k \in K$.

Definition 2.29 ([24]). Let K be either $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}_{>0}$ or $\mathbb{R}_{\geq 0}$. Let $(X_k)_{k\in K}$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(\mathcal{F}_k)_{k\in K}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ such that X_k is $\mathcal{F}_k$-measurable for all $k \in K$. Then, $(X_k)_{k\in K}$ is an $(\mathcal{F}_k)_{k\in K}$-martingale (resp. X_k is an $\mathcal{F}_k$-martingale) if for all $k \in K$

(i) $\mathbb{E}(|X_k|) < \infty$, i.e. $X_k \in L^1(\mathbb{P})$;

(ii) $\mathbb{E}(X_k \mid \mathcal{F}_l) = X_l$ for all $l \leq k$.

If K is countable, we talk about a discrete time martingale, if it is uncountable, we call the process a continuous time martingale.

Definition 2.30 ([16])**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let S be the state space, i.e. for all $n \in \mathbb{N}_0$, X_n is a map from Ω to S and assume that S is countable. This process is said to be a Markov chain (in discrete time), if for all $n \in \mathbb{N}_0$,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \tag{2.1}$$

where $j, i, i_{n-1}, \ldots, i_0 \in S$. Equation 2.1 is referred to as the Markov property. If $X_n = j$, we say that the process is in state j.
The transition probabilities are the numbers $\mathbb{P}(X_1 = j \mid X_0 = i)$. If they do not depend on n, the Markov chain is called homogeneous (or it has stationary transition probabilities). In this case, the transition probabilities satisfy

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

for all $n \in \mathbb{N}_0$ and the transition matrix of the Markov chain is $P = (p_{ij})_{i,j\in S}$. It follows that $\sum_{j\in S} p_{ij} = 1$ for every $i \in S$.

Definition 2.31 ([37])**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov chain. Let S be the state space and let $i \in S$. Let $T_i = \inf\{n \in \mathbb{N} \mid X_n = i\}$. Then, i is said to be recurrent if $\mathbb{P}(T_i < \infty \mid X_0 = i) = 1$. Otherwise, i.e. if $\mathbb{P}(T_i < \infty \mid X_0 = j) < 1$, it is called transient.

Definition 2.32 ([3])**.** Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let S be the state space, i.e. for all $t \in \mathbb{R}_{\geq 0}$, X_t is a map from Ω to S and assume that S is countable. This process is a continuous time Markov chain, if for all $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$, and $j, i, i_{n-1}, \ldots, i_1 \in S$ such that $\mathbb{P}(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1) > 0$, we have

$$\mathbb{P}(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1) = \mathbb{P}(X_{t_{n+1}} = j \mid X_{t_n} = i). \tag{2.2}$$

Equation 2.2 is referred to as the Markov property. If for all $s, t \in \mathbb{R}_{\geq 0}$, $s \leq t$, and all $i, j \in S$, $\mathbb{P}(X_t = j \mid X_s = i)$ only depends on $t - s$ and not on s and t individually, the process is called homogeneous (or it has stationary transition probabilities). In this case, we have

$$\mathbb{P}(X_t = j \mid X_s = i) = \mathbb{P}(X_{t-s} = j \mid X_0 = i)$$

and

$$p_{ij}(t) = \mathbb{P}(X_t = j \mid X_0 = i),\ i, j \in S,\ t \in \mathbb{R}_{\geq 0},$$

is called the transition function of the process.

Definition 2.33 ([3])**.** Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a continuous time Markov chain with state space S. Let $i \in S$ and we assume that $X_0 = i$. The holding time in state i is $T_i = \inf\{t \in \mathbb{R}_{\geq 0} \mid X_t \neq i\}$, with the convention that $\inf(\emptyset) = \infty$.

Remark 2.34 ([3])**.** Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a continuous time Markov chain with state space S. Let $i \in S$ and we assume that $X_0 = i$. Then the holding time T_i in state i is exponentially distributed if i is not an absorbing state (see Lemma I.45)
Let us assume the chain jumps at time T_i to the state j, i.e. $X_{T_i} = j$. The definition and the homogeneity of the process imply that the behaviour of the process after time T_i is

(i) independent of the behaviour of the process before time T_i, and

(ii) as if the time axis is relabelled such that T_i was zero and the chain starts at state j.

If j is not an absorbing state, the chain stays in j for an exponential amount T_j of time, where T_j is independent of T_i, and at time $T_i + T_j$, the chain jumps to a state $X_{T_i+T_j} = k \in S$.
We define

$$\tau_n = \begin{cases} 0 & \text{if } n = 0; \\ \inf\{t > \tau_{n-1} \mid X_t \neq \lim_{s\to t,\ s<t} X_s\} & \text{if } n \in \mathbb{N}. \end{cases}$$

Then τ_1 is the time of the first jump and τ_n is the time of the n-th jump. If $\lim_{n\to\infty} \tau_n = \infty$, we have described the complete behaviour of the Markov chain (all continuous time Markov chains in this thesis satisfy this limit).
The Markov chain $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ is completely determined by the jump times $(\tau_n)_{n\in\mathbb{N}_0}$ (they are exponentially distributed) and the states $(X_{\tau_n})_{n\in\mathbb{N}_0}$ the Markov chain passes through.

2.4 Analytical notations and preliminaries

Definition 2.35 ([28])**.** We denote by $\Gamma(z)$ the Gamma function, which is defined for $x \in \mathbb{R}_{>0}$ by

$$\Gamma(x) \coloneqq \int_0^\infty u^{x-1}e^{-u}\,du.$$

Definition 2.36 ([28])**.** Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be non-zero sequences. We say that these sequences are asymptotically equal and write $a_n \sim b_n$, if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

Notation 2.37 ([28])**.** Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be two functions.
We write $f(x) = o(g(x))$ for $x \to \infty$, if for every $\varepsilon > 0$ there exists a $m \in \mathbb{R}$ such that $|f(x)| \leq \varepsilon|g(x)|$, for all $x \geq m$. If $g(x) \neq 0$ for $x \geq m$, then the above condition is equivalent to

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

We write $f(x) = O(g(x))$ for $x \to \infty$, if there exist constants $C \in R_{>0}$ and $m \in \mathbb{R}$ such that $|f(x)| \leq C|g(x)|$, for all $x \geq m$.

3 Introduction to trees

In this section, we give a set-theoretical introduction to trees in terms of words and alphabets. We consider in particular binary trees as well as binary search trees. Moreover, we develop and study randomness in trees.

An interested reader finds in subsection I.1 a graph-theoretical introduction.

3.1 Words and alphabets

Definition 3.1 ([25])**.** An alphabet is a set which is non-empty and usually finite. Its elements are distinguishable.

Definition 3.2 ([25])**.** Let $\mathcal{A}$ be a finite alphabet. An element $a \in \mathcal{A}$ is said to be a letter. We define a letter to be of size 1.
A word or string is a finite sequence of letters. We usually write a word without separators between the letters. The length of a word is the number of letters that are contained in the word.

Definition 3.3 ([16])**.** The empty word is $\mathcal{E} = \emptyset$.

Definition 3.4 ([7])**.** We define $\mathcal{A}^* = \bigcup_{n\in\mathbb{N}_0} \mathcal{A}^n$ (where $\mathcal{A}^0$ is by convention the set reduced to the empty word $\emptyset$), the set of all words made by an alphabet $\mathcal{A}$.

Definition 3.5 ([25])**.** A subset of $\mathcal{A}^*$ is a called a language.

Definition 3.6 ([16])**.** We define the set

$$\mathbb{U}_{\mathbb{N}_0} \coloneqq \{\mathcal{E}\} \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}_0^n$$

consisting of the words made by the alphabet $\mathbb{N}_0$.
The length of a sequence $u \in \mathbb{U}_{\mathbb{N}_0}$ is $|u|$ and for the empty word we have $|\mathcal{E}| = 0$.
Let u and v be two words in $\mathbb{U}_{\mathbb{N}_0}$. Then we define the sequence uv by concatenation of u with v. If a word u can be written as $u = vw$, then we call v a prefix and w a suffix of u.

3.2 Ordered trees

Definition 3.7 ([16])**.** We say that a subset $\mathcal{T}$ of $\mathbb{U}_{\mathbb{N}_0}$ is an ordered tree if it satisfies the following conditions:

(i) $\mathcal{E} \in \mathcal{T}$;

(ii) $\forall u, v \in \mathbb{U}_{\mathbb{N}_0}$, if $uv \in \mathcal{T}$, then $u \in \mathcal{T}$;

(iii) $\forall u \in \mathbb{U}_{\mathbb{N}_0}$, if $u \in \mathcal{T}$, there exists a $M_u(\mathcal{T}) \in \mathbb{N}_0$ such that for all $j \geq 0$, $uj \in \mathcal{T}$ if and only if $0 \leq j \leq M_u(\mathcal{T}) - 1$.

The element $\mathcal{E}$ is called the root of the tree. The elements in $\mathcal{T}$ are called nodes. The children of a node $u \in \mathcal{T}$ are the elements $u0, \ldots, u(M_u(\mathcal{T}) - 1)$, if they exist, i.e. if $M_u(\mathcal{T}) \geq 1$. The child $u0$ is called the first child of u, $u1$ the second child etc. up to $u(M_u(\mathcal{T}) - 1)$ for the $(M_u(\mathcal{T}))$-th child of u.

Definition 3.8 ([16])**.** Let $\mathcal{T} \subseteq \mathbb{U}_{\mathbb{N}_0}$ be an ordered tree. The path from u to uv in $\mathcal{T}$ for $u, uv \in \mathcal{T}$ is the set P, where $w \in P$ if and only if there exist $k, l \in \mathbb{U}_{\mathbb{N}_0}$ such that $w = uk$ and $wl = uv$.

Remark 3.9 ([16])**.** Condition (i) in Definition 3.7 says that an ordered tree is non-empty. Condition (ii) implies that there exists a path to the root and condition (iii) means that there exists a maximal number of children for every node.

Definition 3.10 ([40])**.** Let $\mathcal{T}$ be an ordered tree. Let $m = \max\{M_u(\mathcal{T}) \mid u \in \mathcal{T}\}$. We say that m is the branching factor of $\mathcal{T}$; it is the largest number of children a node of $\mathcal{T}$ may have. A tree with branching factor $m \geq 3$ is called an m-ary tree.

Definition 3.11 ([16], [21])**.** Let $\mathcal{T}$ be an ordered tree. The empty tree is $\mathcal{T} = \{\mathcal{E}\}$. The number of elements in the tree $\mathcal{T}$ is called the size of $\mathcal{T}$. The tree $\mathcal{T}$ is finite if its size is finite.

Definition 3.12 ([16])**.** Let $\mathcal{T}$ be an ordered tree. We say that the level, the depth or the generation of the node $u \in \mathcal{T}$ is the length of u as a word. When $\mathcal{T}$ is finite, then the maximum of levels in $\mathcal{T}$ is the height of $\mathcal{T}$.

Definition 3.13 ([16])**.** Let $\mathcal{T}$ be an ordered tree. The internal nodes are all the nodes with at least one child, the external nodes or leaves are the nodes having no children. The set of leaves of a tree $\mathcal{T}$ is denoted by $\partial\mathcal{T}$ and the set of internal nodes is $\mathcal{T} \setminus \partial\mathcal{T}$.

Definition 3.14 ([16])**.** Let $\mathcal{T}$ be an ordered tree. The descendants of a node $u \in \mathcal{T}$ are the elements of the tree having u as a prefix.
The subtree starting at u is the set of elements in $\mathbb{U}_{\mathbb{N}_0}$ which we obtain by taking all descendants of u together with u and removing the prefix u from them. We denote it by $\mathcal{T}^{(u)}$. This is again a tree. The subtrees of u are all subtrees starting at the children of u. If a node u can be written as $u = vw$, then every $v \in \mathcal{T}$ is an ancestor of u. The root is sometimes referred to as the ancestor.

3.3 Binary trees

Definition 3.15 ([16])**.** We define

$$\mathbb{U}_{\{0,1\}} \coloneqq \{\mathcal{E}\} \cup \bigcup_{n \in \mathbb{N}} \{0,1\}^n.$$

If $\mathcal{T}$ is the empty tree or a subset $\mathcal{T}$ of $\mathbb{U}_{\{0,1\}}$ such that for all $u, v \in \mathbb{U}_{\{0,1\}}$, if $uv \in \mathcal{T}$, then $u \in \mathcal{T}$, it is called a binary tree. Consequently, a node in $\mathcal{T}$ can only have 0, 1 or 2 children.

Definition 3.16 ([16])**.** A complete binary tree is a subset $\mathcal{T}$ of $\mathbb{U}_{\{0,1\}}$ such that

(i) $\mathcal{E} \in \mathcal{T}$;

(ii) $\forall u, v \in \mathbb{U}_{\{0,1\}}$, if $uv \in \mathcal{T}$, then $u \in \mathcal{T}$;

(iii) $\forall u \in \mathcal{T}$, $u0 \in \mathcal{T} \iff u1 \in \mathcal{T}$.

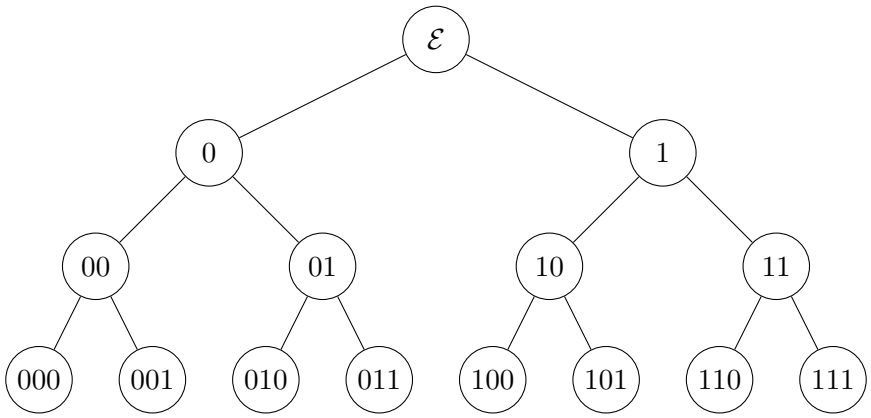

Figure 1: A complete binary tree of 15 nodes, where all leaves have the same level.

Lemma 3.17 ([16])**.** *A complete binary tree $\mathcal{T}$ with $n \in \mathbb{N}_0$ internal nodes has $n+1$ leaves.*

Proof. We proceed by induction. When the tree is reduced to a root $\mathcal{E}$, the root is a leaf and the number of internal nodes is 0. Now, we assume that the statement holds for $0 \leq k \leq n$, with $n \in \mathbb{N}_0$, and we prove it for $n+1$. So, let $\mathcal{T}$ be a complete binary tree with $n+1$ internal nodes. The root is an internal node, so consider the two subtrees of the root. Assume that the left subtree has $k < n+1$ internal nodes, then the right subtree has $n+1-k-1 = n-k$ internal nodes. Therefore, the left subtree has by induction $k+1$ leaves and the right subtree has $n-k+1$ leaves. Since the root was an internal node, the number of leaves did not change when only considering the subtrees. So $\mathcal{T}$ has $k+1+n-k+1 = n+2$ leaves, which is what we wanted to show. Therefore it holds for all $n \in \mathbb{N}_0$. □

Definition 3.18 ([16])**.** A labelled tree is a pair $\overline{\mathcal{T}} = (\mathcal{T}, (\gamma_u)_{u \in \mathcal{T}})$, where $\mathcal{T}$ is an ordered tree and the labels $(\gamma_u)_{u \in \mathcal{T}}$ are elements in a set Γ.
The unlabelled tree $\mathcal{T}$ is called the form of the labelled tree $\overline{\mathcal{T}}$. It is denoted by $\pi(\overline{\mathcal{T}})$ and it is the set of nodes in the labelled tree.

Remark 3.19 ([16])**.** The set of labels can be anything, usually we take a set of numbers. If this set is totally ordered, we are often only interested in the order of the labels, not in the value of the labels.

Definition 3.20 ([16])**.** Let $\mathcal{T}$ be a finite labelled tree and assume that the labels belong to a totally ordered set. Let $\pi(\mathcal{T})$ be the form of the tree and let k be the number of distinct labels in $\mathcal{T}$. The ranked tree of $\mathcal{T}$ is the labelled tree $C(\mathcal{T})$ obtained by labelling the nodes in the tree $\pi(\mathcal{T})$ by the numbers $1, 2, \ldots, k$ and by keeping the labels in the same order as in $\mathcal{T}$. We call this procedure the canonical ranking.

Definition 3.21 ([16])**.** Let $\mathcal{T}$ be a finite labelled tree of size k and assume that the labels $x_1, x_2, \ldots, x_k$ belong to a totally ordered set and all are distinct. The reordering of $x_1, x_2, \ldots, x_k$ is the permutation $\sigma_k \in S_k$ (see Definition 2.13) defined by

$$x_{\sigma_k(1)} < x_{\sigma_k(2)} < \cdots < x_{\sigma_k(k)}.$$

The permutation σ_k^{-1} gives the ranks of the labels.

Remark 3.22 ([16])**.** The canonical ranking does not change the form of the tree, so we have $\pi(C(\mathcal{T})) = \pi(\mathcal{T})$.

Definition 3.23 ([16])**.** A binary search tree is either an empty tree or a labelled binary tree with the following properties.

(i) The labels are elements in a totally ordered set; we call them keys.

(ii) For every internal node u, the keys in the left subtree of u, if it is non-empty, are less than or equal to the key of u.

(iii) For every internal node u, the keys in the right subtree of u, if it is non-empty, are greater than the key of u.

As a consequence, the two subtrees of any node in the tree are again binary search trees.

Construction 3.24 ([16])**.** Let $n \in \mathbb{N}$ and let $x_1, \ldots, x_n$ be n (distinct) elements in a totally ordered set Γ, the set of keys. The binary search tree which we construct from this sequence of elements is a binary tree of size n, where every node is equipped with one key in the following way.
The first key x_1 is placed at the root ($\mathcal{E}$) of the tree. The second key x_2 is placed at the left child (0) of the root if it is less than or equal to x_1. If it is greater than x_1 it is assigned to (1), the right child of the root. The third key x_3 is first of all compared to the key at the root. If it is less than or equal to x_1, it is placed at the left child (0) if this position is not occupied. If the left child (0) is already occupied, we assign x_3 to its left (00) or right (01) child, depending on whether x_3 is less than or equal to or greater than the key at (0). If x_3 is greater than x_1, it goes to the right subtree of the root. We continue this insertion process recursively until we inserted all the keys in the tree. After n insertions, we have a binary tree of size n, where every node contains a key and satisfies the definition of a binary search tree.

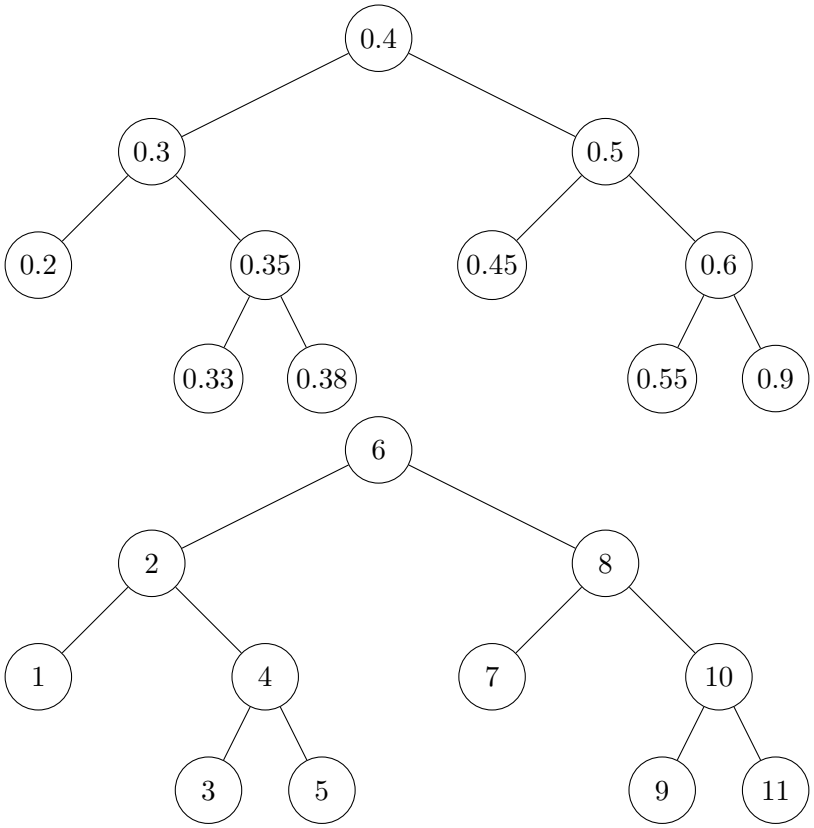

Figure 2: A binary search tree constructed from the keys $x_1 = 0.4$, $x_2 = 0.5$, $x_3 = 0.3$, $x_4 = 0.6$, $x_5 = 0.9$, $x_6 = 0.2$, $x_7 = 0.35$, $x_8 = 0.55$, $x_9 = 0.45$, $x_{10} = 0.33$, $x_{11} = 0.38$ and below its corresponding ranked tree.

Remark 3.25 ([16])**.** If the keys are inserted in a different order, we may still obtain the same binary search tree.

Remark 3.26 ([16])**.** It is possible to add all the $n + 1$ leaves to $\mathcal{T}$, which then correspond to the insertion possibilities for the next key. By doing this, we obtain a complete binary tree.

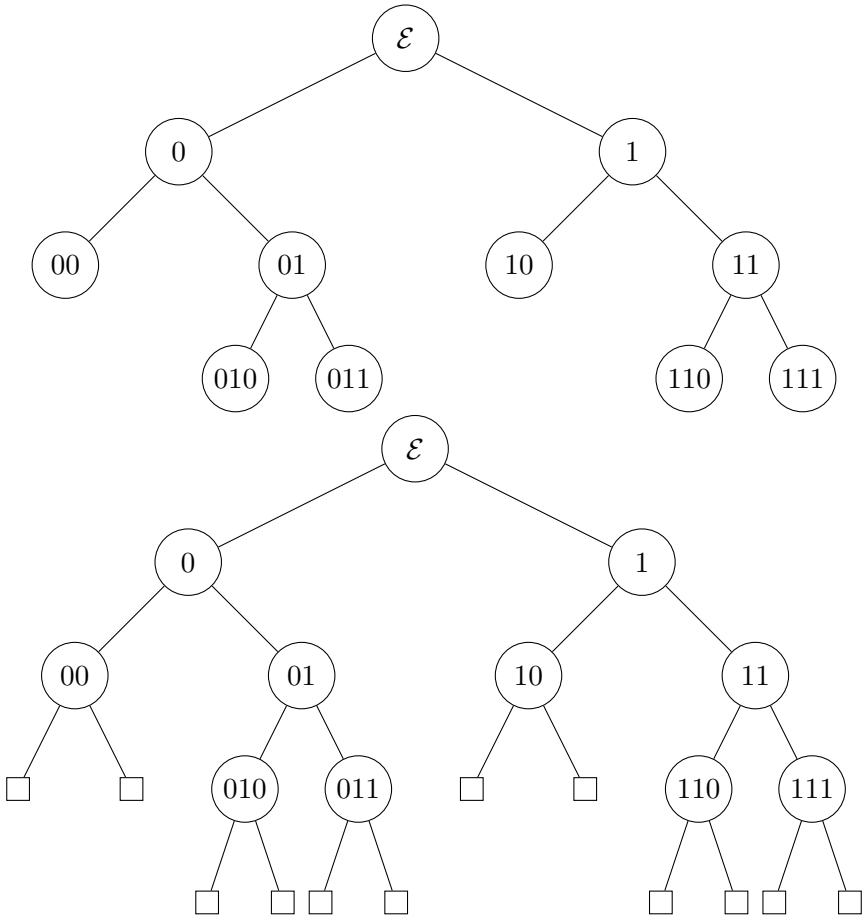

Figure 3: In a binary tree of 11 nodes, we can add the 12 leaves that correspond to the possibilities where the next key can be inserted.

3.4 Randomness on binary trees

Definition 3.27 ([16])**.** We define a stochastic process of complete binary trees $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ in discrete time in the following way. At time $n = 0$, we start with just one leaf (the root) $\mathcal{T}_0$. At time $n \in \mathbb{N}_0$, we choose uniformly one of the $n + 1$ leaves in the tree $\mathcal{T}_n$ and replace it with an internal node and two leaves. This gives us the tree $\mathcal{T}_{n+1}$.
By convention, we say that the size of a complete binary tree is the number of its internal nodes, so $|\mathcal{T}_0| = 0$ and $|\mathcal{T}_n| = n$.
Said differently, let V_n be the leaf of $\mathcal{T}_n$, which we choose uniformly among the $n + 1$ leaves. Then, the tree grows two new leaves V_n0 and V_n1, which gives us the tree $\mathcal{T}_{n+1}$.
The sequence $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ is a discrete time Markov chain with values in the set of all complete binary trees, defined by $\mathcal{T}_0 = \{\mathcal{E}\}$ and $\mathcal{T}_{n+1} = \mathcal{T}_n \cup \{V_n0, V_n1\}$ with $\mathbb{P}(V_n = u \mid \mathcal{T}_n) = \frac{1}{n+1}$, $u \in \partial\mathcal{T}_n$.
We call such a process a burgeoning complete binary tree process.

Proposition 3.28 ([16] Proposition 2.4)**.** *Let $x_1, x_2, \ldots, x_n, \ldots$ be a sequence of independent and identically distributed random variables with values in a totally ordered set. We furthermore assume, that the random variables have a continuous distribution. For every $n \in \mathbb{N}$, let $\sigma_n \in S_n$ be the reordering of $x_1, x_2, \ldots, x_n$, given by*

$$x_{\sigma_n(1)} < x_{\sigma_n(2)} < \cdots < x_{\sigma_n(1)}.$$

Then, σ_n is uniformly distributed on S_n.

Proof. The assumption of a continuous distribution assures that the random variables are almost surely distinct. Since the random variables are i.i.d., they are interchangeable, which implies that the reordering is uniformly distributed. The exact values of the random variables are not important, only their order. This implies that σ_n is uniformly distributed on S_n. □

Definition 3.29 ([16])**.** If a tree is labelled with independent and identically distributed keys that all have the same continuous distribution, we talk about the uniform permutation model.

Definition 3.30 ([16])**.** A random binary search tree is a binary search tree where the labels of the nodes are random variables with values in a totally ordered set.
We only treat the case where the random variables are i.i.d. and their distribution is continuous, so we just consider random binary search trees under the uniform permutation model.

Proposition 3.31 ([16] Proposition 2.8)**.** *Let $x_1, x_2, \ldots, x_n$ be a sequence of $n \in \mathbb{N}$ i.i.d. random variables having the continuous distribution F on the interval $[0, 1]$. Let $\mathcal{T}_n^{(F)}$ be the binary search tree associated to the first n keys, constructed as in Construction 3.24. Then, we have the following properties.*

(i) The sequence $(\mathcal{T}_n^{(F)})_{n\in\mathbb{N}}$ is a Markov chain of labelled trees.

(ii) The form of the binary search tree $\mathcal{T}_n^{(F)}$ does not depend on the distribution F and without loss of generality we can assume that F is the uniform distribution on $[0, 1]$ when we study parameters in the tree which only depend on the form.

(iii) The distribution of the form of the binary search tree $\mathcal{T}_n^{(F)}$ is the same as the distribution of the binary search tree constructed by successively inserting n permuted integers (where the permutation is uniformly distributed on S_n).

Proof.

(i) The random tree $\mathcal{T}_n^{(F)}$ is a value in the set of labelled complete binary trees because every internal node contains a key which is a real number in the interval $[0, 1]$. Since $\mathcal{T}_n^{(F)}$ only depends on $\mathcal{T}_{n-1}^{(F)}$, we have a Markov chain.

(ii) Under the uniform permutation model, the form of the tree $\mathcal{T}_n^{(F)}$, which is constructed from the first n keys, only depends on the relative order of the keys. Said differently, the form only depends on the reordering of the keys, which has the uniform distribution on S_n. This proves also (iii).

□

Definition 3.32 ([16])**.** The distribution on the binary search trees under the uniform permutation model (see Definition 3.29) is called Ord. The keys which we insert into the leaves of the tree are i.i.d. random variables with the same uniform distribution on the interval $[0, 1]$.

Remark 3.33 ([16])**.** We only treat the case of a random binary search tree under the uniform permutation model, so under the distribution Ord. We just write *random binary search tree*, *binary search tree* or *BST*, but the randomness is always there.

Remark 3.34 ([16])**.** We want to study the insertion of the $(n+1)$-th key x_{n+1} in a tree with n internal nodes. We calculate the probability that the $(n+1)$-th key x_{n+1} lies in the interval $(x_{\sigma_n(j)}, x_{\sigma_n(j+1)})$, i.e. the probability that x_{n+1} is the $(j+1)$-th leaf of $\mathcal{T}_n$. We are interested in the form of the tree, so we condition on the reordering σ_n:

$$\mathbb{P}(x_{n+1} \in (x_{\sigma_n(j)}, x_{\sigma_n(j+1)}) \mid \sigma_n) = \mathbb{P}(\sigma_{n+1}(j+1) = n+1) = \frac{1}{n+1}.$$

We can rewrite this equation in terms of the rank R_k of a key x_k, where

$$R_k = \sum_{j=1}^{k} \mathbb{1}_{x_j \leq x_k},\ k \geq 1.$$

Since the random variables x_i are independent and have the same distribution, R_k is uniformly distributed on the set $\{1, \ldots, k\}$, so for every $j = 0, \ldots, n$,

$$\mathbb{P}(R_{n+1} = j+1 \mid R_1, \ldots, R_n) = \frac{1}{n+1}.$$

Thus, if we condition on σ_n, the $(n+1)$-th key is uniformly inserted among the $n+1$ possibilities of insertion. We notice that the form of a binary search tree constructed with i.i.d. labels under the distribution Ord is exactly a burgeoning complete binary tree.

Remark 3.35 ([16])**.** Let $n \in \mathbb{N}$ be fixed. We consider now a sequence of trees $(\mathcal{T}_n)_{n\in\mathbb{N}}$, which we call a tree process.
To every sequence of i.i.d. random variables $(x_n)_{n\in\mathbb{N}}$ with uniform distribution on $[0, 1]$ corresponds a sequence of growing trees $(\mathcal{T}_n)_{n\in\mathbb{N}}$, where growing means that $\mathcal{T}_n$ is contained in $\mathcal{T}_{n+1}$. Moreover, the sequence of reorderings $(\sigma_n)_{n\in\mathbb{N}}$ is compatible, meaning that for every $n \in \mathbb{N}$, the permutation σ_{n+1} (which is canonically represented by an $(n+1)$-tuple) is equal to σ_n after omitting the number $n+1$.
Let $\mathbb{P}_n$ be the distribution on the set of complete binary trees of size n induced by $x_1, \ldots, x_n$ i.i.d. with uniform distribution on $[0, 1]$. Then,

(i) the sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$ is a sequence of growing complete binary trees, where for every $n \in \mathbb{N}$, $\mathcal{T}_n$ has the distribution $\mathbb{P}_n$;

(ii) the sequence of the reorderings $(\sigma_n)_{n\in\mathbb{N}}$ is compatible and for every n, σ_n is uniformly distributed on S_n. As a consequence, in the sequence $(\sigma_n^{-1})_{n\in\mathbb{N}}$, every σ_n^{-1} is also uniformly distributed on S_n for every $n \in \mathbb{N}$.

Therefore, the $\mathbb{P}_n$, for all $n \in \mathbb{N}$, are compatible (meaning that $\mathbb{P}_{n+1}$ restricted to binary trees of size n is equal to $\mathbb{P}_n$) and Kolmogorov's existence theorem (see Theorem I.48) allows us to define a probability measure $\mathbb{P}$ on the set of complete binary trees, where $\mathbb{P}$ is equal to $\mathbb{P}_n$ when restricted to complete binary trees of size n. Hence, we work from now on with $\mathbb{P}$.

4 A theorem about martingales

In this section, we state and prove a theorem about elaborating martingales from a specific behaviour of stochastic processes.

Theorem 4.1 ([14]). *Let $(X_n)_{n\in\mathbb{N}_0}$ be a stochastic process adapted to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Assume further that this process satisfies $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = A_n \cdot X_n$, where A_n is $\mathcal{F}_n$-measurable and almost surely non-zero.*
Let $\nu(0) = 1$ and

$$\nu(n) = \prod_{k=0}^{n-1} A_k$$

for $n \in \mathbb{N}$. Then the process

$$Y_n = \frac{X_n}{\nu(n)}$$

is an $\mathcal{F}_n$-martingale if $\mathbb{E}(|Y_n|) < \infty$ for all $n \in \mathbb{N}_0$.

Proof. First of all, Y_n is $\mathcal{F}_n$-adapted by assumption and we assume $\mathbb{E}(|Y_n|) < \infty$ for all $n \in \mathbb{N}_0$.
Moreover, we have for $n \in \mathbb{N}_0$,

$$\begin{aligned}
\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left(\frac{X_{n+1}}{\nu(n+1)} \mid \mathcal{F}_n\right) \\
&= A_n \cdot \frac{1}{\nu(n+1)} \cdot X_n \\
&= \frac{X_n}{\nu(n)} \\
&= Y_n,
\end{aligned}$$

which proves the statement. □

5 The binary search tree martingale

We define the profile and the level polynomial of a binary search tree and prove several properties of these notions. We develop the binary search tree martingale from the level polynomial of a binary search tree process.

Remark 5.1. We consider a growing binary search tree process. We have seen in section 3 that the form of a growing binary search tree process constructed from a sequence of i.i.d. random variables with a uniform distribution on $[0,1]$ under the distribution Ord is a burgeoning complete binary tree process. So if we are interested in the form of a binary search tree, we can just consider the burgeoning complete binary tree instead (see Definition 3.27).
From now on, we just talk about a binary search tree for simplicity and since we do not lose mathematical precision.
We recall that in a binary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$, the tree $\mathcal{T}_n$ has size n, meaning it has n internal nodes and $n+1$ leaves. The process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ is a discrete time Markov chain with values in the set of all complete binary trees, defined by $\mathcal{T}_0 = \{\mathcal{E}\}$ and $\mathcal{T}_{n+1} = \mathcal{T}_n \cup \{V_n 0, V_n 1\}$, where V_n is a leaf of $\mathcal{T}_n$, uniformly chosen among the $n+1$ leaves, hence $\mathbb{P}(V_n = u \mid \mathcal{T}_n) = \frac{1}{n+1}$ for $u \in \partial\mathcal{T}_n$. Said differently, the process at time $n = 0$ is reduced to a leaf. For $n \in \mathbb{N}_0$, we obtain $\mathcal{T}_{n+1}$ by taking $\mathcal{T}_n$ and replacing a uniformly chosen leaf by an internal node and its two leaves.

5.1 The level polynomial

Definition 5.2 ([16])**.** Let $\mathcal{T}_n$ be a binary search tree of size n. The distribution of the nodes in this tree can be described by

$$U_k(\mathcal{T}_n) \coloneqq \text{number of leaves on level } k \text{ in the tree } \mathcal{T}_n.$$

We call $(U_k(\mathcal{T}_n))_{k\in\mathbb{N}_0}$ the profile of the tree. It contains all the informations about the form of the tree.

Remark 5.3 ([16])**.** We are working here with just the external nodes, but there is an analogue way to do it with just the internal nodes. However, the branching of the tree can be seen more explicit when considering the leaves.

Definition 5.4 ([16])**.** A tree $\mathcal{T}_n$ of size n has $n+1$ leaves, so $\frac{U_k(\mathcal{T}_n)}{n+1}$ is the proportion of leaves on level k. We define the probability measure

$$\mu_{\mathcal{T}_n} \coloneqq \sum_{k=0}^{\infty} \frac{U_k(\mathcal{T}_n)}{n+1} \delta_{\{k\}},$$

where $\delta_{\{k\}}$ is the Dirac measure of the point k (see Definition 2.22).

Remark 5.5 ([16])**.** All informations on the profile of the tree $\mathcal{T}_n$ is contained in $\mu_{\mathcal{T}_n}$ and it shows how the leaves are distributed on the levels of the tree. The sum is finite since there are at most n terms.

We remember that the $(n+1)$-th node is inserted uniformly at random in one of the $n+1$ leaves of a tree $\mathcal{T}_n$. Then the depth of a newly inserted key in $\mathcal{T}_n$ is denoted by D_{n+1} and conditioned on knowing $\mathcal{T}_n$, we have

$$\mathbb{P}(D_{n+1} = k \mid \mathcal{T}_n) = \frac{U_k(\mathcal{T}_n)}{n+1} = \mu_{\mathcal{T}_n}(k). \tag{5.1}$$

By taking the expectation, we get

$$\mathbb{E}(\mathbb{P}(D_{n+1} = k)) = \mathbb{E}\left(\frac{U_k(\mathcal{T}_n)}{n+1}\right) = \mathbb{E}(\mu_{\mathcal{T}_n}(k)). \tag{5.2}$$

Definition 5.6 ([16])**.** We define the level polynomial of the binary search tree $\mathcal{T}_n$ as the polynomial

$$W_{\mathcal{T}_n}(z) \coloneqq \sum_{k=0}^{\infty} U_k(\mathcal{T}_n) z^k.$$

Remark 5.7 ([16])**.** If k is larger than the height of $\mathcal{T}_n$, we have $U_k(\mathcal{T}_n) = 0$, so $W_{\mathcal{T}_n}$ is actually a well-defined polynomial. Moreover, it is a random variable since the $U_k(\mathcal{T}_n)$ are random variables. For $n = 0$, we have $W_{\mathcal{T}_n} = 1$ and if $z = 1$, we get $W_{\mathcal{T}_n} = n+1$, the total number of leaves in $\mathcal{T}_n$.
Even though we do not know its expectation at the moment, we already know from the definition that $\mathbb{E}(|W_{\mathcal{T}_n}|) < \infty$, due to the finiteness of the sum.

Definition 5.8 ([16])**.** The Stirling numbers of the first kind count the number of permutations which can be decomposed into k disjoint cycles in the group S_n (see Definition 2.13). We denote them by

$$\begin{bmatrix} n \\ k \end{bmatrix},$$

where $n, k \in \mathbb{N}_0$. They are defined by the recurrence relation

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} + (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix}; \quad \begin{bmatrix} n \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ k \end{bmatrix} = 0; \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1.$$

The so-called generating function of these numbers is given by

$$f(x) = \sum_{k=0}^{\infty} z^k \begin{bmatrix} n \\ k \end{bmatrix} = z(z+1)\dots(z+n-1).$$

Lemma 5.9 ([16] Théorème 6.6)**.** *Let $\mathcal{T}_n$ be a binary search tree of size n. The expectation of the number of leaves on level k is*

$$\mathbb{E}(U_k(\mathcal{T}_n)) = \frac{2^k}{n!} \begin{bmatrix} n \\ k \end{bmatrix},$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ are the Stirling numbers of the first kind and $n, k \in \mathbb{N}_0$.

The following proof is based on results in [16] and [38].

Proof. We set

$$a_{n,k} = \frac{n!}{2^k}\mathbb{E}(U_k(\mathcal{T}_n))$$

and show that $a_{n,k}$ satisfies the recurrence relation of the Stirling numbers of the first kind. First of all we consider the simple cases:

$$a_{0,0} = \frac{0!}{2^0}\mathbb{E}(U_0(\mathcal{T}_0)) = \mathbb{E}(1) = 1,$$

$$a_{n,0} = \frac{n!}{2^0}\mathbb{E}(U_0(\mathcal{T}_n)) = n! \cdot 0 = 0 \text{ for } n \in \mathbb{N},$$

$$a_{0,k} = \frac{0!}{2^k}\mathbb{E}(U_k(\mathcal{T}_0)) = \frac{1}{2^k} \cdot 0 = 0 \text{ for } k \in \mathbb{N}.$$

Now, we assume $n, k \in \mathbb{N}$. It remains to show that $a_{n,k} = a_{n-1,k-1} + (n-1)a_{n-1,k}$.
To prove this, we use properties of subtrees in binary search trees. $\mathbb{P}(D_{n+1} = k)$ is the the probability that the $(n+1)$-th inserted key has level k, but also the probability that we need to make exactly k comparisons to insert the $(n+1)$-th element into $\mathcal{T}_n$.
Let $\mathbb{P}(D_{n+1} = k \mid x)$ be the probability of making k comparisons given that the key of the root has value x (we recall that all keys are uniformly distributed on [0,1]).
Furthermore, let $\mathbb{P}(D_{n+1} = k \mid x, i)$ be the probability of making k comparisons given that the key of the root has value x and there are precisely i nodes to the left of the root (and hence $n-i-1$ nodes to the right).
Then

$$\mathbb{P}(D_{n+1} = k \mid x, i) = x \cdot \mathbb{P}(D_{i+1} = k-1) + (1-x)\mathbb{P}(D_{n-i-1} = k-1),$$

hence

$$\mathbb{P}(D_{n+1} = k \mid x) = \sum_{i=0}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1} \mathbb{P}(D_{n+1} = k \mid x, i),$$

and thus

$$\begin{aligned}
\mathbb{P}(D_{n+1} = k) &= \int_0^1 \mathbb{P}(D_{n+1} = k \mid x)\, dx \\
&= \int_0^1 \sum_{i=0}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1} \\
&\quad \cdot (x \cdot \mathbb{P}(D_{i+1} = k-1) + (1-x)\mathbb{P}(D_{n-i-1} = k-1))\, dx \\
&= \sum_{i=0}^{n-1} \mathbb{P}(D_{i+1} = k-1) \binom{n-1}{i} \int_0^1 x^{i+1}(1-x)^{n-i-1}\, dx \\
&\quad + \sum_{i=0}^{n-1} \mathbb{P}(D_{n-i-1} = k-1) \binom{n-1}{i} \int_0^1 x^i (1-x)^{n-i}\, dx.
\end{aligned}$$

Now using $\int_0^1 x^i(1-x)^{n-i}\,dx = \frac{k!(n-k)!}{(n+1)!}$, we get

$$\begin{aligned}\mathbb{P}(D_{n+1}=k) &= \sum_{i=0}^{n-1}\frac{i+1}{n(n+1)}\mathbb{P}(D_{i+1}=k-1)\\ &\quad \sum_{i=0}^{n-1}\frac{n-i}{n(n+1)}\mathbb{P}(D_{n-i-1}=k-1)\\ &= \frac{2}{n(n+1)}\sum_{i=0}^{n-1}(i+1)\mathbb{P}(D_{i+1}=k-1).\end{aligned}$$

Reducing n by 1 gives us

$$\mathbb{P}(D_n=k) = \frac{2}{n(n+1)}\sum_{i=0}^{n-2}(i+1)\mathbb{P}(D_{i+1}=k-1)$$

and thus

$$\mathbb{P}(D_{n+1}=k) = \frac{2}{n+1}\mathbb{P}(D_n=k-1) + \frac{n-1}{n+1}\mathbb{P}(D_n=k).$$

In the next step, we use that $\mathbb{P}(D_{n+1}=k) = \mathbb{E}\left(\frac{U_k(\mathcal{T}_n)}{n+1}\right)$ (see Equation 5.2) and get

$$\mathbb{E}\left(\frac{U_k(\mathcal{T}_n)}{n+1}\right) = 2\mathbb{E}\left(\frac{U_{k-1}(\mathcal{T}_{n-1})}{n}\right) + (n-1)\mathbb{E}\left(\frac{U_k(\mathcal{T}_{n-1})}{n}\right).$$

Multiplying by $n!$ and dividing by 2^k finally leads to

$$\frac{n!}{2^k}\mathbb{E}(U_k(\mathcal{T}_n)) = \frac{(n-1)!}{2^{k-1}}\mathbb{E}(U_{k-1}(\mathcal{T}_{n-1})) + (n-1)\frac{(n-1)!}{2^k}\mathbb{E}(U_k(\mathcal{T}_{n-1})),$$

which is what we wanted to prove. □

Remark 5.10 ([16])**.** This lemma and Equation 5.2 imply that

$$\mathbb{P}(D_{n+1}=k) = \frac{2^k}{(n+1)!}\begin{bmatrix}n\\k\end{bmatrix}.$$

Proposition 5.11 ([16] Proposition 6.7)**.** *The distribution of D_{n+1}, the depth of the insertion of the $(n+1)$-th key in a binary search tree $\mathcal{T}_n$ of size n, is given by its generating function*

$$d_n(z) := \sum_{k=0}^{\infty}\mathbb{P}(D_{n+1}=k)z^k = \frac{1}{(n+1)!}\prod_{j=0}^{n-1}(j+2z)$$

Proof. Using Lemma 5.9 and the generating function of the Stirling numbers of the first kind (see Definition 5.8), we can calculate the expectation of the level polynomial

$$\mathbb{E}(W_{\mathcal{T}_n}(z)) = \sum_{k=0}^{\infty}\mathbb{E}(U_k(\mathcal{T}_n))z^k = \frac{1}{n!}\sum_{k=0}^{\infty}2^k\begin{bmatrix}n\\k\end{bmatrix}z^k \tag{5.3}$$

$$= \frac{1}{n!} 2z(2z+1)\dots(2z+n-1). \tag{5.4}$$

From Equation 5.2, we know $\mathbb{P}(D_{n+1}=k) = \frac{1}{n+1}\mathbb{E}(U_{\mathcal{T}_n}(z))$, so

$$d_n(z) = \frac{1}{n+1}\mathbb{E}(W_{\mathcal{T}_n}(z)) = \frac{1}{(n+1)!}\prod_{j=0}^{n-1}(j+2z).$$

□

5.2 The binary search tree martingale

Theorem 5.12 ([16] Théorème 6.8)**.** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be a growing binary search tree process, where $\mathcal{T}_n$ is of size n. Let $W_{\mathcal{T}_n}$ be the level polynomial of the corresponding tree $\mathcal{T}_n$. For every $z \in \mathbb{C}$ such that $2z \neq -k$, $k \in \mathbb{N}$, we define $C_0(z) = 1$ and for $n \in \mathbb{N}$,*

$$C_n(z) \coloneqq \prod_{j=0}^{n-1}\left(\frac{j+2z}{j+1}\right).$$

Let

$$M_n(z) \coloneqq \frac{W_{\mathcal{T}_n}(z)}{\mathbb{E}(W_{\mathcal{T}_n}(z))} = \frac{W_{\mathcal{T}_n}(z)}{C_n(z)}$$

be the normalised level polynomial.
Then $(M_n(z))_{n\in\mathbb{N}_0}$ is an $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$-martingale, where $\mathcal{F}_n = \sigma(\mathcal{T}_0,\dots,\mathcal{T}_n)$ is the filtration generated by $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$, and its expectation is 1.

Remark 5.13 ([16])**.** The normalised level polynomial can also be written as

$$M_n(z) = \frac{1}{C_n(z)}\sum_{u\in\partial\mathcal{T}_n} z^{|u|}.$$

Proof (of Theorem 5.12). When the tree grows from $\mathcal{T}_n$ to $\mathcal{T}_{n+1}$ by insertion of a key x on level k, we remove a leaf on level $k-1$ while gaining two leaves on level k, so

$$U_k(\mathcal{T}_{n+1}) = U_k(\mathcal{T}_n) - \mathbb{1}_{\{D_{n+1}=k\}} + 2\mathbb{1}_{\{k\geq 1\}}\mathbb{1}_{\{D_{n+1}=k-1\}}.$$

With this equation and the fact that a binary search tree process is a Markov chain, we get

$$\begin{aligned}
\mathbb{E}(W_{\mathcal{T}_{n+1}}(z) \mid \mathcal{F}_n) &= \mathbb{E}(W_{\mathcal{T}_{n+1}}(z) \mid \mathcal{T}_n) \\
&= \mathbb{E}\left(\sum_{k=0}^{\infty} U_k(\mathcal{T}_{n+1})z^k \mid \mathcal{T}_n\right) \\
&= \sum_{k=0}^{\infty}\mathbb{E}(U_k(\mathcal{T}_n) - \mathbb{1}_{\{D_{n+1}=k\}} + 2\mathbb{1}_{\{k\geq 1\}}\mathbb{1}_{\{D_{n+1}=k-1\}} \mid \mathcal{T}_n)z^k
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} \mathbb{E}(U_k(\mathcal{T}_n) - \mathbb{P}(D_{n+1} = k \mid \mathcal{T}_n) \\
&\quad + 2\mathbb{1}_{\{k\geq 1\}}\mathbb{P}(D_{n+1} = k-1 \mid \mathcal{T}_n))z^k \\
&= W_{\mathcal{T}_n} - \sum_{k=0}^{\infty} \frac{U_k(\mathcal{T}_n)}{n+1} z^k + 2\sum_{k=1}^{\infty} \frac{U_{k-1}(\mathcal{T}_n)}{n+1} z^k,
\end{aligned}
$$

where we used Equation 5.1. This leads to

$$
\mathbb{E}(W_{\mathcal{T}_{n+1}}(z) \mid \mathcal{F}_n) = \frac{n+2z}{n+1} W_{\mathcal{T}_n}(z). \tag{5.5}
$$

We notice that Equation 5.4 implies

$$
\mathbb{E}(W_{\mathcal{T}_n}(z)) = \prod_{j=0}^{n-1} \frac{j+2z}{j+1} = C_n(z).
$$

The $\mathcal{F}_n$-measurability of $M_n(z)$ is clear. Moreover,

$$
\mathbb{E}(M_n(z)) = \mathbb{E}\left(\frac{W_{\mathcal{T}_n}(z)}{\mathbb{E}(W_{\mathcal{T}_n}(z))}\right) = \frac{\mathbb{E}(W_{\mathcal{T}_n}(z))}{\mathbb{E}(W_{\mathcal{T}_n}(z))} = 1,
$$

and $\mathbb{E}(|M_n(z)|) < \infty$ follows from $\mathbb{E}(|W_{\mathcal{T}_n}(z)|) < \infty$ (see Remark 5.7). Finally, Equation 5.5 together with Theorem 4.1 imply that $M_n(z)$ is an $\mathcal{F}_n$-martingale. □

6 The bisection martingale

We introduce point processes and develop the discrete time branching random walk in $\mathbb{R}$, from which we derive a discrete time martingale. The random bisection problem is defined as a particular case of a discrete time branching random walk in $\mathbb{R}$.

6.1 Point processes

Notation 6.1. We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the usual probability space.

Definition 6.2 ([36])**.** Let $(\mathbb{X}, \mathcal{X})$ be a measurable space. We denote by $N_{<\infty}(\mathbb{X}) \equiv N_{<\infty}$ the space of all measures μ on $\mathbb{X}$ such that $\mu(B) \in \mathbb{N}_0$ for all $B \in \mathcal{X}$, and let $N(\mathbb{X}) \equiv N$ be the space of all measures that can be written as a countable sum of measures from $N_{<\infty}$.

Example 6.3 ([36])**.** Let $x \in \mathbb{X}$. The Dirac measure δ_x at x is $\delta_x(B) = \mathbb{1}_B(x)$ for $B \in \mathcal{X}$ (see Definition 2.22). It satisfies $\delta_x \in N$.
Let $(x_k)_{1 \le k \le n}$ be a finite or infinite sequence of elements in $\mathbb{X}$, with $n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, the number of elements in the sequence. We can define a measure

$$\mu = \sum_{k=1}^{n} \delta_{x_k}. \tag{6.1}$$

Then $\mu \in N$ and $\mu(B) = \sum_{k=1}^{n} \mathbb{1}_B(x_k)$ for $B \in \mathcal{X}$.
For some measurable function $f \colon \mathbb{X} \to \mathbb{R}_{\ge 0}$, we have $\int f \, d\mu = \sum_{k=1}^{n} f(x_k)$.
If we allow $n = 0$, then μ would be the zero measure, which is the most trivial example.

Remark 6.4 ([36])**.** The points $x_1, x_2, \ldots$ in the example before do not have to be distinct. If we have $x_i = x_j$ for some $i \neq j$, then μ is said to have multiplicities. To be more precise, $\operatorname{card}\{j \le n \mid x_j = x_i\}$ is the multiplicity of x_i.

Remark 6.5 ([36])**.** Not every measure $\mu \in N$ can be written in the form of Equation 6.1, but only weak assumptions on $(\mathbb{X}, \mathcal{X})$ and μ are required to achieve it.

Definition 6.6 ([36])**.** Let ν be a measure on $\mathbb{X}$. We say that ν is s-finite if it is a countable sum of finite measures.

Remark 6.7 ([36])**.** By definition, every element of N is s-finite. We remember that a measure ν on $\mathbb{X}$ is σ-finite if there exists a sequence $B_m \in \mathcal{X}$, $m \in \mathbb{N}$, such that $\bigcup_{m \in \mathbb{N}} B_m = \mathbb{X}$ and $\nu(B_m) < \infty$ for all $m \in \mathbb{N}$. This implies that a σ-finite measure is also an s-finite measure. Any σ-finite measure taking values in $\overline{\mathbb{N}_0} := \mathbb{N}_0 \cup \{\infty\}$ is contained in N. Also, a countable sum of s-finite measures is s-finite again.

Definition 6.8 ([36])**.** Let $\mathcal{N}(\mathbb{X}) \equiv \mathcal{N}$ be the σ-algebra generated by the collection of all subsets of N of the form $\{\mu \in N | \mu(B) = k\}$, for $B \in \mathcal{X}$, $k \in \mathbb{N}_0$. $\mathcal{N}$ is the smallest σ-algebra on N such that the map $\mu \mapsto \mu(B)$ is measurable for all $B \in \mathcal{X}$.

Definition 6.9 ([36])**.** A point process on $\mathbb{X}$ is a random element η of $(N, \mathcal{N})$, i.e. a measurable map $\eta\colon \Omega \to N$.
When η is a point process on $\mathbb{X}$ and $B \in \mathcal{X}$, we denote by $\eta(B)$ the map $w \mapsto \eta(w, B) \coloneqq \eta(w)(B)$. By the definition of η and the σ-algebra $\mathcal{N}$, these are random variables taking values in $\overline{\mathbb{N}_0}$. This means that

$$\{\eta(B) = k\} \equiv \{w \in \Omega \mid \eta(w, B) = k\} \in \mathcal{F},\ B \in \mathcal{X},\ k \in \overline{\mathbb{N}_0}. \tag{6.2}$$

Conversely, a mapping $\eta\colon \Omega \to N$ is a point process if Equation 6.2 holds. In this case, we call $\eta(B)$ the number of points of η in B.

Definition 6.10 ([36])**.** Let η and ζ be two point processes on $\mathbb{X}$. They are said to be almost surely equal if there exists an $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$ such that $\eta(w) = \zeta(w)$ for every $w \in A$.

Definition 6.11 ([36])**.** A point process η on $\mathbb{X}$ is called a proper point process if there exist random elements $X_1, X_2, \ldots$ in $\mathbb{X}$ and an $\overline{\mathbb{N}_0}$-valued random variable κ such that almost surely

$$\eta = \sum_{k=1}^{\kappa} \delta_k.$$

If $\kappa = 0$, we view η as the zero measure on $\mathbb{X}$.

Remark 6.12 ([36])**.** Not all point processes are proper.

Definition 6.13 ([36])**.** The intensity measure of the point process η on $\mathbb{X}$ is the measure λ given by

$$\lambda(B) = \mathbb{E}(\eta(B)),\ B \in \mathcal{X}.$$

It follows from the properties of the expectation that the intensity measure of a point process is indeed a measure.

Notation 6.14 ([36])**.** We define $\overline{\mathbb{R}} \coloneqq [-\infty, \infty]$ and denote by $\mathbb{R}(\mathbb{X})$ (resp. $\overline{\mathbb{R}}(\mathbb{X})$) the set of all measurabe functions $u\colon \mathbb{X} \to \mathbb{R}$ (resp. $u\colon \mathbb{X} \to \overline{\mathbb{R}}$).

Notation 6.15 ([36])**.** For a point process η on $\mathbb{X}$ and a function $u \in \overline{\mathbb{R}}(\mathbb{X})$, we denote by $\eta(u) \equiv \int u\, d\eta$ the mapping $w \mapsto \int u(x)\, \eta(w, dx)$.

Proposition 6.16 (Campbell's formula, [36] Proposition 2.7)**.** *Let η be a point process on $(\mathbb{X}, \mathcal{X})$ with intensity measure λ. Let $u \in \overline{\mathbb{R}}(\mathbb{X})$. Then $\int u(x)\, \eta(dx)$ is a random variable with the property that*

$$\mathbb{E}\left(\int_{\mathbb{R}} u(x)\, \eta(dx)\right) = \int_{\mathbb{R}} u(x)\, \lambda(dx) \tag{6.3}$$

whenever $u \geq 0$ or $\int |u(x)|\, \lambda(dx) < \infty$.

Sketch of the proof. The proof of this proposition relies on standard techniques of measure theory. Details can be found in [36]. □

6.2 A martingale in the discrete time branching random walk in $\mathbb{R}$

Definition 6.17 ([11], [20])**.** We define the discrete time branching random walk in $\mathbb{R}$ recursively as follows. At time $n = 0$, an initial ancestor is at the origin. It has children, whose positions on the real line $\mathbb{R}$ are given by a point process $Z^{(1)}$ at time $n = 1$. All the children then produce offspring in the same way as the ancestor. Thus, the position of a new family relative to the corresponding parent is given by an independent copy of $Z^{(1)}$ at time $n = 2$. We obtain the point process $Z^{(2)}$ of the second generation individuals which gives the position of these individuals. We continue this process recursively. The point process of the n-th generation is given by $Z^{(n)}$ at time $n \in \mathbb{N}$. We let $\{Z_r^{(n)} \mid r\}$ be an enumeration of all the positions of the n-th generation.

Remark 6.18 ([11])**.** Let μ be the intensity measure of $Z^{(1)}$ and let m be the Laplace transform of μ (see Definition 2.23). Then,

$$\begin{aligned} m(\lambda) &= \int_{\mathbb{R}} e^{-\lambda x}\, \mu(dx) = \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(1)}(dx) \right) \\ &= \mathbb{E}\left(\sum_r e^{-\lambda Z_r^{(1)}} \right), \end{aligned}$$

where we used Campbell's formula (Proposition 6.16) in the second equation.

Remark 6.19 ([9], [11])**.** We want the branching random walk to be supercritical, which means we assume $\mu(\mathbb{R}) = m(0) > 1$ and $\mu(\mathbb{R}) < \infty$. It is important to notice that $m(0)$ is the expected number of children of any individual. The condition on the branching random walk to be supercritical means that the event of individuals occurring in every generation is positive.
We further assume that $m(\lambda)$ is finite. However, we only consider particular cases of the branching random walk afterwards, where all these conditions are naturally satisfied.

Lemma 6.20 ([11])**.** *Let $n \in \mathbb{N}$. The intensity measure of the point process of the n-th generation is given by the n-fold convolution of μ, i.e. the intensity measure of $Z^{(n)}$ is μ^{n*} (see Definition 2.26).*

Proof. We argue by induction. For $n = 1$, this is trivial. So let us assume it holds for $n \in \mathbb{N}$ and we prove it for $n + 1$ The definition of the branching process leads to the following relation. For a measurable set B, we have

$$Z^{(n+1)}(B) = \sum_r Z_{n,r}^{(1)}(B - Z_r^{(n)}), \tag{6.4}$$

with $\{Z_{n,r}^{(1)} \mid r\}$ independent copies of $Z^{(1)}$. $Z_{n,r}^{(1)}$ gives the relative position of the family of $Z_r^{(n)}$.
We write μ_{n+1} for the intensity measure of the point process of the $(n + 1)$-th generation, $Z^{(n+1)}$. Then,

$$\mu_{n+1}(B) = \mathbb{E}(Z^{(n+1)}(B)) = \int_{\mathbb{R}} \mathbb{1}_B(x)\, Z^{(n+1)}(dx)$$

$$
\begin{aligned}
&= \sum_r \int_{\mathbb{R}} \mathbb{1}_B(x)\, Z_{n,r}^{(1)}(dx - Z_r^{(n)}) \\
&= \sum_r \int_{\mathbb{R}} \mathbb{1}_B(x + Z_r^{(n)})\, Z_{n,r}^{(1)}(dx) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(x+y)\, Z^{(1)}(dx)\, Z^{(n)}(dy) \\
&= \int_{\mathbb{R}} \mathbb{1}_B(x)\, Z^{(1)} * Z^{(n)}(dx) = \mathbb{E}(Z^{(1)} * Z^{(n)}(B)) = \mu * \mu_n(B).
\end{aligned}
$$

By induction, we have $\mu * \mu_n = \mu * \mu^{n*} = \mu^{(n+1)*}$ and we are done. □

Lemma 6.21 ([11])**.** *Let $n \in \mathbb{N}$. The Laplace transform of the point process of the n-th generation is given by m^n, where m is the Laplace transform of $Z^{(1)}$.*

Proof. We denote the Laplace transform of $Z^{(n)}$ by m_n. Then, using Lemma 6.20 and the definition of convolution of probability measures (see Definition 2.26),

$$
\begin{aligned}
m_n(\lambda) &= \int_{\mathbb{R}} e^{-\lambda x}\, \mu^{n*}(dx) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{-\lambda(x_1+x_2+\cdots+x_n)}\, \mu(dx_1)\, \mu(dx_2) \ldots \mu(dx_n) \\
&= \left(\int_{\mathbb{R}} e^{-\lambda x}\, \mu(dx) \right)^n = m^n(\lambda),
\end{aligned}
$$

where we used independence of $Z^{(1)}$ resp. μ. □

Remark 6.22 ([11])**.** Lemma 6.20 and 6.21 imply now, that

$$
\begin{aligned}
m(\lambda)^n = \int_{\mathbb{R}} e^{-\lambda x}\, \mu^{n*}(dx) &= \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx) \right) \\
&= \mathbb{E}\left(\sum_r e^{-\lambda Z_r^{(n)}} \right).
\end{aligned}
$$

Theorem 6.23 ([11])**.** *We consider the discrete time supercritical branching random walk in $\mathbb{R}$ with point processes $(Z^{(n)})_{n\in\mathbb{N}}$, intensity measures and their corresponding Laplace transforms defined and denoted as before. Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be the filtration with $\mathcal{F}_n = \sigma(Z^{(1)}, \ldots, Z^{(n)})$.*
Then

$$
W^{(n)}(\lambda) = m(\lambda)^{-n} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx)
$$

is an $\mathcal{F}_n$-martingale with expectation $\mathbb{E}(W^{(n)}(\lambda)) = 1$.

Proof. We recall Equation 6.4. For a measurable set B, we have

$$
Z^{(n+1)}(B) = \sum_r Z_{n,r}^{(1)}(B - Z_r^{(n)}),
$$

with $\{Z_{n,r}^{(1)} \mid r\}$ being independent copies of $Z^{(1)}$. $Z_{n,r}^{(1)}$ gives the relative position of the family of $Z_r^{(n)}$.
Equation 6.4 implies now

$$\begin{aligned}
\mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n+1)}(dx) \mid \mathcal{F}_n\right) &= \sum_r \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z_{n,r}^{(1)}(dx - Z_r^{(n)}) \mid \mathcal{F}_n\right) \\
&= \sum_r \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda (x+Z_r^{(n)})}\, Z_{n,r}^{(1)}(dx) \mid \mathcal{F}_n\right) \\
&= \sum_r \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z_{n,r}^{(1)}(dx) \mid \mathcal{F}_n\right) \cdot e^{-\lambda Z_r^{(n)}} \\
&= \sum_r \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(1)}(dx) \mid \mathcal{F}_n\right) \cdot e^{-\lambda Z_r^{(n)}} \\
&= \sum_r m(\lambda) \cdot e^{-\lambda Z_r^{(n)}} \\
&= m(\lambda) \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx).
\end{aligned}$$

We assumed furthermore that $m(\lambda)$ is finite and non-zero. Therefore,

$$\begin{aligned}
\mathbb{E}(W^{(n)}(\lambda)) = \mathbb{E}(W^{(1)}(\lambda)) &= \mathbb{E}\left(m(\lambda)^{-1} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(1)}(dx)\right) \\
&= \mathbb{E}(m(\lambda)^{-1} \cdot m(\lambda)) = 1.
\end{aligned}$$

Moreover, $W^{(n)}$ is non-negative, so $W^{(n)} \in L^1$ for all $n \in \mathbb{N}$. The $\mathcal{F}_n$-adaptation is trivial. Therefore, we can apply Theorem 4.1 (thanks to Lemma 6.20 and 6.21) to conclude that $W^{(n)}(\lambda)$ is an $\mathcal{F}_n$-martingale. □

Remark 6.24 ([9], [10], [20])**.** Let

$$\Lambda(\lambda) = \log\left(\mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(1)}(dx)\right)\right) = \log(m(\lambda))$$

and

$$W^{(n)}(\lambda) = m(\lambda)^{-n} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx).$$

We know that $W^{(n)}(\lambda)$ is non-negative and therefore the limit

$$W^{\infty}(\lambda) = \lim_{n\to\infty} W^{(n)}(\lambda)$$

exists almost surely (see Theorem I.39).
It was proven in [9] and [10] using mostly analytical methods that this limit depends on the value of λ. If $\mathbb{E}(W^{(1)} \cdot |W^{(1)}|) < \infty$ then $W^{(n)}(\lambda)$ converges in L^1 and $\mathbb{E}(W^{\infty}(\lambda)) = 1$ whenever

$$\lambda\Lambda'(\lambda) - \Lambda(\lambda) < 0.$$

These values of λ are called supercritical.
On the other hand, $W^\infty(\lambda) = 0$ almost surely if

$$\lambda\Lambda'(\lambda) - \Lambda(\lambda) \geq 0.$$

6.3 The bisection problem and the associated martingale

Definition 6.25 ([23])**.** The random bisection problem is defined as follows. We consider an interval of a certain length s and cut it into two pieces $s_1 = U \cdot s$ and $s_2 = (1-U) \cdot s$, with U being a random variable which is uniformly distributed on $[0,1]$. Independent of each other, each of the two fragments s_1 and s_2 is cut into two pieces again. After the k-th step, there are 2^k pieces (resp. fragments). The lengths of the fragments are random variables and for a given length s at the beginning, one problem is now to determine the probability that the length of each of the 2^k fragments is smaller than a certain length l. Another problem is what the lengths of the 2^k pieces are. We notice that the lengths s and l depend on each other, so we can just assume for instance that $s = 1$. We only consider the case where the interval has length 1.

Construction 6.26 ([20])**.** When we take logarithms in the bisection problem, we get a discrete time branching random walk in $\mathbb{R}$. The point process is $Z^{(1)} = \delta_{\log(U)} + \delta_{\log(1-U)}$, with U a random variable, uniformly distributed over $[0,1]$. Then

$$\begin{aligned} m(\lambda) &= \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(1)}(dx)\right) \\ &= \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, (\delta_{\log(U)}(dx) + \delta_{\log(1-U)}(dx))\right) \\ &= \mathbb{E}\left(e^{-\lambda \log(U)} + e^{-\lambda \log(1-U)}\right) \\ &= \mathbb{E}\left(U^{-\lambda} + (1-U)^{-\lambda}\right) \\ &= \frac{2}{1-\lambda}. \end{aligned}$$

Now we set $z = \frac{1-\lambda}{2}$. Thus we get $m(\lambda)^{-1} = z$ and $m(\lambda)^{-n} = z^n$.
We denote the individuals of the discrete time branching process by u, where $u \in \mathbb{U}_{\{0,1\}} = \{\mathcal{E}\} \cup \bigcup_{n\geq 1}(\{0,1\})^n$ and $|u| = n$ if u is an individual of generation n, just like in the case of a binary tree. Let X_u be the position of the individual u. Then

$$\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx) = \sum_{|u|=n} e^{-\lambda X_u},$$

and now combining all to a branching random walk martingale, we get

$$\begin{aligned} m(\lambda)^{-n} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(n)}(dx) &= z^n \cdot \sum_{|u|=n} e^{-\lambda X_u} \\ &= \sum_{|u|=n} e^{(2z-1)X_u} \cdot z^n. \end{aligned}$$

Definition 6.27 ([20])**.** We define, with the notations as before, the bisection martingale

$$B_n(z) \coloneqq \sum_{|u|=n} e^{(2z-1)X_u} \cdot z^n, z \in \mathbb{R}_{>0}.$$

Theorem 6.28 ([20])**.** *Let $(B_n(z))_{n\in\mathbb{N}_0}$ be the bisection martingale and assume $z \in \mathbb{R}_{>0}$. Then $(B_n(z))_{n\in\mathbb{N}_0}$ converges almost surely to a limit denoted by $B_\infty(z) \coloneqq \lim_{n\to\infty} B_n(z)$.*

Proof. For $z \in \mathbb{R}_{>0}$, $B_n(z)$ is a positive martingale. It is a well-known result in probability theory that the limit of a positive martingale exists almost surely (see Theorem I.39). □

Remark 6.29 ([9], [10], [20])**.** In our case of a bisection martingale, the inequalities of Remark 6.24 mean that $B_n(z)$ converges in L^1 for $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$ and $B_\infty(z) = 0$ otherwise. Here, $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.
The computation goes as follows.
$\Lambda(\lambda) = \log\left(\frac{2}{1-\lambda}\right)$, $\Lambda'(\lambda) = \frac{1}{1-\lambda}$. Therefore

$$\begin{aligned}
&\lambda\Lambda'(\lambda) - \Lambda(\lambda) = 0 \\
\iff &\frac{\lambda}{1-\lambda} - \log\left(\frac{2}{1-\lambda}\right) = 0 \\
\iff &\lambda - (1-\lambda) \cdot \log\left(\frac{2}{1-\lambda}\right) = 0 \mid z = \frac{1-\lambda}{2},\ \lambda = 1-2z \\
\iff &1 - 2z - 2z \cdot \log(z^{-1}) = 0 \\
\iff &1 - 2z + 2z \cdot \log(z) = 0 \mid z = \frac{x}{2} \\
\iff &1 - x + x \cdot \log\left(\frac{x}{2}\right) = 0.
\end{aligned}$$

7 The Yule tree martingale

We use point processes to define the continuous time branching random walk in $\mathbb{R}$ and elaborate a martingale. The Yule tree process is introduced as a particular case of a continuous time branching random walk in $\mathbb{R}$.

7.1 A martingale in the continuous time branching random walk in $\mathbb{R}$

Definition 7.1 ([10], [20])**.** We define a continuous time branching random walk in $\mathbb{R}$ recursively as follows. We consider a continuous time setting, so the time is denoted by t and we have $t \in \mathbb{R}_{\geq 0}$. We start with an ancestor at the origin at time $t = 0$. This ancestor has a lifetime which is exponentially distributed, we can think of a clock attached to this individual and when the clock rings, the individual dies. We want the event when the clock rings to be exponentially distributed with parameter β. The ancestor does not move and when it dies, it creates children, whose positions on the real line $\mathbb{R}$ are given by a point process Z. All the children have again an independent exponentially distributed lifetime with parameter β and they do not move during their lifetime. At their death, they are each replaced by children according to a copy of the point process Z. We denote by $Z^{(t)}$ the point process of the individuals alive at time t. Furthermore, we enumerate the individuals by $r \in \mathbb{N}$. Let
$A_t = \{\text{individuals alive at time } t\}$ be the set of all individuals alive at time t, i.e. $r \in A_t$ if individual r is alive at time t. Let $(Z_r^{(t)})_{r \in A_t}$ be an enumeration of the positions of all individuals alive at time t. The time of the first death is denoted by $\tau \in \mathbb{R}_{\geq 0}$.

Remark 7.2 ([10])**.** This process is a continuous time Markov chain (see Definition 2.32), since the lifetime is exponentially distributed.

Remark 7.3 ([11])**.** Let μ be the intensity measure of Z and let $m(\lambda)$ be the Laplace transform of μ. Then

$$m(\lambda) = \int_{\mathbb{R}} e^{-\lambda x}\, \mu(dx) = \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z(dx) \right),$$

where we used Campbell's formula (Proposition 6.16) in the second equation.

Remark 7.4 ([20])**.** We assume (as in the discrete time case) that $1 < m(\lambda) < \infty$. The particular case we are studying afterwards satisfies this condition. We say that the process is supercritical if $1 < \mu(\mathbb{R}) < \infty$, with $m(0) = \mu(\mathbb{R})$.

Lemma 7.5 ([10])**.** *Let $m_t(\lambda) = \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx) \right)$ be the Laplace transform of the intensity measure of the point process $Z^{(t)}$. Then $m_t(\lambda) = m(\lambda)^t$.*

The method used in the following proof is inspired by the proof of Theorem 8.1.1 in [46].

Proof. Let $\varepsilon > 0$ and $t \in \mathbb{R}_{\geq 0}$. We want to study the behaviour of

$$m_{t+\varepsilon}(\lambda) = \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t+\varepsilon)}(dx) \right)$$

when $\varepsilon \to 0$. Our goal is to find its derivative. We look at our branching process at time t and consider the occurring changes in the process when the time moves forward a little bit. There are three possibilities that can occur.

(i) One individual dies, giving birth to some children, the expected number is $m(\lambda)$. This happens with probability $\beta \cdot \varepsilon + o(\varepsilon)$.

(ii) No individual dies, meaning the number of individuals alive at time $t + \varepsilon$ stays the same as the number of individuals at time t. Therefore the point process is the same at both times, i.e. $Z^{(t+\varepsilon)} = Z^{(t)}$ and hence $m_{t+\varepsilon}(\lambda) = m_t(\lambda)$. This happens with probability $1 - \beta \cdot \varepsilon + o(\varepsilon)$.

(iii) More than one individual dies. This happens with probability $O(\varepsilon^2)$.

Taking all theses conditions together, we can state

$$m_{t+\varepsilon}(\lambda) = m_t(\lambda) \cdot m(\lambda) \cdot \beta \cdot \varepsilon + m_t(\lambda) \cdot (1 - \beta \cdot \varepsilon) + O(\varepsilon^2).$$

We define $f(t) \coloneqq m_t(\lambda)$ for some fixed $\lambda \in \mathbb{R}$. Then the above equation is written

$$f(t + \varepsilon) = f(t) \cdot m \cdot \beta \cdot \varepsilon + f(t) \cdot (1 - \beta \cdot \varepsilon) + O(\varepsilon^2).$$

We rearrange the equation and get

$$\frac{f(t + \varepsilon) - f(t)}{\varepsilon} = f(t) \cdot \beta \cdot (m - 1) + O(\varepsilon).$$

Now taking $\lim_{\varepsilon \to 0}$, we have

$$\frac{df}{dt} = f(t) \cdot \beta \cdot (m - 1),$$

which is a differential equation. Solving this leads to

$$f(t) = e^{\beta(m(\lambda)-1) \cdot t + c},$$

where $c \in \mathbb{R}$ is a constant. Since $f(0) = 1 = e^c$, we get that $c = 0$. Conditioning on the first death, and doing the same computations for $m(\lambda)$, we obtain $m(\lambda) = e^{\beta(m(\lambda)-1)}$. Therefore, $m_t(\lambda) = m(\lambda)^t$ for $t \in \mathbb{R}_{\geq 0}$ as stated. □

Remark 7.6. We notice that for the time of the first death $\tau \geq 0$, we have $Z^{(\tau)} = Z$ and therefore $m(\lambda) = m_\tau(\lambda) = m(\lambda)^\tau$.

We finally state and prove the theorem about a martingale in the continuous time branching random walk in $\mathbb{R}$.

Theorem 7.7 ([10]). *We consider the continuous time supercritical branching random walk in $\mathbb{R}$ with point processes Z conditioned on the first death and $(Z^{(t)})_{t\in\mathbb{R}_{\geq 0}}$. The intensity measures and their corresponding Laplace transforms are defined and denoted as before. Let $(\mathcal{F}_t)_{t\in\mathbb{R}_{\geq 0}}$ be the filtration defined by $\mathcal{F}_t = \sigma(Z^{(s)} \mid s \leq t, s \in \mathbb{R}_{\geq 0})$. Then,*

$$W^{(t)}(\lambda) = m(\lambda)^{-t} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx)$$

is an $\mathcal{F}_t$-martingale with expectation $\mathbb{E}(W^{(t)}(\lambda)) = 1$.

Proof. The definition of the branching process leads to the following relation. Let $\tau \in \mathbb{R}_{\geq 0}$ be the time of the first death and let $t \geq \tau$, $s \in \mathbb{R}_{\geq 0}$. For a set B, we have

$$Z^{(t+s)}(B) = \sum_{r\in A_t} Z^{(s)}_{t,r}(B - Z^{(t)}_r), \tag{7.1}$$

with $(Z^{(s)}_{t,r})_{r\in A_t}$ independent copies of $Z^{(s)}$. $Z^{(s)}_{t,r}$ gives the relative position of the family of $Z^{(t)}_r$.
Equation 7.1 implies now

$$\begin{aligned}
\mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t+s)}(dx) \mid \mathcal{F}_t\right) &= \sum_{r\in A_t} \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(s)}_{t,r}(dx - Z^{(s)}_r) \mid \mathcal{F}_t\right) \\
&= \sum_{r\in A_t} \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda (x+Z^{(t)}_r)}\, Z^{(s)}_{t,r}(dx) \mid \mathcal{F}_t\right) \\
&= \sum_{r\in A_t} \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(s)}_{t,r}(dx) \mid \mathcal{F}_t\right) \cdot e^{-\lambda Z^{(t)}_r)} \\
&= \sum_{r\in A_t} \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(s)}(dx) \mid \mathcal{F}_t\right) \cdot e^{-\lambda Z^{(t)}_r} \\
&= \sum_{r\in A_t} \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z^{(s)}(dx)\right) \cdot e^{-\lambda Z^{(t)}_r} \\
&= \sum_{r\in A_t} m(\lambda)^s \cdot e^{-\lambda Z^{(t)}_r} \\
&= m(\lambda)^s \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx).
\end{aligned}$$

We assumed furthermore that $m(\lambda)$ is finite and non-zero. Therefore,

$$\begin{aligned}
\mathbb{E}(W^{(t)}(\lambda)) = \mathbb{E}(W^{(\tau)}(\lambda)) &= \mathbb{E}\left(m(\lambda)^{-\tau} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(\tau)}(dx)\right) \\
&= \mathbb{E}\left(m(\lambda)^{-1} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z(dx)\right) \\
&= \mathbb{E}(m(\lambda)^{-1} \cdot m(\lambda)) = 1.
\end{aligned}$$

Moreover, $W^{(t)}$ is non-negative, so $W^{(t)} \in L^1$ for all $t \in \mathbb{R}_{\geq 0}$. Therefore we have

$$\begin{aligned}\mathbb{E}(W^{(t+s)}(\lambda) \mid \mathcal{F}_t) &= \mathbb{E}\left(m(\lambda)^{-(t+s)} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t+s)}(dx) \mid \mathcal{F}_t\right) \\ &= m(\lambda)^{-(t+s)} \cdot m(\lambda)^s \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx) \\ &= m(\lambda)^{-t} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx) = W^{(t)}(\lambda).\end{aligned}$$

The $\mathcal{F}_t$-adaptation is clear and we conclude that $W^{(t)}(\lambda)$ is an $\mathcal{F}_t$-martingale. □

Remark 7.8 ([10], [20])**.** Let

$$L(\lambda) = \beta\left(\mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z(dx)\right) - 1\right)$$

and

$$W^{(t)}(\lambda) = m(\lambda)^{-t} \cdot \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx).$$

We know that $W^{(t)}(\lambda)$ is non-negative and therefore the limit

$$W^{\infty}(\lambda) = \lim_{t\to\infty} W^{(t)}(\lambda)$$

exists almost surely (see Theorem I.39).
The results from Remark 6.24 can be extended to the continuous time branching walk. Biggins proved in [10] that the behaviour of the martingale depends again on the value of λ. If $\mathbb{E}(W^{(t)} \cdot |W^{(t)}|) < \infty$ then $W^{(t)}(\lambda)$ converges in L^1 and $\mathbb{E}(W^{\infty}(\lambda)) = 1$ whenever

$$\lambda L'(\lambda) - L(\lambda) < 0.$$

These are the supercritical values of λ.
Otherwise, we have $W^{\infty}(\lambda) = 0$ almost surely if

$$\lambda L'(\lambda) - L(\lambda) \geq 0.$$

7.2 The Yule tree martingale

Definition 7.9 ([16])**.** The Yule tree process is recursively defined as follows. We consider a continuous time complete binary tree process, denoted by $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$. The behaviour of each leaf is independent of the others. We begin at time $t = 0$, the tree $\mathbb{T}_0$ is reduced to a leaf. The lifetime of a leaf is an exponentially distributed random variable with parameter 1. When the leaf dies, it gives birth to two children and becomes an internal node. The set of leaves at time t is $\partial\mathbb{T}_t$.
We can also describe this process with individuals. At time $t = 0$, one individual is alive and its lifetime is Exp(1)-distributed. When it dies, it disappears and gives birth to two children. The set of individuals alive at time t is $\partial\mathbb{T}_t$.

Remark 7.10 ([20])**.** The definition of a Yule tree process implies that it is a particular case of a supercritical continuous time branching process in $\mathbb{R}$. The ancestor is located at 0 at time $t = 0$. The positions of the individuals are positive integers and the lifetime parameter β is 1.
The point process is given by $Z = 2\delta_{\log(2)}$ and the position of an individual $u \in \mathbb{T}_t$ alive at time t is given by $X_u(t) = |u| \cdot \log(2)$ (recall that the complete binary tree $\mathbb{T}$ is a subset of $\mathbb{U}_{\{0,1\}} = \{\mathcal{E}\} \cup \bigcup_{n\geq 1}\{0,1\}^n$).

Construction 7.11 ([20])**.** We want to construct a martingale from the Yule tree process using the martingale developed in Theorem 7.7. Using the given point process $Z = 2\delta_{\log(2)}$, we have that

$$\begin{aligned} m(\lambda) &= \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, Z(dx)\right) = \mathbb{E}\left(\int_{\mathbb{R}} e^{-\lambda x}\, 2\delta_{\log(2)}(dx)\right) \\ &= \mathbb{E}\left(e^{-\lambda \log(2)} \cdot 2\right) = 2^{-\lambda+1}. \end{aligned}$$

Now, given that the position of individual u alive at time t is $X_u(t) = |u| \cdot \log(2)$, we get

$$\begin{aligned} \int_{\mathbb{R}} e^{-\lambda x}\, Z^{(t)}(dx) &= \sum_{u\in\partial\mathbb{T}_t} e^{-\lambda \cdot X_u(t)} \\ &= \sum_{u\in\partial\mathbb{T}_t} e^{-\lambda\cdot|u|\cdot\log(2)} \\ &= \sum_{u\in\partial\mathbb{T}_t} 2^{-\lambda\cdot|u|} \\ &= \sum_{u\in\partial\mathbb{T}_t} z^{|u|}, \end{aligned}$$

where we set $z = 2^{-\lambda}$. Finally, we use $m(\lambda) = e^{(m(\lambda)-1)}$ to get

$$m(\lambda) = 2^{-\lambda+1} = 2z = e^{(m(\lambda)-1)} = e^{(2z-1)},$$

respectively

$$m(\lambda)^t = e^{(2z-1)t}.$$

Everything combined gives us the following continuous time branching random walk martingale

$$\sum_{u\in\partial\mathbb{T}_t} z^{|u|} \cdot e^{(1-2z)t}.$$

Definition 7.12 ([20])**.** We define, with the notations as before, the Yule tree martingale

$$\mathscr{M}_t(z) \coloneqq \sum_{u\in\partial\mathbb{T}_t} z^{|u|} \cdot e^{(1-2z)t},\ z \in \mathbb{R}_{>0},\ t \in \mathbb{R}_{\geq 0},$$

where the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_{\geq 0}}$ is defined by $\mathcal{F}_t = \sigma(\mathbb{T}_s \mid s \leq t, s \in \mathbb{R}_{\geq 0})$.

Theorem 7.13 ([20]). *Let $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$ be the martingale induced from a Yule tree process $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ and assume $z \in \mathbb{R}_{>0}$.*
Then, $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$ converges almost surely to a limit denoted by $\mathscr{M}_\infty(z) \coloneqq \lim_{t\to\infty} \mathscr{M}_t(z)$.

Proof. For $z \in \mathbb{R}_{>0}$, $\mathscr{M}_t(z)$ is a positive martingale. It is a well-known result in probability theory that the limit of a positive martingale exists almost surely (see Theorem I.39). □

Remark 7.14 ([10], [20]). In the case of a Yule tree martingale, the inequalities of Remark 7.8 mean that $\mathscr{M}_t(z)$ converges in L^1 for $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$ and $\mathscr{M}_\infty(z) = 0$ otherwise, where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.
The computation goes as follows.
$L(\lambda) = 2^{1-\lambda} - 1$, $L'(\lambda) = -2^{1-\lambda} \cdot \log(2)$. Therefore

$$\begin{aligned}
&\lambda L'(\lambda) - L(\lambda) = 0 \\
\iff &\lambda \cdot (-2^{1-\lambda}) \cdot \log(2) - 2^{1-\lambda} + 1 = 0 \\
\iff &(-\lambda) \cdot \log(2) \cdot 2^{1-\lambda} - 2^{1-\lambda} + 1 = 0 \mid z = 2^{-\lambda},\ \log(z) = (-\lambda) \cdot \log(z) \\
\iff &2z \cdot \log(z) - 2z + 1 = 0 \mid z = \frac{x}{2} \\
\iff &x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0.
\end{aligned}$$

We see that this is the same equation as in the discrete time case for the bisection.

8 The asymptotic behaviour of the BST martingale

We prove that the limit of the binary search tree martingale exists and derive a 0-1 law for the limit. L^1-convergence is proven for a specific interval and that the limit is zero outside this interval.

Notation 8.1. In this section, we use the same notation as in section 5. We write $(M_n(z))_{n\in\mathbb{N}_0}$ for the martingale induced from a binary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$. The level polynomial is $W_{\mathcal{T}_n}(z)$. The number of leaves on level k in the tree $\mathcal{T}_n$ is $U_k(\mathcal{T}_n)$ and the corresponding probability measure is

$$\mu_{\mathcal{T}_n} = \sum_{k=0}^{\infty} \frac{U_k(\mathcal{T}_n)}{n+1}\delta_{\{k\}}.$$

8.1 First results

Theorem 8.2 ([16])**.** *Let $(M_n(z))_{n\in\mathbb{N}_0}$ be the martingale induced from a binary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ and assume $z \in \mathbb{R}_{\geq 0}$.*
Then $(M_n(z))_{n\in\mathbb{N}_0}$ converges almost surely to a limit denoted by $M_\infty(z) := \lim_{n\to\infty} M_n(z)$.

Proof. For $z \in \mathbb{R}_{\geq 0}$, $M_n(z)$ is a positive martingale. It is a well-known result in probability theory that the limit of a positive martingale exists almost surely (see Theorem I.39). □

Remark 8.3 ([21], [30])**.** Let $H_n = \max\{|u| \mid u \in \partial\mathcal{T}_n\}$ be the height of the tree $\mathcal{T}_n$ and let $h_n = \min\{|u| \mid u \in \partial\mathcal{T}_n\}$. A known result about binary search trees is that (see [22])

$$\lim_{n\to\infty} \frac{h_n}{\log(n)} = c' = 0.3733\ldots \text{ and } \lim_{n\to\infty} \frac{H_n}{\log(n)} = c' = 4.31107\ldots,$$

where c' and c are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.

Remark 8.4 ([30])**.** It follows from the convergence in Remark 8.3 that the support of $\mu_{\mathcal{T}_n}$ lies almost surely in the interval $[c'\log(n) + o(\log(n)), c\log(n) + o(\log(n))]$.

Lemma 8.5 ([30])**.** *Let $C_n(z) = \prod_{j=0}^{n-1}\left(\frac{j+2z}{j+1}\right)$, $n \in \mathbb{N}_0$, as defined in Theorem 5.12. Then*

$$C_n(z) \sim \frac{n^{2z-1}}{\Gamma(2z)} \tag{8.1}$$

and

$$\frac{C_n(z_1)}{C_n(z_2)} \sim n^{2(z_2-z_1)}\frac{\Gamma(2z_2)}{\Gamma(2z_1)}, \tag{8.2}$$

for z, z_1, z_2 bounded in $\mathbb{R}_{\geq 0}$ (see Definition 2.35 and subsection I.7 for results about the Gamma function).

Proof. We notice that

$$\prod_{j=0}^{n-1} \frac{j+2z}{j+1} = \frac{2z \cdot (2z+1) \cdots (2z+n-1)}{1 \cdot 2 \cdots (n-1) \cdot n} = \frac{1}{n!} \cdot \prod_{k=0}^{n-1} (j+2z).$$

Now, we know that for $n \in \mathbb{N}$, $\Gamma(n+1) = n!$ and $\frac{\Gamma(n+x)}{\Gamma(x)} = x \cdot (x+1) \cdots (x+n-1)$. So, we get that

$$C_n(z) = \frac{1}{\Gamma(n+1)} \cdot \frac{\Gamma(2z+n)}{\Gamma(2z)} = \frac{\Gamma(2z-1+n+1)}{\Gamma(2z) \cdot \Gamma(n+1)}.$$

Using that $\Gamma(x+\alpha) \sim \Gamma(x) \cdot x^{\alpha}$ for $x \to \infty$ and $\alpha \in \mathbb{R}$, we get

$$C_n(z) = \frac{\Gamma(2z-1+n+1)}{\Gamma(2z) \cdot \Gamma(n+1)} \sim \frac{(n+1)^{2z-1}}{\Gamma(2z)} \sim \frac{n^{2z-1}}{\Gamma(2z)}.$$

Using this result for the second equation of the lemma, we get

$$\frac{C_n(z_1)}{C_n(z_2)} \sim \frac{\frac{n^{2z_1-1}}{\Gamma(2z_1)}}{\frac{n^{2z_2-1}}{\Gamma(2z_2)}} = n^{2(z_2-z_1)} \frac{\Gamma(2z_2)}{\Gamma(2z_1)},$$

which proves the statement. □

8.2 A relation between subtrees

Remark 8.6 ([30])**.** We consider the set of all finite binary search trees. Let $\mathbb{P}_n$ be the distribution of $\mathcal{T}_n$ over this set (see also Remark 3.35). We can characterize the law of $\mathcal{T}_n$ in the following recursive way. If we denote by $\mathcal{T}^{(0)}(n)$ and $\mathcal{T}^{(1)}(n)$ the left and right subtree of $\mathcal{T}_n$ respectively, then the number of internal nodes $\mathrm{card}(\mathcal{T}_n^{(0)})$ of $\mathcal{T}_n^{(0)}$ is an integer chosen uniformly at random over the set $\{0, 1, \ldots, n-1\}$. It follows that

$$\mathcal{L}(\mathcal{T}_n^{(0)}, \mathcal{T}_n^{(1)} \mid \mathrm{card}(\mathcal{T}_n^{(0)}) = k)) = \mathbb{P}_k \otimes \mathbb{P}_{n-1-k}.$$

Said differently, if we know that $\mathrm{card}(\mathcal{T}_n^{(0)}) = k$, then $\mathcal{T}_n^{(0)}$ and $\mathcal{T}_n^{(1)}$ are independent binary search trees of respective size k and $n-1-k$. This property is also called the burst-property.

Remark 8.7 ([30])**.** We denote by $M_{\mathcal{T}_n^{(u)}}$ the binary search tree martingale of the subtree $\mathcal{T}_n^{(u)}$ and by $M_n(z)$ the binary search tree martingale induced by a tree of size n as usual.

Lemma 8.8 ([30] Lemma 4.1)**.** *Let $t \in \mathbb{R}$ and $z \in \mathbb{R}_{\geq 0}$. Then the following recursive relation holds:*

$$\mathbb{E}\left(e^{-tM_{n+1}(z)}\right) = \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{E}\left(e^{-tz\frac{C_k(z)}{C_{n+1}(z)} M_k(z)}\right) \cdot \mathbb{E}\left(e^{-tz\frac{C_{n-k}(z)}{C_{n+1}(z)} M_{n-k}(z)}\right). \tag{8.3}$$

Proof. First of all, we have for every $k \in \mathbb{N}$, $U_k(\mathcal{T}_n) = U_{k-1}(\mathcal{T}_n^{(0)}) + U_{k-1}(\mathcal{T}_n^{(1)})$. With this, we also have a relation between the martingales $M_n(z)$, $M_{\mathcal{T}_n^{(0)}}(z)$, $M_{\mathcal{T}_n^{(1)}}(z)$ in the following way:

$$W_{\mathcal{T}_{n+1}}(z) = \sum_{k=1}^{\infty} U_k(\mathcal{T}_{n+1}) z^k = \sum_{k=1}^{\infty} (U_{k-1}(\mathcal{T}_{n+1}^{(0)}) + U_{k-1}(\mathcal{T}_{n+1}^{(1)})) z^k,$$

which implies

$$W_{\mathcal{T}_{n+1}}(z) = z(W_{\mathcal{T}_{n+1}^{(0)}}(z) + W_{\mathcal{T}_{n+1}^{(1)}}(z)).$$

Now taking $K = \operatorname{card}(\mathcal{T}_{n+1}^{(0)}) \in [0, n]$, we get

$$M_{n+1}(z) = z \cdot \left(\frac{C_K(z)}{C_{n+1}(z)} M_{\mathcal{T}_{n+1}^{(0)}}(z) + \frac{C_{n-K}(z)}{C_{n+1}(z)} M_{\mathcal{T}_{n+1}^{(1)}}(z) \right), \tag{8.4}$$

and thus

$$\mathbb{E}\left(e^{-tM_{n+1}(z)}\right) = \mathbb{E}\left(e^{-tz\left(\frac{C_K(z)}{C_{n+1}(z)} M_{\mathcal{T}_{n+1}^{(0)}}(z) + \frac{C_{n-K}(z)}{C_{n+1}(z)} M_{\mathcal{T}_{n+1}^{(1)}}(z) \right)} \right). \tag{8.5}$$

We apply the burst-property (see Remark 8.6). So the martingales of the subtrees are independent and their distribution is equal to the one of a binary search tree martingale of size $k = \operatorname{card}(\mathcal{T}_{n+1}^{(0)})$ resp. $n - k = \operatorname{card}(\mathcal{T}_{n+1}^{(1)})$. Knowing that $k \in [0, n]$ uniformly at random gives us finally

$$\mathbb{E}\left(e^{-tM_{n+1}(z)}\right) = \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{E}\left(e^{-tz \frac{C_k(z)}{C_{n+1}(z)} M_k(z)} \right) \cdot \mathbb{E}\left(e^{-tz \frac{C_{n-k}(z)}{C_{n+1}(z)} M_{n-k}(z)} \right).$$

□

Lemma 8.9 ([30] Lemma 4.2)**.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. Then for all $z \in \mathbb{R}_{\geq 0}$,*

$$\frac{1}{n+1} \sum_{k=0}^{n+1} f\left(\frac{C_k(z)}{C_{n+1}(z)} \right) \xrightarrow[n\to\infty]{} \int_0^1 f(x^{2z-1})\, dx.$$

In a more general setting, let $g\colon \mathbb{R}^4 \to \mathbb{R}$ be bounded and contiuous, then for all $z \in \mathbb{R}_{\geq 0}$,

$$\frac{1}{n+1} \sum_{k=0}^{n+1} g\left(M_k(z), \frac{C_k(z)}{C_{n+1}(z)}, M_{n-k}(z), \frac{C_{n-k}(z)}{C_{n+1}(z)} \right) \xrightarrow[n\to\infty]{} \int_0^1 g\left(M_\infty(z), (1-x)^{2z-1}, M_\infty(z), x^{2z-1} \right) dx \text{ a.s.} \tag{8.6}$$

Sketch of the proof. The proof relies on uniform continuity and Lemma 8.5. All the details can be found in [30] in the proof of Lemma 4.2. □

8.3 The 0-1 law of the BST martingale limit

We state and prove the 0-1 law for the limit of the binary search tree martingale.

Proposition 8.10 ([30] Proposition 4.1)**.** *For every $z \in \mathbb{R}_{\geq 0}$, $\mathbb{P}(M_\infty(z) = 0) = 0$ or $\mathbb{P}(M_\infty(z) = 0) = 1$.*

Proof. We use Equation 8.3 in Lemma 8.8 and let $n \to \infty$, this leads to

$$\mathbb{E}\left(e^{-tM_\infty(z)}\right) = \int_0^1 \mathbb{E}\left(e^{-tzx^{2z-1}M_\infty(z)}\right) \cdot \mathbb{E}\left(e^{-tz(1-x)^{2z-1}M_\infty(z)}\right) dx, \tag{8.7}$$

where we used Lemma 8.9. Now, we want to consider this equation when taking $t \to \infty$. We remember that $M_\infty(z) \geq 0$ and as $t \to \infty$, $e^{-tM_\infty(z)}$ goes to 0 on $\{M_\infty(z) > 0\}$, whereas $e^{-tM_\infty(z)} = 1$ for $t \to \infty$ on $\{M_\infty(z) = 0\}$. Said differently, we have $\lim_{t\to\infty} e^{-tM_\infty(z)} = \mathbb{1}_{\{M_\infty(z)=0\}}$. Using the dominated convergence theorem, we get for $0 < x < 1$

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}\left(e^{-tM_\infty(z)}\right) &= \lim_{t\to\infty} \mathbb{E}\left(e^{-tzx^{2z-1}M_\infty(z)}\right) = \lim_{t\to\infty} \mathbb{E}\left(e^{-tz(1-x)^{2z-1}M_\infty(z)}\right) \\ &= \mathbb{P}(M_\infty(z) = 0).\end{aligned}$$

Thus, for $t \to \infty$ in Equation 8.7 and using dominated convergence again gives us

$$\mathbb{P}(M_\infty(z) = 0) = \int_0^1 \mathbb{P}(M_\infty(z) = 0) \cdot \mathbb{P}(M_\infty(z) = 0)\, dx = \mathbb{P}(M_\infty(z) = 0)^2.$$

This implies that $\mathbb{P}(M_\infty(z) = 0) = 0$ or $\mathbb{P}(M_\infty(z) = 0) = 1$ a.s. □

8.4 L^p-boundedness and L^p-convergence

The following lemma is a generalization of a lemma proven in [42].

Lemma 8.11 ([30] Lemma 4.3)**.** *Let $\mathcal{F}$ be a σ-algebra and let L^{p+} be the set of all the non-negative L^p-integrable random variables. We define for $X \in L^{p+}$*
$\mathrm{Var}_p(X \mid \mathcal{F}) \coloneqq \mathbb{E}(X^p \mid \mathcal{F}) - (\mathbb{E}(X \mid \mathcal{F}))^p$.
If $p \in (1, 2]$ and $X, Y \in L^{p+}$ are $\mathcal{F}$-conditionally independent, then

$$\mathrm{Var}_p(X + Y \mid \mathcal{F}) \leq \mathrm{Var}_p(X \mid \mathcal{F}) + \mathrm{Var}_p(Y \mid \mathcal{F})$$

Proof. It is enough to show that for any couple of random variables X, Y such that $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$, the inequality $\mathrm{Var}_p(X + Y) \leq \mathrm{Var}_p(X) + \mathrm{Var}_p(Y)$ holds. The function $f(x) = (x + a)^p - x^p$ is concave on $\mathbb{R}_+$ for $a > 0$ and $p \in (1, 2]$, so

$$\begin{aligned}\mathbb{E}((X + Y)^p - X^p \mid Y) &= \mathbb{E}((X + Y)^p \mid Y) - \mathbb{E}(X^p \mid Y) \\ &\leq (\mathbb{E}(X + Y \mid Y))^p - (\mathbb{E}(X \mid Y))^p \\ &= (\mathbb{E}(X) + Y))^p - (\mathbb{E}(X))^p.\end{aligned}$$

Taking expectation gives us

$$\mathbb{E}((X+Y)^p) - \mathbb{E}(X^p) \leq \mathbb{E}((\mathbb{E}(X)+Y)^p) - (\mathbb{E}(X))^p$$

and

$$\mathbb{E}((Y+\mathbb{E}(X))^p) - \mathbb{E}(Y^p) \leq (\mathbb{E}(Y)+\mathbb{E}(X))^p - (\mathbb{E}(Y))^p.$$

Adding these two inequalities together results in

$$\begin{aligned}\mathbb{E}((X+Y)^p) - \mathbb{E}(X^p) + \mathbb{E}((Y+\mathbb{E}(X))^p) - \mathbb{E}(Y^p) \\ \leq \mathbb{E}((\mathbb{E}(X)+Y)^p) - (\mathbb{E}(X))^p + (\mathbb{E}(Y)+\mathbb{E}(X))^p - (\mathbb{E}(Y))^p,\end{aligned}$$

and thus

$$\mathbb{E}((X+Y)^p) - (\mathbb{E}(Y)+\mathbb{E}(X))^p \leq \mathbb{E}(X^p) - (\mathbb{E}(X))^p + \mathbb{E}(Y^p) - (\mathbb{E}(Y))^p,$$

which is the same as $\mathrm{Var}_p(X+Y) \leq \mathrm{Var}_p(X) + \mathrm{Var}_p(Y)$. □

We state and prove the results about the asymptotic behaviour of the binary search tree martingale in terms of L^p-convergence.

Theorem 8.12 ([30] Proposition 4.1)**.** *Let $z \in \mathbb{R}_{\geq 0}$. For $\frac{c'}{2} < z < \frac{c}{2}$, there exists a $p > 1$ such that $(M_n(z))_{n\in\mathbb{N}_0}$ is bounded in L^p, where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.*

Corollary 8.13 ([30] Proposition 4.1)**.** *Let $z \in \mathbb{R}_{\geq 0}$. For $\frac{c'}{2} < z < \frac{c}{2}$, there exists $p > 1$ such that $M_n(z)$ converges to $M_\infty(z) > 0$ in L^p as $n \to \infty$.*

Proof (of Corollary 8.13). The convergence follows directly from Theorem 8.12 using Doob's martingale convergence theorem (see Theorem I.40). □

Proof (of Theorem 8.12). We know that $z \cdot \frac{C_K(z)}{C_{n+1}(z)} M_{\mathcal{T}_n^{(0)}}(z)$ and $z \cdot \frac{C_{n-K}(z)}{C_{n+1}(z)} M_{\mathcal{T}_n^{(1)}}(z)$ in relation 8.4 are independent when conditioned on $K = \mathrm{card}(\mathcal{T}_{n+1}^{(0)})$, so we apply Lemma 8.11. Thus,

$$\begin{aligned}\mathrm{Var}_p(M_{n+1}(z) \mid K) \leq \left(\frac{zC_K(z)}{C_{n+1}(z)}\right)^p \mathrm{Var}_p(M_{\mathcal{T}_n^{(0)}}(z) \mid K) \\ + \left(\frac{zC_{n-K}(z)}{C_{n+1}(z)}\right)^p \mathrm{Var}_p(M_{\mathcal{T}_n^{(1)}}(z) \mid K).\end{aligned}$$

Now, given that $K = k$, we know that the left and right subtree martingales have the same distribution as the binary search tree martingales of corresponding sizes. Therefore, the above equation becomes

$$\begin{aligned}\mathbb{E}(M_{n+1}^p(z) \mid K = k) - (\mathbb{E}(M_{n+1}(z) \mid K = k))^p \\ \leq \left(\frac{zC_k(z)}{C_{n+1}(z)}\right)^p (\mathbb{E}(M_k^p(z)) - (\mathbb{E}(M_k(z)))^p)\end{aligned}$$

$$+\left(\frac{zC_{n-k}(z)}{C_{n+1}(z)}\right)^p (\mathbb{E}(M_{n-k}^p(z)) - (\mathbb{E}(M_{n-k}(z)))^p).$$

We remember that that $\mathbb{E}(M_n) = 1$ for all $n \in \mathbb{N}_0$. Hence,

$$\begin{aligned}\mathbb{E}(M_{n+1}^p(z) \mid K = k) \leq & \left(\frac{zC_k(z)}{C_{n+1}(z)}\right)^p \mathbb{E}(M_k^p(z)) \\ & + \left(\frac{zC_{n-k}(z)}{C_{n+1}(z)}\right)^p \mathbb{E}(M_{n-k}^p(z)) \\ & + \frac{z^p}{C_{n+1}^p(z)} \cdot ((C_k(z) + C_{n-k}(z))^p - C_k^p(z) - C_{n-k}^p(z)).\end{aligned}$$

The law of K is uniform over $\{0, 1, \ldots, n\}$, so we deduce

$$\begin{aligned}&\mathbb{E}(M_{n+1}^p(z)) \leq \\ &\qquad \frac{2z^p}{n+1} \sum_{k=0}^{n} \left(\frac{C_k(z)}{C_{n+1}(z)}\right)^p \mathbb{E}(M_k^p(z)) + \frac{(2^p-2)z^p}{n+1} \sum_{k=0}^{n} \left(\frac{C_k(z)}{C_{n+1}(z)}\right)^p,\end{aligned}$$

where we used $2^p \cdot \left(\left(\frac{C_k(z)}{C_{n+1}(z)}\right)^p + \left(\frac{C_{n-k}(z)}{C_{n+1}(z)}\right)^p\right)$ as an upper bound for $\left(\frac{C_k(z)}{C_{n+1}(z)} + \frac{C_{n-k}(z)}{C_{n+1}(z)}\right)^p$.
Since $(M_n(z))_{n\in\mathbb{N}_0}$ is a martingale, $(M_n^p(z))_{n\in\mathbb{N}}$ is a sub-martingale, and so we can find an upper bound on $\mathbb{E}(M_k^p(z))$ using $\mathbb{E}(M_n^p(z))$ for $k \leq n$. This leads to

$$\mathbb{E}(M_{n+1}^p(z)) \leq \alpha_n(z,p)\mathbb{E}(M_n^p(z)) + \beta_n(z,p),$$

where the sums $\alpha_n(z,p)$ and $\beta_n(z,p)$ are given by the kind of sums in Lemma 8.9, so we know they converge as $n \to \infty$,

$$\alpha_n(z,p) \to \alpha(z,p) = 2z^p \int_0^1 x^{(2z-1)p}\, dx = \frac{2z^p}{1+p(2z-1)},$$

$$\beta_n(z,p) \to \beta(z,p) = \frac{(2^p-2)z^p}{1+p(2z-1)}.$$

It suffices to have $\alpha(z,p) < 1$ for L^p-boundedness, so we define $f(z,p) = 2z^p - (1 + p(2z-1))$. Then

$$\alpha(z,p) < 1 \iff f(z,p) < 0. \tag{8.8}$$

We notice from computations that $f(z,1) = 0$ and

$$\frac{\partial}{\partial p} f(z,p)\big|_{p=1} = 2z \log\left(\frac{2z}{2e}\right).$$

Moreover, we have that $\frac{\partial}{\partial p} f(z,p)\big|_{p=1} < 0$ when $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$. So whenever z is contained in this interval, we can find $p \in (1,2]$ such that $f(z,p) < 0$. The equivalence in Equation 8.8 now yields that $(M_n(z))_{n\in\mathbb{N}_0}$ is bounded in L^p for $\frac{c'}{2} < z < \frac{c}{2}$. As $\frac{\partial^2}{\partial p^2} f(z,p) \geq 0$, we can not extend this result to the interval $\left[\frac{c'}{2}, \frac{c}{2}\right]$. □

Remark 8.14. The proof of Theorem 8.12 shows, that the L^p-boundedness holds for $p \in (1,2]$. Consequently, the L^p-convergence holds also in L^p for $p \in (1,2]$. In particular, the convergence holds in L^1.

8.5 When the martingale limit is zero

Remark 8.15 ([30])**.** We let $\Lambda_{\mathcal{T}_n}(\theta) = \int e^{\theta x}\, d\mu_{\mathcal{T}_n}(x)$, where $\theta \in \mathbb{R}$, and notice that

$$\Lambda_{\mathcal{T}_n}(\theta) = \frac{W_{\mathcal{T}_n}(e^\theta)}{n+1} = M_n(e^\theta)\frac{C_n(e^\theta)}{n+1} \tag{8.9}$$

$$= M_n(e^\theta)\frac{n^{2(e^\theta-1)}}{\Gamma(e^\theta)}(1+o(1)) \tag{8.10}$$

By Corollary 8.13, we have the following asymptotic behaviour for $\Lambda_{\mathcal{T}_n}(\theta)$ when $\frac{c'}{2} < e^\theta < \frac{c}{2}$:

$$\frac{1}{\log(n)}\log(\Lambda_{\mathcal{T}_n}(\theta)) \xrightarrow[n\to\infty]{} \Lambda(\theta) = 2(e^\theta - 1)\ a.s. \tag{8.11}$$

This property is usually called the Gartner-Ellis condition.
We have the following duality

$$\theta \in \left(\log\left(\frac{c'}{2}\right), \log\left(\frac{c}{2}\right)\right) \Longleftrightarrow \alpha \in (c', c), \text{ with } \alpha = \Lambda'(\theta) = 2e^\theta. \tag{8.12}$$

Lemma 8.16 ([30] Lemma 5.1)**.** *We define for all $\theta \in \mathbb{R}$*

$$\mu_{\mathcal{T}_n,\theta}(x) = \frac{e^{\theta x} d\mu_{\mathcal{T}_n}(x)}{\int e^{\theta x'} d\mu_{\mathcal{T}_n}(x')}.$$

If θ is in duality with α, i.e. if $\alpha = \alpha = \Lambda'(\theta)$, then as $n \to \infty$,

$$\forall \epsilon > 0,\ \mu_{\mathcal{T}_n,\theta}([(\alpha-\epsilon)\log(n), (\alpha+\epsilon)\log(n)]) \to 1\ a.s. \tag{8.13}$$

Sketch of the proof. The proof is based on the Chebycheff inequality. To prove Equation 8.13 for a given θ_0, we need to prove Equation 8.12 for θ_0 and

$$\limsup_{n\to\infty} \frac{1}{\log(n)}\log(\Lambda_{\mathcal{T}_n}(\theta)) \le \Lambda(\theta)\ a.s. \tag{8.14}$$

for all θ in a neighbourhood of θ_0. More details can be found in [30] in the proof of Lemma 5.1. □

We state and prove the final theorem about the asymptotic behaviour of the binary search tree martingale.

Theorem 8.17 ([30] Proposition 5.1)**.** *If $z \in \mathbb{R}_{\ge 0}$ and*

$$z > \frac{c}{2} \text{ or } 0 < z < \frac{c'}{2},$$

then $M_\infty(z) = 0$ almost surely, where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.

Proof. We argue by contradiction. We assume that $M_\infty(z_0) > 0$ for some $z_0 < \frac{c'}{2}$. Then, the Garner-Ellis condition (Equation 8.11) holds for $z_0 = e^{\theta_0}$, i.e. it holds for $\theta_0 = \log(z_0)$. Moreover, Lemma 8.16 holds for $\theta_0 = \log(z_0)$ and $\alpha_0 = \Lambda'(\theta_0)$ as soon as the asymptotic upper bound 8.14 holds for all θ in a neighbourhood of θ_0. This is due to the reason that for all θ, we have $M_n(e^\theta) = O(1)$ almost surely, which implies the asymptotic bounds (Equation 8.14) for all $\theta \in \mathbb{R}$. A consequence is the law of large numbers (Equation 8.13) for $\theta_0 = \log(z_0)$ and $\alpha_0 = 2z_0 < c'$.
We now get a contradiction since the support of $\mu_{\mathcal{T}_n,\theta}$ is included in the support of $\mu_{\mathcal{T}_n}$, so it is included almost surely in $[c' \log(n), c \log(n)]$ for n large enough (see Remark 8.4). The case where $z_0 > \frac{c}{2}$ works in a similar way. □

Remark 8.18 ([30])**.** $M_\infty(z)$ is positive with expectation equal 1 in the area $\left(\frac{c'}{2}, \frac{c}{2}\right)$ and apart from the critical points $\frac{c'}{2}$, $\frac{c}{2}$, it is 0. But we still have $\mathbb{E}(M_n(z)) = 1$ for all $n \in \mathbb{N}_0$. Therefore the martingale $(M_n(z))_{n\in\mathbb{N}_0}$ can not possibly be bounded in L^p for $0 < z < \frac{c'}{2}$ or $z > \frac{c}{2}$.

9 Connections between the binary search tree and Yule tree

In this section, we show a relation between the binary search tree and Yule tree process and extend this relation to a martingale and limit martingale connection. The connection is used to state and prove a result about the profile of a binary search tree.

Notation 9.1. In this section, we use the same notations as in section 5 and 7. We write $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ for a binary search tree process and its corresponding binary search tree martingale is $(M_n(z))_{n\in\mathbb{N}_0}$, where

$$M_n(z) = \frac{1}{C_n(z)} \sum_{u\in\partial\mathcal{T}_n} z^{|u|}$$

and

$$C_0(z) = 1,\ C_n(z) = \prod_{j=0}^{n-1} \left(\frac{j+2z}{j+1}\right) \text{ for } n \in \mathbb{N},\ z \in \mathbb{R}_{>0}.$$

The Yule tree process is denoted by $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ and its martingale is $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$, where

$$\mathscr{M}_t(z) = \sum_{u\in\partial\mathbb{T}_t} z^{|u|} \cdot e^{(1-2z)t},\ z \in \mathbb{R}_{>0}.$$

9.1 The embedding principle

Definition 9.2 ([16])**.** We recall the Yule tree process $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ and $\partial\mathbb{T}_t$ the set of individuals alive at time $t \in \mathbb{R}_{\geq 0}$. We define $N_t \coloneqq \operatorname{card}(\partial\mathbb{T}_t)$, the number of individuals alive at time t. This is a random variable.
We consider further the successive jump times $\tau_0 = 0 < \tau_1 < \tau_2 < \dots$ which are also random variables and defined as

$$\tau_n \coloneqq \inf\{t \in \mathbb{R}_{\geq 0},\ N_t = n+1\}$$

and dually we have

$$N_t = 1 + \sup\{n \in \mathbb{N}_0,\ \tau_n \leq t\}.$$

Remark 9.3 ([16])**.** We remember that the individuals in the Yule tree have an Exp(1)-distributed lifetime. When we consider the Yule tree process at the times τ_n, we find the underlying structure of a binary search tree process. The reason for this is that when n individuals are alive, each with Exp(1)-distributed lifetime and independent of each other, the first one to die is uniformly distributed among these n individuals (see Remark I.35). Said differently, to get from $\mathbb{T}_{\tau_n}$ to $\mathbb{T}_{\tau_{n+1}}$, we choose one leaf uniformly at random and turn it into an internal node and two leaves. This is exactly the process of a binary search tree. Therefore we get

$$(\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0} \overset{\mathcal{L}}{=} (\mathcal{T}_n)_{n\in\mathbb{N}_0}.$$

Moreover, we have for all $n \in \mathbb{N}_0$ and $t \in \mathbb{R}_{\geq 0}$

$$\{N_t = n+1\} = \{\tau_n \leq t < \tau_{n+1}\}$$

which also gives us

$$(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}} \stackrel{\mathcal{L}}{=} (\mathcal{T}_{N_t})_{t\in\mathbb{R}_{\geq 0}}.$$

Proposition 9.4 ([16] Proposition 6.26)**.** *The binary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ and the the Yule tree process $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ considered at the jump times $(\tau_n)_{n\in\mathbb{N}_0}$ satisfy the following properties.*

(i) The time intervals $\tau_n - \tau_{n-1}$ are independent and $\tau_n - \tau_{n-1}$ is an Exp(n)*-distributed random variable.*

(ii) The processes $(\tau_n)_{n\in\mathbb{N}_0}$ and $(\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0}$ are independent.

(iii) The processes $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ and $(\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0}$ have the same distribution.

(iv) The processes $(\mathcal{T}_{N_t})_{t\in\mathbb{R}_{\geq 0}}$ and $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$, where N_t is the number of individuals alive at time t, have the same distribution.

Proof.

(i) By construction, τ_n is the first jump time (or the first death) after τ_{n-1}, so $\tau_n - \tau_{n-1}$ is the minimum of n independent Exp(1)-distributed random variables, which means it is Exp(n)-distributed (see Remark I.35).

(ii) We notice that the process $(\tau_n)_{n\in\mathbb{N}_0}$ describes at which points in time the Yule tree changes, whereas the process $(\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0}$ describes how the Yule tree changes. Therefore they are independent.

(iii) The two considered processes have the same initial state, i.e. a root node, and their evolution behaves the same as we just noticed in the remark before.

(iv) The two considered processes have the same initial state, i.e. a root node, and their evolution behaves the same as we just noticed in the remark before.

□

Remark 9.5 ([21])**.** From now on, we consider the binary search tree process and the Yule tree process in the same probability space in which

$$(\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0} = (\mathcal{T}_n)_{n\in\mathbb{N}_0} \ a.s. \tag{9.1}$$

We say that the binary search tree process is embedded in the Yule tree process and we call this property the embedding principle. The embedding principle is very useful, because it allows us to use the independence properties between the subtrees in the Yule tree process.

9.2 The martingale connection

The following lemma is a classical result and part of a larger theorem. It can be found in [22] (Theorem 5.4) together with a full proof.

Lemma 9.6 ([22] Theorem 5.4)**.** *Let N_t be the number of individuals alive in the Yule tree $\mathbb{T}_t$ at time t. Then $e^{-t} \cdot N_t$ converges almost surely for $t \to \infty$ to a limit*

$$\xi \coloneqq \lim_{t\to\infty} e^{-t} \cdot N_t, \tag{9.2}$$

where ξ is Exp(1)*-distributed.*

Corollary 9.7 ([21])**.** *Let τ_n be the n-th jump time of the Yule tree process $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$. Then*

$$\lim_{n\to\infty} n \cdot e^{-\tau_n} = \xi \ a.s. \tag{9.3}$$

Proof. We have $\lim_{n\to\infty} \tau_n = \infty$ almost surely. Therefore

$$\lim_{n\to\infty} n \cdot e^{-\tau_n} = \lim_{t\to\infty} e^{-t} \cdot N_t = \xi.$$

□

Proposition 9.8 (Martingale connection, [21] Proposition 2.1)**.** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be a binary search tree process and $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ the Yule tree process such that*

$$(\mathcal{T}_n)_{n\in\mathbb{N}_0} = (\mathbb{T}_{\tau_n})_{n\in\mathbb{N}_0} \ a.s.$$

Let $(M_n(z))_{n\in\mathbb{N}_0}$ and $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$ be the corresponding binary search tree martingale and Yule tree martingale respectively. For $z \in \mathbb{R}_{>0}$, we have

$$\mathscr{M}_{\tau_n}(z) = e^{(1-2z)\tau_n} \cdot C_n(z) \cdot M_n(z), \tag{9.4}$$

where τ_n is independent of $M_n(z)$.

Proof. Using the embedding principle (Equation 9.1), we get

$$\sum_{u\in\partial\mathbb{T}_{\tau_n}} z^{|u|} = \sum_{u\in\partial\mathcal{T}_n} z^{|u|}.$$

Now starting with one side of the desired equality gives us

$$\begin{aligned}
\mathscr{M}_{\tau_n}(z) &= \sum_{u\in\partial\mathbb{T}_{\tau_n}} z^{|u|} \cdot e^{(1-2z)\tau_n} = e^{(1-2z)\tau_n} \cdot \sum_{u\in\partial\mathbb{T}_{\tau_n}} z^{|u|} \\
&= e^{(1-2z)\tau_n} \cdot C_n(z) \cdot \frac{1}{C_n(z} \cdot \sum_{u\in\partial\mathbb{T}_{\tau_n}} z^{|u|} \\
&= e^{(1-2z)\tau_n} \cdot C_n(z \cdot \frac{1}{C_n(z)} \cdot \sum_{u\in\partial\mathcal{T}_n} z^{|u|} \\
&= e^{(1-2z)\tau_n} \cdot C_n(z) \cdot M_n(z).
\end{aligned}$$

The reason for the independence is the same as with τ_n and $\mathbb{T}_{\tau_n}$. τ_n describes when the tree changes but $M_n(z)$ describes how it changes, implying they are independent. □

Lemma 9.9 ([21] Proposition 2.1)**.** *For $z \in \mathbb{R}_{>0}$, we have*

$$\lim_{n\to\infty} e^{(1-2z)\tau_n} \cdot C_n(z) = \frac{\xi^{2z-1}}{\Gamma(2z)} \ a.s.,$$

where ξ is Exp(1)*-distributed, as defined in Equation 9.2.*

Proof. We know from Lemma 8.5 that $C_n(z) \sim \frac{n^{2z-1}}{\Gamma(2z)}$. Now using this and Corollary 9.7, we find

$$\lim_{n\to\infty} (n \cdot e^{-\tau_n})^{2z-1} = \lim_{n\to\infty} n^{2z-1} \cdot e^{(1-2z)\tau_n} = \xi^{2z-1} \ a.s.,$$

and thus

$$\lim_{n\to\infty} e^{(1-2z)\tau_n} \cdot C_n(z) = \lim_{n\to\infty} \frac{e^{(1-2z)\tau_n} \cdot n^{2z-1}}{\Gamma(2z)} = \frac{\xi^{2z-1}}{\Gamma(2z)} \ a.s.$$

□

Theorem 9.10 (Limit martingale connection, [21] Theorem 2.4)**.** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be a binary search tree process and $(\mathrm{I\!T}_t)_{t\in\mathbb{R}_{\geq 0}}$ the Yule tree process such that*

$$(\mathcal{T}_n)_{n\in\mathbb{N}_0} = (\mathrm{I\!T}_{\tau_n})_{n\in\mathbb{N}_0} \ a.s.$$

Let $(M_n(z))_{n\in\mathbb{N}_0}$ and $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$ be the corresponding binary search tree martingale and Yule tree martingale respectively. For $z \in \mathbb{R}_{>0}$, we have

$$\mathscr{M}_\infty(z) = \frac{\xi^{(2z-1)}}{\Gamma(2z)} \cdot M_\infty(z) \ a.s., \tag{9.5}$$

where ξ is Exp(1)*-distributed as defined in Equation 9.2 and independent of $M_\infty(z)$.*

Proof. Since $\mathscr{M}_t(z)$ converges almost surely as $t \to \infty$ and since $\lim_{n\to\infty} \tau_n = \infty$ almost surely, we have $\lim_{n\to\infty} \mathscr{M}_{\tau_n}(z) = \mathscr{M}_\infty(z)$. Using Lemma 9.9 and Proposition 9.8, we get

$$\begin{aligned}
\mathscr{M}_\infty(z) &= \lim_{n\to\infty} \mathscr{M}_{\tau_n}(z) \\
&= \lim_{n\to\infty} e^{(1-2z)\tau_n} \cdot C_n(z) \cdot M_n(z) \\
&= \frac{\xi^{(2z-1)}}{\Gamma(2z)} \cdot M_\infty(z) \ a.s.
\end{aligned}$$

The independence of ξ and $M_\infty(z)$ comes from the independence of τ_n and $M_n(z)$ in Proposition 9.8. □

Theorem 9.11 ([15] Theorem 2.2)**.** *Let $(\mathrm{I\!T}_t)_{t\in\mathbb{R}_{\geq 0}}$ be a Yule tree process and let $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$ be the corresponding Yule tree martingale. Let $z \in \mathbb{R}_{>0}$. Then $\mathscr{M}_t(z)$ converges in L^1 to a positive limit $\mathscr{M}_\infty(z)$ for all $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$ and for $z \notin \left[\frac{c'}{2}, \frac{c}{2}\right]$ it converges almost surely to 0, where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.*

Proof. Using the limit martingale connection of the binary search tree and Yule tree martingales (Theorem 9.10) together with the results from Theorem 8.17, Corollary 8.13 and Remark 8.14, we get the statement we wanted to prove. □

Remark 9.12 ([21])**.** We used the limit martingale connection (Theorem 9.10) together with the results we gained from the study of the binary search tree martingale to get results about the behaviour of the Yule tree martingale. We see directly, that the behaviour can not be described on the critical points $\frac{c'}{2}$ and $\frac{c}{2}$. When we go the other way around and use Remark 7.14 where we found results on the Yule tree martingale and then apply the limit martingale connection, we would get the same results about the binary search tree martingale. This method is especially useful because it follows that $\mathscr{M}_\infty(z) = M_\infty(z) = 0$ for $z \in \{\frac{c'}{2}, \frac{c}{2}\}$.

9.3 The profile of a binary search tree

Remark 9.13 ([21])**.** We recall from Definition 5.2 that $(U_k(\mathcal{T}_n))_{k\in\mathbb{N}_0}$ is the profile of a binary search tree $\mathcal{T}_n$, where $U_k(\mathcal{T}_n)$ is the number of leaves on level k of the tree $\mathcal{T}_n$. We have already computed its expectation in Lemma 5.9 using the Stirling numbers of the first kind.

Remark 9.14 ([21], [29])**.** The following asymptotic estimate for the expectation of $U_k(\mathcal{T}_n)$ can be found in [29] and relies on an asymptotic estimate of the Stirling numbers of the first kind. For $l > 0$ as $n \to \infty$ and $k \to \infty$ such that $r = \frac{k}{\log(n)} \leq l$, we have

$$\mathbb{E}(U_k(\mathcal{T}_n)) = \frac{(2\log(n))^k}{k! n \Gamma(r)}(1 + o(1)).$$

We can deduce from this, that for any $l > 0$,

$$\mathbb{E}(U_k(\mathcal{T}_n)) = \frac{n^{1-\eta\left(\frac{k}{\log(n)}\right)}}{\Gamma\left(\frac{k}{\log(n)}\right)\sqrt{2\pi k}}(1 + o(1)), \tag{9.6}$$

where $o(1)$ is uniform for $\frac{k}{\log(n)} \leq l$ and $\eta(x) = x\log\left(\frac{x}{2}\right) - x + 2$ for $x \in \mathbb{R}_{\geq 0}$.

Definition 9.15 ([21])**.** Let $\mathcal{T}_n$ be a binary search tree. We recall the probability measure which gives us all the information on the profile of $\mathcal{T}_n$:

$$\mu_{\mathcal{T}_n} = \sum_{k=0}^{\infty} \frac{U_k(\mathcal{T}_n)}{n+1}\delta_{\{k\}}.$$

Let $\mathbb{T}_t$ be a Yule tree, $t \in \mathbb{R}_{\geq 0}$. Similarly to the binary search tree, we define the random measure counting the levels of leaves in $\mathbb{T}_t$ as

$$\rho_t := \sum_{u\in\partial\mathbb{T}_t} \delta_{|u|},$$

for $t \in \mathbb{R}_{\geq 0}$.

Remark 9.16 ([21])**.** We notice that $\rho_{\tau_n}(k) = U_k(\mathcal{T}_n)$.

Lemma 9.17 ([21] Lemma 3.2)**.** *For any compact subset K of $\left(\frac{c'}{2}, \frac{c}{2}\right)$, we have*

$$\lim_{t\to\infty} \sup_{k\geq 1, z\in K} z^k \sqrt{t} e^{t(1-2z)} \left(\rho_t(k) - \mathscr{M}_\infty(z) e^{-t} \frac{(2t)^k}{k!}\right) = 0 \text{ a.s.}, \tag{9.7}$$

where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.

Sketch of the proof. The proof relies on a continuous time version of Lemma 5 in [10], the Fourier inversion formula and the Cauchy formula. Details can be found in [21] in the proof of Lemma 3.2. □

The convergence of the profile of a binary search tree process is given by the following theorem.

Theorem 9.18 ([21] Theorem 3.1)**.** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be a binary search tree process with corresponding martingale $(M_n(z))_{n\in\mathbb{N}_0}$. For any compact subset K of $\left(\frac{c'}{2}, \frac{c}{2}\right)$, we have*

$$\lim_{n\to\infty} \sup_{k:\, \frac{k}{\log(n)}\in K} \left(\frac{U_k(\mathcal{T}_n)}{\mathbb{E}(U_k(\mathcal{T}_n))} - M_\infty\left(\frac{k}{2\log(n)}\right)\right) = 0 \text{ a.s.}, \tag{9.8}$$

where $c' = 0.373\ldots$ and $c = 4.311\ldots$ are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.

Proof. Let K be a compact subset of $\left(\frac{c'}{2}, \frac{c}{2}\right)$. Lemma 9.17 implies that

$$z^k \sqrt{t} e^{t(1-2z)} \left(\rho_t(k) - \mathscr{M}_\infty(z) e^{-t} \frac{(2t)^k}{k!}\right) = o(1),$$

where o(1) is uniform in k and in $z \in K$. This can be rearranged to

$$\rho_t(k) = \mathscr{M}_\infty(z) e^{-t} \frac{(2t)^k}{k!} + o(1) z^{-k} t^{-\frac{1}{2}} e^{-t(1-2z)}.$$

We write $\mathcal{P}^{(\lambda)}$ for the Poisson law with parameter λ (see Definition 2.18). We see that $\mathcal{P}^{(2t)}$ appears in the above equation and change the probability from $\mathcal{P}^{(2t)}$ to $\mathcal{P}^{(2tz)}$, which gives us

$$\begin{aligned}
\rho_t(k) &= \mathscr{M}_\infty(z) e^t e^{-2t} \frac{(2t)^k}{k!} + o(1) z^{-k} t^{-\frac{1}{2}} e^{-t(1-2z)} && (9.9)\\
&= \mathscr{M}_\infty(z) e^t z^{-k} e^{-2t} e^{2tz} e^{-2tz} \frac{(2tz)^k}{k!} + o(1) z^{-k} t^{-\frac{1}{2}} e^{-t(1-2z)} && (9.10)\\
&= (t^{\frac{1}{2}} \cdot \mathscr{M}_\infty(z) \cdot \mathcal{P}^{(2tz)} + o(1)) z^{-k} t^{-\frac{1}{2}} e^{-t(1-2z)}. && (9.11)
\end{aligned}$$

The local limit theorem ([43] VII Theorem 12) applied to our case of a Poisson random variable, says that

$$\lim_{\lambda\to\infty}\sup_k\left|\sqrt{2\pi\lambda}\cdot\mathcal{P}^{(\lambda)}(k)-e^{-\frac{(k-\lambda)^2}{2\lambda}}\right|=0,$$

and for $\lambda=2tz$ with $z\in K$, we get

$$\lim_{t\to\infty}\sup_{z\in K}\sup_k\left|\sqrt{4\pi tz}\cdot\mathcal{P}^{(2tz)}(k)-e^{-\frac{(k-2tz)^2}{4tz}}\right|=0,$$

which we can rewrite as

$$\left|\sqrt{4\pi tz}\cdot\mathcal{P}^{(2tz)}(k)-e^{-\frac{(k-2tz)^2}{4tz}}\right|=o(1).$$

Multiplying by $\frac{\mathscr{M}_\infty(z)}{\sqrt{4\pi z}}$ and rearranging gives us

$$\mathscr{M}_\infty(z)\cdot t^{\frac{1}{2}}\cdot\mathcal{P}^{(2tz)}(k)=e^{-\frac{(k-2tz)^2}{4tz}}\cdot\frac{\mathscr{M}_\infty(z)}{\sqrt{4\pi z}}+o(1)\cdot\frac{\mathscr{M}_\infty(z)}{\sqrt{4\pi z}}.$$

We insert this equation into Equation 9.11 and get

$$\begin{aligned}\rho_t(k)&=z^{-k}t^{-\frac{1}{2}}e^{-t(1-2z)}\left(e^{-\frac{(k-2tz)^2}{4tz}}\cdot\frac{\mathscr{M}_\infty(z)}{\sqrt{4\pi z}}+o(1)\cdot\frac{\mathscr{M}_\infty(z)}{\sqrt{4\pi z}}+o(1)\right)\\&=\frac{e^{-t(1-2z)}}{z^kt^{\frac{1}{2}}\sqrt{4\pi z}}\left(e^{-\frac{(k-2tz)^2}{4tz}}\mathscr{M}_\infty(z)+o(1)\left(\mathscr{M}_\infty(z)+\sqrt{4\pi z}\right)\right),\end{aligned}$$

where we set

$$A_t(k,z):=\frac{e^{-t(1-2z)}}{z^kt^{\frac{1}{2}}\sqrt{4\pi z}}.$$

We recall that $\rho_{\tau_n}(k)=U_k(\mathcal{T}_n)$ and let $t=\tau_n$, $z=\frac{k}{2\log(n)}$. Using Corollary 9.7 and the estimate 9.6, we get

$$\frac{A_{\tau_n}(k,z)}{\mathbb{E}(U_k(\mathcal{T}_n))\cdot\xi^{1-2z}\cdot\Gamma(2z)}=1+o(1)\text{ and }e^{-\frac{(k-2\tau_nz)^2}{4\tau_nz}}=1+o(1).$$

Furthermore, we have that

$$m:=\sup_{z\in K}\left(\sqrt{4\pi z}+\mathscr{M}_\infty(z)\right)<\infty.$$

Now combining all these results together with the limit martingale connection (Equation 9.5) gives us

$$\begin{aligned}U_k(\mathcal{T}_n)&=\rho_{\tau_n}(k)\\&=A_{\tau_n}(k,z)\cdot\left(e^{-\frac{(k-2\tau_nz)^2}{4\tau_nz}}\frac{1}{\xi^{(2z-1)}\cdot\Gamma(2z)}\cdot M_\infty(z)+o(1)\cdot m\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{A_{\tau_n}(k,z)}{\xi^{1-2z}\Gamma(2z)} \cdot ((1+o(1))M_\infty(z) + \xi^{1-2z} \cdot \Gamma(2z) \cdot o(1)) \\
&= (1+o(1)) \cdot \mathbb{E}(U_k(\mathcal{T}_n)) \cdot ((1+o(1))M_\infty(z) + o(1)) \\
&= \mathbb{E}(U_k(\mathcal{T}_n)) \cdot M_\infty(z) \cdot (1+o(1)),
\end{aligned}
$$

where $z = \frac{k}{2\log(n)}$ and $o(1)$ is uniform in $z \in K$. □

10 Limit connection between the BST and bisection martingales

We study properties of subtrees in a Yule tree and develop a splitting formula for the Yule tree martingale and the binary search tree martingale. A bisection martingale is constructed from a binary search tree process and we prove a martingale limit connection between the binary search tree martingale and the constructed bisection martingale.

Notation 10.1. In this section, we use the same notations as in section 5 and 7. We write $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ for a binary search tree process and its corresponding binary search tree martingale is $(M_n(z))_{n\in\mathbb{N}_0}$. The Yule tree process is denoted by $(\mathbb{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ and its martingale is $(\mathscr{M}_t(z))_{t\in\mathbb{R}_{\geq 0}}$. The interval $\left(\frac{c'}{2}, \frac{c}{2}\right)$ is given by $c' = 0.373\ldots$ and $c = 4.311\ldots$, which are the solutions of the equation $x \cdot \log\left(\frac{x}{2}\right) - x + 1 = 0$.

10.1 The bisection martingale of a binary search tree

Remark 10.2 ([21])**.** We recall that all elements in a Yule tree $\mathbb{T}_t$ are elements in $\mathbb{U}_{\{0,1\}} = \{\mathcal{E}\} \cup \bigcup_{n\geq 1}(\{0,1\})^n$.
Let $u \in \mathbb{U}_{\{0,1\}}$, let $\tau^{(u)} := \inf\{t \mid u \in \mathbb{T}_t\}$ be the (almost surely finite) time when u appears in the Yule tree. For $t > 0$, we set $\mathbb{T}_t^{(u)} := \{v \in \mathbb{U}_{\{0,1\}} \mid uv \in \mathbb{T}_{t+\tau^{(u)}}\}$. This is the tree process that starts at u and grows from there. Therefore, we also need $N_t^{(u)} := \operatorname{card}(\partial\mathbb{T}_t^{(u)})$.
When $t > \tau^{(u)}$, the number of leaves at time t in the subtree grown from the node u is $n_t^{(u)} := N_{t-\tau^{(u)}}^{(u)}$. Since the behaviour of this subtree is the same as the behaviour of the Yule tree, we have together with Equation 9.2 almost surely for every $u \in \mathbb{U}_{\{0,1\}}$,

$$\lim_{t\to\infty} e^{-t} \cdot N_t^{(u)} = \xi_u,$$

and by changing t to $t - \tau^{(u)}$, this gives us almost surely for every $u \in \mathbb{U}_{\{0,1\}}$,

$$\lim_{t\to\infty} e^{-t} \cdot n_t^{(u)} = \xi_u \cdot e^{-\tau^{(u)}}.$$

Here, ξ_u is Exp(1)-distributed as ξ in Equation 9.2.
We assume $u, v \in \mathbb{U}_{\{0,1\}}$ are not in the same line of descent, which means that the random variables ξ_u and ξ_v are independent. We let $t > \tau^{(u0)}$, then

$$n_t^{(u)} = n_t^{(u0)} + n_t^{(u1)} \text{ and } \tau^{(u0)} = \tau^{(u1)},$$

and thus

$$\lim_{t\to\infty} \frac{n_t^{(u0)}}{n_t^{(u)}} = \lim_{t\to\infty} \frac{e^{-t} \cdot n_t^{(u0)}}{e^{-t} \cdot (n_t^{(u0)} + n_t^{(u1)})} = \frac{\xi_{u0} \cdot e^{-\tau^{(u0)}}}{(\xi_{u0} + \xi_{u1}) \cdot e^{-\tau^{(u0)}}} = \frac{\xi_{u0}}{\xi_{u0} + \xi_{u1}} =: U^{(u0)} \text{ a.s.}$$

Since ξ_{u0} and ξ_{u1} are two independent Exp(1)-distributed random variables, it follows that $U^{(u0)}$ is uniformly distributed on $[0,1]$ (see Lemma I.36). Doing the same computation for $u1$ gives us

$$\lim_{t\to\infty} \frac{n_t^{(u1)}}{n_t^{(u)}} = \frac{\xi_{u1}}{\xi_{u0}+\xi_{u1}} = 1 - U^{(u0)} =: U^{(u1)} \ a.s.$$

Therefore, we can now attach a Uni($[0,1]$)-distributed random variable to every node in $\mathbb{U}_{\{0,1\}}$.
We set

$$U := U^{(0)} = \frac{\xi_0}{\xi_0+\xi_1}, \tag{10.1}$$

so we have

$$\xi := \xi_{\mathcal{E}} = e^{-\tau_1}\cdot(\xi_0+\xi_1),\ \xi_0 = U\cdot\xi\cdot e^{\tau_1},\ \xi_1 = (1-U)\cdot\xi\cdot e^{\tau_1}. \tag{10.2}$$

Now, by embedding, the properties we have stated also hold for the nodes in the binary search tree, as $n\to\infty$.

Theorem 10.3 ([21] Theorem 2.4)**.** *Let $z\in\left(\frac{c'}{2},\frac{c}{2}\right)$. We have the following splitting formulas.*

(i) The Yule tree martingale satisfies

$$\mathscr{M}_\infty(z) = z\cdot e^{(1-2z)\tau_1}(\mathscr{M}_{\infty,(0)}(z)+\mathscr{M}_{\infty,(1)}(z))\ a.s.,$$

where $\mathscr{M}_{\infty,(0)}(z)$ and $\mathscr{M}_{\infty,(1)}(z)$ are independent, distributed as $\mathscr{M}_\infty(z)$ and independent of τ_1.

(ii) The binary search tree martingale satisfies

$$M_\infty(z) = z\cdot(U^{2z-1}\cdot M_{\infty,(0)}(z)+(1-U)^{2z-1}\cdot M_{\infty,(1)}(z))\ a.s.,$$

where U is Uni($[0,1]$)*-distributed and defined in Equation 10.1, $M_{\infty,(0)}(z)$, $M_{\infty,(1)}(z)$ are independent, distributed as $M_\infty(z)$ and independent of U.*

Proof. (i) When $t>\tau_1$, we have the decomposition

$$\mathscr{M}_t(z) = z\cdot e^{(1-2z)\tau_1}(\mathscr{M}^{(0)}_{t-\tau_1}(z)+\mathscr{M}^{(1)}_{t-\tau_1}(z)), \tag{10.3}$$

with

$$\mathscr{M}^{(i)}_s(z) = \sum_{u\in\partial\mathbb{T}^{(i)}_s} z^{|u|}\cdot e^{s(1-2z)},$$

where we have defined $\mathbb{T}^{(i)}_s$ in Remark 10.2.

(ii) We now take $t = \tau_n$ in Equation 10.3, condition on the first splitting time τ_1 and apply the branching property. We let $n \to \infty$ and use the limit martingale connection, i.e. Equation 9.5. Then we get

$$\frac{\xi^{2z-1}}{\Gamma(2z)} \cdot M_\infty(z) = z \cdot e^{(1-2z)\tau_1} \left(\frac{\xi_0^{2z-1}}{\Gamma(2z)} \cdot M_{\infty,(0)}(z) + \frac{\xi_1^{2z-1}}{\Gamma(2z)} \cdot M_{\infty,(1)}(z) \right), \quad (10.4)$$

with ξ_0 and ξ_1 defined as in Remark 10.2. The Equations 10.2, give us

$$\begin{aligned} \xi^{2z-1} \cdot M_\infty(z) = z \cdot e^{(1-2z)\tau_1} (&U^{2z-1} \cdot \xi^{2z-1} \cdot e^{(2z-1)\tau_1} \cdot M_{\infty,(0)}(z) \\ &+ (1-U)^{2z-1} \cdot \xi^{2z-1} e^{(2z-1)\tau_1} \cdot M_{\infty,(1)}(z)), \end{aligned}$$

which after dividing by $\xi^{2z-1} \cdot e^{(2z-1)\tau_1}$ is what we wanted to prove.

□

Construction 10.4 ([20])**.** Let $u \in \mathbb{U}_{\{0,1\}}$ be an individual of the binary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$. By Remark 10.2, using the same notation for the discrete time case, the random variables $U^{(u0)}$ and $U^{(u1)}$, are Uni([0, 1])-distributed, independent and $U^{(u0)} + U^{(u1)} = 1$ for all u.
This implies that we can build a bisection branching random walk with these random variables. Using the notation from the definition of the bisection martingale (see Definition 6.27), we have $X_u = \sum_{v \prec u} \log(U^{(v)})$, where we write $v \prec u$ if v is an ancestor of u. Using this, we get the associated bisection martingale

$$B_n(z) := \sum_{|u|=n} \left(\prod_{v \prec u} (U^{(v)})^{2z-1} \right) z^n, \ z \in \mathbb{R}_{>0}. \quad (10.5)$$

10.2 Martingale limit connection

Remark 10.5 ([20])**.** We have already mentioned in Theorem 10.3 that for the binary search tree martingale, the following splitting formula holds for $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$,

$$M_\infty(z) = z \cdot (U^{2z-1} \cdot M_{\infty,(0)}(z) + (1-U)^{2z-1} \cdot M_{\infty,(1)}(z)) \ a.s. \quad (10.6)$$

Now, we iterate this equation and get

$$M_\infty(z) = z^n \cdot \sum_{|u|=n} \left(\prod_{v \prec u} (U^{(v)})^{2z-1} \right) M_{\infty,(u)}(z) \ a.s. \quad (10.7)$$

Moreover, using the previously defined associated bisection martingale (Equation 10.5) and conditioned on the first generation of the bisection process, we have

$$B_\infty(z) = z \cdot (U^{2z-1} \cdot B_{\infty,(0)}(z) + (1-U)^{2z-1} \cdot B_{\infty,(1)}(z)) \ a.s. \quad (10.8)$$

This is exactly the same splitting formula as for the binary search tree martingale.

Definition 10.6 ([20])**.** We define for $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$,

$$j^{BST}(z,x) \coloneqq \mathbb{E}(e^{-xM_\infty(z)})$$

and

$$j^{BIS}(z,x) \coloneqq \mathbb{E}(e^{-xB_\infty(z)}),$$

the Laplace transforms of the binary search tree martingale and the bisection martingale respectively.

Proposition 10.7 ([20] Proposition 4.1)**.** *Let $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$. Then we have*

$$j^{BST}(z,x) = j^{BIS}(z,x),$$

or equivalently,

$$M_\infty(z) \stackrel{\mathcal{L}}{=} B_\infty(z). \tag{10.9}$$

Proof. Equation 10.6 implies for $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$ that

$$\begin{aligned} j^{BST}(z,x) &= \mathbb{E}\left(e^{-xM_\infty(z)}\right) \\ &= \mathbb{E}\left(e^{-xzM_{\infty,(0)}(z)U^{2z-1} - xzM_{\infty,(1)}(1-U)^{2z-1}}\right) \\ &= \int_0^1 \mathbb{E}\left(e^{-xzM_{\infty,(0)}(z)u^{2z-1}}\right) \cdot \mathbb{E}\left(e^{-xzM_{\infty,(1)}(z)(1-u)^{2z-1}}\right) du \\ &= \int_0^1 j^{BST}(z, xu^{2z-1}) \cdot j^{BST}(z, x(1-u)^{2z-1})\, du \end{aligned}$$

and since the behaviour of the bisection martingale is the same (see Equation 10.8), we also have

$$j^{BIS}(z,x) = \int_0^1 j^{BIS}(z, xu^{2z-1}) \cdot j^{BIS}(z, x(1-u)^{2z-1})\, du.$$

Moreover,

$$j^{BST}(z,0) = 1 = j^{BIS}(z,0)$$

and $\mathbb{E}(M_\infty(z)) = \mathbb{E}(B_\infty(z)) = 1$.
Together with some results about the Laplace transform of the limit of the martingale of a general supercritical branching random walk in $\mathbb{R}$ (further details can be found in [20]), we get that the equations stated above imply $j^{BST}(z,x) = j^{BIS}(z,x)$. □

We finally state and prove the theorem about the connection of the limits of the binary search tree martingale and its associated bisection martingale.

Theorem 10.8 ([20] Theorem 3.1)**.** *Let $z \in \left(\frac{c'}{2}, \frac{c}{2}\right)$ and $n \in \mathbb{N}_0$. Let $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be the natural filtration of the bisection martingale. Then*

$$\mathbb{E}(M_\infty(z) \mid \mathcal{F}_n) = B_n(z) \tag{10.10}$$

and

$$M_\infty(z) = B_\infty(z). \tag{10.11}$$

Proof. We rewrite $\mathbb{E}(M_\infty(z) \mid \mathcal{F}_n)$ using Equation 10.7, the definition of the associated bisection martingale, independence of $M_{\infty,(u)}(z)$ with respect to the $(U^{(v)}, v \prec u)$ and the fact that $\mathbb{E}(M_{\infty,(u)}(z)) = 1$.

$$\begin{aligned}\mathbb{E}(M_\infty(z) \mid \mathcal{F}_n) &= \mathbb{E}\left(z^n \cdot \sum_{|u|=n}\left(\prod_{v\prec u}(U^{(v)})^{2z-1}\right) M_{\infty,(u)}(z) \mid \mathcal{F}_\infty\right)\\ &= \mathbb{E}\left(B_n(z) \mid \mathcal{F}_n\right) \cdot \mathbb{E}(M_{\infty,(u)}(z)) = B_n(z) \text{ a.s.},\end{aligned}$$

which gives us the first equation of the theorem.
Now we let $n \to \infty$ to get

$$\mathbb{E}(M_\infty(z) \mid \mathcal{F}_\infty) = B_\infty(z).$$

We set $X \coloneqq M_\infty(z)$, $\mathcal{A} \coloneqq \mathcal{F}_\infty$ and $Y \coloneqq B_\infty(z)$. Then the above equation gives us $\mathbb{E}(X \mid \mathcal{A}) = Y$ and from Equation 10.9 we have $X \overset{\mathcal{L}}{=} Y$. Now using Lemma I.37 implies $X = Y$ almost surely. □

11 A discrete time martingale in Pólya urn processes

We give an introduction to Pólya urns, find a martingale associated to them and discuss some properties of this process concerning eigenvalues and eigenvectors.

11.1 Definition of the process and a martingale

Definition 11.1 ([16])**.** We define a discrete time stochastic process in the following way. We consider an urn which contains balls of two different colours, let us say α white and β blue balls, where $(\alpha, \beta)^T \in \mathbb{N}_0^2 \setminus \{0\}$. We now draw one ball uniformly at random and put it back into the urn. If the colour of the ball is white, we add a white and b blue balls. If the drawn ball is blue, we add c white and d blue balls. We continue drawing uniformly and adding balls in this way. Furthermore, we want $a, b, c, d \in \mathbb{N}_0$. Sometimes this process is also defined with a, b, c, d negative. Then it could go extinct, but we do not consider this case. We call this process a Pólya urn process.
The composition vector is $U_n \in \mathbb{N}_0^2$, with

$$U_n = \begin{pmatrix} U_n^{(1)} \\ U_n^{(2)} \end{pmatrix} = \begin{pmatrix} \text{card}\{\text{white balls at time n}\} \\ \text{card}\{\text{blue balls at time n}\} \end{pmatrix},$$

and the replacement matrix is

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

In our case, we only consider a balanced urn, meaning the total number of balls added each time stays the same, i.e. $S \coloneqq a + b = c + d$ is a constant, called the balance. Let

$$U_0 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

be the initial composition vector. Then after n drawings, the total number of balls in the urn is $\alpha + \beta + n \cdot S = u + n \cdot S$, where $\alpha + \beta = u$ is the total number of initial balls in the urn.
We only treat the case where $S > m \coloneqq a - c = d - b$.

Remark 11.2 ([16])**.** It follows immediately from this definition that a Pólya urn process is a discrete time Markov chain.

Theorem 11.3 ([19])**.** *Let $(U_n)_{n\in\mathbb{N}_0}$ be a Pólya urn process with replacement matrix R, $A = R^T$ and notations as in the definition before.*
Let $\gamma_{x,0}(t) = 1$ and for $n \in \mathbb{N}$,

$$\gamma_{x,n}(t) \coloneqq \prod_{k=0}^{n-1} \left(1 + \frac{t}{x+k}\right).$$

Let $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be the filtration defined by $\mathcal{F}_n = \sigma(U_0, \ldots, U_n)$.

Then

$$\left(\gamma_{\frac{u}{S},n}\left(\frac{A}{S}\right)^{-1}U_n\right)_{n\in\mathbb{N}_0}$$

is a vectorial $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$-martingale.

Proof. We consider the composition vector U_n of the process and assume that

$$U_n = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\begin{cases} U_{n+1} = U_n + \begin{pmatrix} a \\ b \end{pmatrix} & \text{with probability } \frac{x}{x+y}; \\ U_{n+1} = U_n + \begin{pmatrix} c \\ d \end{pmatrix} & \text{with probability } \frac{y}{x+y}. \end{cases}$$

With $x + y = u + n \cdot S$ we have

$$\begin{aligned} \mathbb{E}(U_{n+1} \mid \mathcal{F}_n) &= \frac{x}{x+y}\left(U_n + \begin{pmatrix} a \\ b \end{pmatrix}\right) + \frac{y}{x+y}\left(U_n + \begin{pmatrix} c \\ d \end{pmatrix}\right) \\ &= U_n + \frac{x}{x+y}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{y}{x+y}\begin{pmatrix} c \\ d \end{pmatrix} \\ &= U_n + \frac{1}{x+y}\begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix}. \end{aligned}$$

Using that

$$\begin{aligned} \left(I + \frac{A}{u+n\cdot S}\right)U_n &= U_n + \frac{1}{x+y}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= U_n + \frac{1}{x+y}\begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix}, \end{aligned}$$

we can compute the conditional expectation as

$$\mathbb{E}(U_{n+1} \mid \mathcal{F}_n) = \left(I + \frac{A}{u+n\cdot S}\right)U_n,$$

where we used the fact that the process is a Markov chain. The $\mathcal{F}_n$-adaptation is clear and applying Theorem 4.1 leads to the required equation for a martingale. Note that $\gamma_{\frac{u}{S},n}\left(\frac{A}{S}\right)^{-1}U_n$ has always positive components and

$$E\left(\gamma_{\frac{u}{S},n}\left(\frac{A}{S}\right)^{-1}U_n\right) = \mathbb{E}\left(\gamma_{\frac{u}{S},0}\left(\frac{A}{S}\right)^{-1}U_0\right) = \mathbb{E}(U_0) < \infty,$$

thus $\gamma_{\frac{u}{S},n}\left(\frac{A}{S}\right)^{-1}U_n$ is an $\mathcal{F}_n$-martingale. □

11.2 Eigenvectors and eigenvalues

Remark 11.4 ([19])**.** Let $(U_n)_{n\in\mathbb{N}_0}$ be a Pólya urn process with replacement matrix R, $A = R^T$ and notations as before. A standard computation shows that the eigenvalues of R resp. A are S and $m = a - c = d - b$. We write u_1 and u_2 for the following two linear eigenforms of A, respectively associated with the eigenvalues S and m, meaning that $u_1 \circ A = Su_1$ and $u_2 \circ A = mu_2$:

$$u_1 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{S}(x + y),\ u_2 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{S}(bx - cy). \tag{11.1}$$

The corresponding vectors v_1 and v_2 of the dual basis of u_1 and u_2 are:

$$v_1 = \frac{S}{b+c} \begin{pmatrix} c \\ b \end{pmatrix},\ v_2 = \frac{S}{b+c} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{11.2}$$

These are eigenvectors of A and the projections on the eigenlines are u_1v_1 and u_2v_2.

Definition 11.5 ([19])**.** A Pólya urn process is called large, if $\sigma_2 \coloneqq \frac{m}{S} > \frac{1}{2}$. Otherwise, we say that the process is small.

12 A discrete time martingale in m-ary search tree processes

In this section, we give an introduction to random m-ary search trees, find a martingale and discuss some properties of this process concerning eigenvalues and eigenvectors. Furthermore, we show a relation to discrete time Pólya urn processes.

12.1 Definition of the process and a martingale

Definition 12.1 ([16])**.** Let $m \in \mathbb{N}$, $m \geq 3$. A complete m-ary tree is an ordered tree $\mathcal{T} \in \mathbb{U}_{\{0,1,\ldots,m-1\}} \coloneqq \{\mathcal{E}\} \cup \bigcup_{n\in\mathbb{N}} \{0,1,\ldots,m-1\}^n$ such that

$$\forall u \in \mathcal{T},\ u0 \in \mathcal{T} \iff u1 \in \mathcal{T} \iff \cdots \iff u(m-1) \in \mathcal{T}.$$

Definition 12.2 ([16])**.** Let $m \in \mathbb{N}$, $m \geq 3$. A sequence $(\mathcal{T}_n)_{n\geq 0}$ of growing random m-ary search trees is a sequence of growing random complete m-ary trees, constructed by successively inserting new keys into the leaves of the tree. Every node of the tree contains at most $m-1$ keys. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of keys, in our case, they are independent and identically distributed random variables taking values in the interval $[0,1]$.
The m-ary search tree $\mathcal{T}_n$ for $n \in \mathbb{N}_0$ is then recursively defined as follows. $\mathcal{T}_0$ is just the empty root. $\mathcal{T}_1$ is the root containing x_1. $\mathcal{T}_2$ is the root containing x_1 and x_2. $\mathcal{T}_{m-1}$ is the root, containing $x_1, \ldots, x_{m-1}$. All these $m-1$ keys are rearranged and stored in the sorted order

$$x_{(1)} < x_{(2)} < \cdots < x_{(m-1)}$$

in the node. When the $(m-1)$-th key is inserted, m empty subtrees of the root are created. Each subtree corresponds to one of the m intervals

$$I_1 = [0, x_{(1)}], I_2 = (x_{(1)}, x_{(2)}], \ldots, I_{m-1} = (x_{(m-2)}, x_{(m-1)}], I_m = (x_{(m-1)}, 1].$$

Each new key $x_m, \ldots$ belongs to a unique interval I_j and is now recursively inserted in the subtree which corresponds to the interval I_j. When there are $m-1$ keys in a node, m empty subtrees of this node are created.

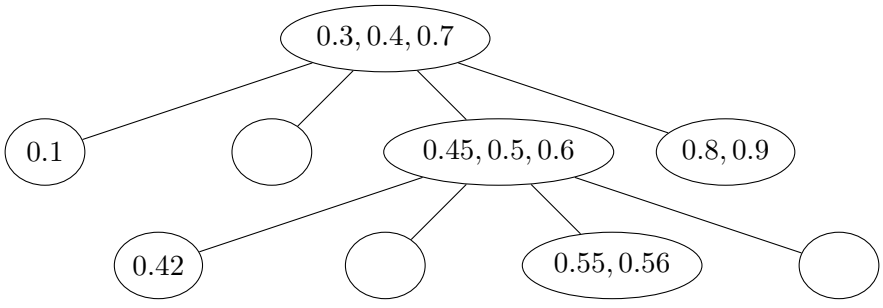

Figure 4: A 4-ary search tree constructed from the following sequence of keys: $x_1 = 0.3$, $x_2 = 0.4$, $x_3 = 0.7$, $x_4 = 0.5$, $x_5 = 0.6$, $x_6 = 0.9$, $x_7 = 0.1$, $x_8 = 0.45$, $x_9 = 0.8$, $x_{10} = 0.55$, $x_{11} = 0.42$, $x_{12} = 0.56$.

Definition 12.3 ([15])**.** Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be an m-ary search tree process. For $i \in \{1,\dots,m-1\}$ and $n \in \mathbb{N}_0$, we define by $X_n^{(i)}$ the number of nodes in $\mathcal{T}_n$ which contain $i-1$ keys. We call such a node a node of type i.

Remark 12.4 ([15])**.** Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be an m-ary search tree process. Let us consider a node of type i in the tree $\mathcal{T}_n$. We assume that the next key we insert in the tree is placed in this node. The key can be smaller or greater than all of the other keys in the node or it can lie between two keys. Altogether, we have i empty places in this particular node. Similar to binary search trees, we concentrate on the leaves of the tree, meaning the nodes of the tree containing at most $m-2$ keys.

Definition 12.5 ([15])**.** We call

$$X_n = \begin{pmatrix} X_n^{(1)} \\ \vdots \\ X_n^{(m-1)} \end{pmatrix}$$

the composition vector of the m-ary search tree $\mathcal{T}_n$.

Definition 12.6 ([16])**.** When the keys are i.i.d. random variables, the tree becomes a random m-ary search tree. We proceed as in the case of a binary search tree and do not mention the *random*, even though the randomness is always there. The insertion of a new key is uniformly among the free places and hence X_n is a random variable.

Remark 12.7 ([17])**.** Since the place of a newly inserted key x_{n+1} only depends on the tree $\mathcal{T}_n$, an m-ary search tree process as well as the corresponding composition vectors are discrete time Markov chains.

Definition 12.8. The size of an m-ary search tree is the number of inserted keys.

Remark 12.9 ([15])**.** We consider an m-ary search tree $\mathcal{T}_n$ of size n and insert the $(n+1)$-th key. In a node of type i, there are $i-1$ keys and i empty places. So we insert the key in a node of type i, $i = 1,\dots,m-1$, with probability

$$\frac{iX_n^{(i)}}{n+1}.$$

Consequently, the node becomes a node of type $i+1$ for $i = 1,\dots,m-2$ and it gives m nodes of type 1 when $i = m-1$.
For $i = 1,\dots,m-1$, let

$$\begin{aligned} \Delta_1 &= \big(-1,1,0,\dots\big)^T \\ \Delta_2 &= \big(0,-1,1,0,\dots\big)^T \\ &\vdots \\ \Delta_{m-2} &= \big(0,\dots,0,-1,1\big)^T \end{aligned}$$

$$\Delta_{m-1} = \big(m, 0, \ldots, 0, -1\big)^T.$$

It follows now that

$$\mathbb{P}(X_{n+1} + \Delta_i \mid X_n) = \frac{iX_n^{(i)}}{n+1}.$$

Definition 12.10 ([15])**.** The composition matrix of an m-ary search tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ is

$$A \coloneqq \begin{pmatrix} -1 & & & & & m(m-1) \\ 1 & -2 & & & & \\ & 2 & -3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & -(m-2) & \\ & & & & m-2 & -(m-1) \end{pmatrix}.$$

Theorem 12.11 ([15] Proposition 4.11)**.** *Let $m \in \mathbb{N}$, $m \geq 3$. Let $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ be an m-ary search tree process with corresponding composition vectors $(X_n)_{n\in\mathbb{N}_0}$ and composition matrix A, all as defined and denoted before. Let $\gamma_{x,0}(t) = 1$ and for $n \in \mathbb{N}$ we define*

$$\gamma_{x,n}(t) \coloneqq \prod_{j=0}^{n-1} \left(1 + \frac{t}{j+x}\right).$$

Then the process

$$\gamma_{1,n}(A)^{-1} X_n,\ n \in \mathbb{N}_0,$$

is a vectorial $\mathcal{F}_n$-martingale, where $(\mathcal{F})_{n\in\mathbb{N}_0}$ is the filtration defined by $\mathcal{F}_n = \sigma(\mathcal{T}_0, \ldots, \mathcal{T}_n)$.

Proof. First, we notice that

$$\sum_{i=1}^{m-1} \mathbb{P}(X_{n+1} + \Delta_i \mid X_n) = 1 = \sum_{i=1}^{m-1} \frac{iX_n^{(i)}}{n+1}.$$

Thus, we take the conditional expectation

$$\begin{aligned} \mathbb{E}(X_{n+1} \mid X_n) &= \sum_{i=1}^{m-1} (X_n + \Delta_i) \frac{iX_n^{(i)}}{n+1} \\ &= X_n \cdot \sum_{i=1}^{m-1} \frac{iX_n^{(i)}}{n+1} + \sum_{i=1}^{m-1} \Delta_i \cdot \frac{iX_n^{(i)}}{n+1} \\ &= X_n + \frac{1}{n+1} \cdot \sum_{i=1}^{m-1} \Delta_i \cdot i \cdot X_n^{(i)}. \end{aligned}$$

We want to show now that $\sum_{i=1}^{m-1} \Delta_i \cdot i \cdot X_n^{(i)} = A \cdot X_n$ to apply Theorem 4.1.

We compute

$$A \cdot X_n = \begin{pmatrix} -X_n^{(1)} + m(m-1) \cdot X_n^{(m-1)} \\ X_n^{(1)} - 2 \cdot X_n^{(2)} \\ 2 \cdot X_n^{(2)} - 3 \cdot X_n^{(3)} \\ \vdots \\ (m-3) \cdot X_n^{(m-3)} - (m-2) \cdot X_n^{(m-2)} \\ (m-2) \cdot X_n^{(m-2)} - (m-1) \cdot X_n^{(m-1)} \end{pmatrix}$$

and

$$\begin{aligned} \sum_{i=1}^{m-1} \Delta_i \cdot i \cdot X_n^{(i)} &= \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot 1 \cdot X_n^{(1)} + \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot 2 \cdot X_n^{(2)} + \dots \\ &\quad + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -1 \\ 1 \end{pmatrix} \cdot (m-2) \cdot X_n^{(m-2)} + \begin{pmatrix} m \\ 0 \\ \vdots \\ 0 \\ 0 \\ -1 \end{pmatrix} \cdot (m-1) \cdot X_n^{(m-1)} \\ &= \begin{pmatrix} -X_n^{(1)} + m(m-1) X_n^{(m-1)} \\ X_n^{(1)} - 2 \cdot X_n^{(2)} \\ 2 \cdot X_n^{(2)} - 3 \cdot X_n^{(3)} \\ \vdots \\ (m-3) \cdot X_n^{(m-3)} - (m-2) \cdot X_n^{(m-2)} \\ (m-2) \cdot X_n^{(m-2)} - (m-1) \cdot X_n^{(m-1)} \end{pmatrix}, \end{aligned}$$

which is what we wanted to show.
Together with the equation before, we can write the conditional expectation as

$$\mathbb{E}(X_{n+1} \mid X_n) = \left(I + \frac{A}{n+1} \right) X_n.$$

Since $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain, this equation becomes

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \left(I + \frac{A}{n+1} \right) X_n. \tag{12.1}$$

The $\mathcal{F}_n$-adaptation is trivial and applying Theorem 4.1 proves that $(\gamma_{1,n}(A)^{-1} X_n)_{n \in \mathbb{N}_0}$ is a vectorial $\mathcal{F}_n$-martingale since $\mathbb{E}(\gamma_{1,n}(A)^{-1} X_n)$ resp. $\mathbb{E}(|\gamma_{1,n}(A)^{-1} X_n|)$ can be written as a finite sum of vectors, so $\mathbb{E}(|\gamma_{1,n}(A)^{-1} X_n|) < \infty$. This completes the proof. □

12.2 Some linear algebra in m-ary search tree processes

Remark 12.12 ([19], [32])**.** A computation shows that the characteristic polynomial of the matrix A is

$$\mathcal{X}_A(\lambda) = \prod_{k=1}^{m-1} (\lambda + k) - m!. \tag{12.2}$$

All (complex) eigenvalues of A are simple and 1 is the eigenvalue with the greatest real part. The simplicity of the eigenvalues implies that A is diagonalizable over $\mathbb{C}$.
The most important result about the eigenvalues of A is that all eigenvalues different from 1 have a real part less than $\frac{1}{2}$ if and only if $m \leq 26$.
The reason for this is that the roots of the characteristic polynomial are regularly spaced points on the algebraic curve defined by the equation $\prod_{k=1}^{m-1} |\lambda + k|^2 = (m!)^2$.
In [32], the eigenvalues with the second largest real part for increasing m were computed, which shows this property computationally.

Notation 12.13 ([17])**.** We write λ_2 for the eigenvalue of the matrix A with largest real part not equal to 1, i.e. $\Re(\lambda_2) \geq \Re(\lambda)$ for all $\lambda \in \mathrm{Sp}(A) \setminus \{1\}$.

Definition 12.14 ([17])**.** We say that an m-ary search tree process is small when $m \leq 26$ resp. if $\Re(\lambda_2) < \frac{1}{2}$ and we call it large if $m \geq 27$ resp. if $\Re(\lambda_2) > \frac{1}{2}$. The case where $\Re(\lambda_2) = \frac{1}{2}$ can not occur due to the characteristic polynomial of A.

Definition 12.15 ([17])**.** We write u_1 and u_2 for the following two linear eigenforms of A, respectively associated with the eigenvalues 1 and λ_2, meaning that $u_1 \circ A = u_1$ and $u_2 \circ A = \lambda_2 u_2$:

$$u_1 \begin{pmatrix} x_1 \\ \vdots \\ x_{m-1} \end{pmatrix} = \sum_{k=1}^{m-1} k x_k, \; u_2 \begin{pmatrix} x_1 \\ \vdots \\ x_{m-1} \end{pmatrix} = \sum_{k=1}^{m-1} \binom{\lambda_2 + k - 1}{k-1} x_k. \tag{12.3}$$

The corresponding vectors v_1 and v_2 of the dual basis of u_1 and u_2 are:

$$v_1 = \frac{1}{H_m(1)} \left(\frac{1}{k(k+1)} \right)_{1 \leq k \leq m-1}, \; v_2 = \frac{1}{H_m(\lambda_2)} \left(\frac{1}{k \binom{\lambda_2 + k}{k}} \right)_{1 \leq k \leq m-1}, \tag{12.4}$$

where

$$H_m(z) = \sum_{k=1}^{m-1} \frac{1}{z+k}.$$

These are eigenvectors of A and the projections on the eigenlines are $u_1 v_1$ and $u_2 v_2$.

Remark 12.16 ([17])**.** The eigenvectors and linear forms satisfy $u_1(v_1) = u_2(v_2) = 1$ and $u_1(v_2) = u_2(v_1) = 0$. Since λ_2 is not real, v_2 and u_2 have coefficients with imaginary parts. In particular, $\overline{v_2}$ is an eigenvector and $\overline{u_2}$ is an eigenform to the eigenvalue $\overline{\lambda_2}$ of A. We can complete $(v_1, v_2, \overline{v_2})$ to a basis of eigenvectors of A. The dual basis is then of the form $(u_1, u_2, \overline{u_2}, \dots)$ and $u_2(\overline{v_2}) = 0$.

12.3 The associated Pólya urn process

Construction 12.17 ([16], [41])**.** We can assign a Pólya urn to an m-ary search tree, resp. we can associate an m-ary search tree with a Pólya urn. The corresponding urn is a Pólya urn consisting of $m-1$ colours. The colour of type i corresponds to the nodes of type i for $i = 1, \ldots, m-1$. The number of balls of type i is given by the number of free places for the keys in the nodes of type i, for $i = 1, \ldots, m-1$.
If we assume that an m-ary search tree has k nodes of type i and each node of type i has i free places, then there are ki free places of type i, meaning the associated Pólya urn contains ki balls of type i, for $i = 1, \ldots, m-1$.
Let us now assume that $i = 1, \ldots, m-2$. If we have drawn a ball of type i, this means we insert a key in the m-ary search tree in a node of type i. The node becomes a node of type $i+1$, having $i+1$ free places. In terms of a Pólya urn, this means that we lose i free places of type i, i.e. we remove i balls of type i and we add $i+1$ balls of type $i+1$. If we draw a ball of type $m-1$, this means that we insert a key in a node containing $m-2$ keys, so this node becomes an internal node with m leaves, which are nodes of type 1. Hence we remove $m-1$ empty places of type $m-1$ and add m free places of type 1, which corresponds to removing $m-1$ balls of type $m-1$ and adding m balls of type 1.

Definition 12.18 ([18])**.** The Pólya urn process $(U_n)_{n\in\mathbb{N}_0}$ associated to the m-ary search tree process given by the composition vectors $(X_n)_{n\in\mathbb{N}_0}$ is defined by

$$U_n^{(i)} = i \cdot X_n^{(i)},\ i = 1, \ldots, m-1,\ n \in \mathbb{N}_0.$$

The corresponding replacement matrix of the urn process is

$$R = \begin{pmatrix} -1 & 2 & & & & & \\ & -2 & 3 & & & & \\ & & -3 & 4 & & & \\ & & & \ddots & \ddots & & \\ & & & & -(m-2) & m-1 & \\ m & & & & & -(m-1) \end{pmatrix}.$$

Remark 12.19 ([18])**.** We notice that the Pólya urn process associated to the m-ary search tree process is balanced, with balance $S = 1$.

Remark 12.20 ([18])**.** Let $P = (p_{ij})_{1\le i,j\le m-1}$ be the matrix defined by

$$p_{ij} = \begin{cases} i \text{ if } i = j; \\ 0 \text{ else.} \end{cases}$$

Then $U_n = P \cdot X_n$. Moreover, it follows that $A = P^{-1}R^TP$, so the composition matrix of an m-ary search tree process and the replacement matrix of the associated Pólya urn process have the same characteristic polynomial and thus the same eigenvalues.

13 Asymptotic behaviour of Pólya urns and m-ary search trees in discrete time part I

We introduce Pólya processes as a generalisation of discrete time Pólya urn processes and hence as a generalisation of discrete time m-ary search tree processes. We prove a statement about the asymptotic behaviour of Pólya processes and apply it afterwards to discrete time Pólya urn processes as well as to m-ary search tree processes.

13.1 The Pólya process

Remark 13.1 ([44])**.** Since we view m-ary search trees as Pólya urns, we consider the case of an s-dimensional Pólya urn process, where $s \in \mathbb{N}$. A Pólya urn process can be seen as a random walk in the first quadrant of $\mathbb{R}^s$, $s \in \mathbb{N}$, with finitely many possible increments (the rows of the replacement matrix R). This random walk is a Markov chain in discrete time. Conditioned on the process at time n, we notice that the transition probabilities (the probabilities of the process to get from its current state at time n to the possible states at time $n+1$) are linear functions of the coordinates of the composition vector at time n.
We generalize this to a so-called Pólya process. We deal with a random walk in $\mathbb{R}^s$, $s \in \mathbb{N}$, with normalized balance (in terms of Pólya urn processes $S = 1$). Instead of counting a number of balls, we have a quantity $l_k(U_n) \in \mathbb{R}_{\geq 0}$ which is associated with each colour $1 \leq k \leq s$ (and which corresponds to the number of balls of colour k at time n), which gives the propensity to pick this colour when colour k has been drawn. We have vectors $w_k \in \mathbb{R}^s$, defined by the fact that, when colour k has been drawn, then for all $j \in \{1, \ldots, s\}$, we add $l_j(w_k)$ balls of colour j to the urn.

Definition 13.2 ([44])**.** Let V be a vector space over $\mathbb{R}$ of finite dimension $s \in \mathbb{N}$. Let X_0, $w_1, \ldots, w_s$ be vectors in V and $(l_k)_{1 \leq k \leq s}$ be a basis of linear forms on V satisfying the following assumptions:

(i) (initialization hypothesis)

$$X_0 \neq 0 \text{ and } \forall k \in \{1, \ldots, s\},\ l_k(X_0) \geq 0; \tag{13.1}$$

(ii) (balance hypothesis) for every $k \in \{1, \ldots, s\}$,

$$\sum_{j=1}^{s} l_j(w_k) = 1; \tag{13.2}$$

(iii) (sufficient conditions of tenability) for all $k, k' \in \{1, \ldots, s\}$,

$$\begin{cases} k \neq k' \longrightarrow l_k(w_{k'}) \geq 0, \\ l_k(w_k) \geq 0 \text{ or } l_k(X_0)\mathbb{Z} + \sum_{j=1}^{s} l_k(w_j)\mathbb{Z} = l_k(w_k)\mathbb{Z}. \end{cases} \tag{13.3}$$

The (discrete and finite dimensional) Pólya process associated with these data is the V-valued random walk $(X_n)_{n\in\mathbb{N}_0}$, with increments in the finite set $\{w_1,\dots,w_s\}$, defined by X_0 and the induction: for every $n\in\mathbb{N}_0$ and $k\in\{1,\dots,s\}$,

$$\mathbb{P}(X_{n+1}=X_n+w_k\mid X_n)=\frac{l_k(X_n)}{n+\tau_0}, \tag{13.4}$$

where τ_0 is the positive real number defined by

$$\tau_0=\sum_{k=1}^{s} l_k(X_0). \tag{13.5}$$

Remark 13.3 ([44])**.** The process is defined on the space of all trajectories of $X_0+\sum_{1\le k\le s}\mathbb{N}_0 w_k$, endowed with the natural filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, where $\mathcal{F}_n=\sigma(X_0,\dots,X_n)$. It is a Markov chain and the transition conditional probabilities between times n and $n+1$ depend linearly on the state at time n, as stated in Equation 13.4. The conditions 13.1 and 13.2 are necessary and sufficient for the random vector X_1 to be well-defined by Equation 13.4. Condition 13.3 guarantees that the process is well-defined, i.e. the numbers $l_k(X_n)$ do not become negative, which means the process does not go extinct.

Lemma 13.4 ([44])**.** *For all $n\in\mathbb{N}_0$, we have*

$$\sum_{k=1}^{s} l_k(X_n)=n+\tau_0. \tag{13.6}$$

Proof. We use induction on n. For $n=0$, this is just the definition, so we assume that it holds for n and prove it for $n+1$. By condition 13.2, we have for all $j\in\{1,\dots,s\}$,

$$\begin{aligned}\sum_{k=1}^{s} l_k(X_{n+1}) &= \sum_{k=1}^{s} l_k(X_n)+l_k(w_j)=n+\tau_0+\sum_{k=1}^{s} l_k(w_j)\\ &= n+1+\tau_0,\end{aligned}$$

which is what we wanted. Therefore, it holds for all $n\in\mathbb{N}_0$. □

Notation 13.5 ([44])**.** For a real vector space V, we denote the complexified space of V by $V_{\mathbb{C}}=V\otimes_{\mathbb{R}}\mathbb{C}$.

Definition 13.6 ([44])**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a Pólya process. Its corresponding replacement endomorphism is the endomorphism $A=\sum_{1\le k\le s} l_k\otimes w_k\in V^*\otimes V\cong\mathrm{End}(V)$, defined as

$$A(v)=\sum_{1\le k\le s} l_k(v)w_k,$$

for all $v\in V$.
By choosing a basis V, we can view this endomorphism as a matrix A.

Lemma 13.7 ([44]). *Let $\gamma_{x,0}(t) = 1$ and for all $n \in \mathbb{N}$,*

$$\gamma_{x,n}(t) \coloneqq \prod_{k=0}^{n-1} \left(1 + \frac{t}{k+x}\right).$$

Then

$$\mathbb{E}(X_n) = \gamma_{\tau_0,n}(A)(X_0).$$

Proof. Similar computations as in the proof of Theorem 11.3 show that

$$\mathbb{E}(X_{n+1} \mid X_n) = \left(I + \frac{A}{n+\tau_0}\right) X_n.$$

We apply the expectation on both sides of the equation, giving us

$$\mathbb{E}(X_{n+1}) = \left(I + \frac{A}{n+\tau_0}\right) \mathbb{E}(X_n).$$

Now, we apply induction and get

$$\mathbb{E}(X_n) = \gamma_{\tau_0,n}(A)(X_0).$$

□

Remark 13.8 ([44]). Let $u_1 = \sum_{k=1}^{s} l_k$. Condition 13.2 in the definition of a Pólya process implies that the linear form u_1 satisfies $u_1 \circ A = u_1$, which shows that 1 is always an eigenvalue of A.

Proposition 13.9 ([44] Proposition 2.2). *Any complex eigenvalue λ of A equals 1 or satisfies $\Re(\lambda) < 1$. Moreover, $\dim(\ker(A-1))$ is equal to the multiplicity of 1 as eigenvalue of A.*

Sketch of the proof. We replace A by its matrix in the dual basis of $(l_k)_{1 \le k \le s}$. We consider first the case where all entries of A are non-negative and show the statement using boundedness. For the second case, we consider Jordan's decomposition of A. More details can be found in the proof of Proposition 2.2 in [31]. □

Notation 13.10 ([44]). For a Pólya process with replacement endomorphism A, we denote by σ_2 the real number ≤ 1 defined by

$$\sigma_2 = \begin{cases} 1 \text{ if 1 is a multiple eigenvalue of } A; \\ \max\{\Re(\lambda),\ \lambda \in \mathrm{Sp}(A),\ \lambda \neq 1\} \text{ otherwise.} \end{cases} \tag{13.7}$$

13.2 A Jordan basis of linear forms of a Pólya process

Definition 13.11 ([44])**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a Pólya process of dimension s. A basis $(u_k)_{1\leq k\leq s}$ of linear forms on $V_{\mathbb{C}}$ is called a Jordan basis of linear forms of this process, or abbreviated a Jordan basis, if

(i) $u_1 = \sum_{1\leq k\leq s} l_k$;

(ii) $u_k \circ A = \lambda_k u_k + \varepsilon u_{k-1}$ for all $k \geq 2$, where the $\lambda_k \in \mathbb{C}$ are the eigenvalues of A and the ε_k are numbers in $\{0, 1\}$ satisfying $\lambda_k \neq \lambda_{k-1} \Longrightarrow \varepsilon_k = 0$.

Remark 13.12 ([44])**.** Said differently, the matrix of the transposed endomorphism A^T in a Jordan basis of linear forms has a block-diagonal form
$\mathrm{Diag}(1, J_{p_1}(\lambda_{k_1}), \ldots, J_{p_t}(\lambda_{k_t}))$, with $J_p(z)$ being the $(p \times p)$-matrix

$$J_p(z) = \begin{pmatrix} z & 1 & & \\ & z & \ddots & \\ & & \ddots & 1 \\ & & & z \end{pmatrix}.$$

Definition 13.13 ([44])**.** We recall that an eigenform of the matrix A is a (real or complex) linear form u_k if $u_k \circ A = \lambda_k u_k$, i.e. if $\varepsilon_k = 0$. An eigenform of the process is an eigenvector of A^T (sometimes it is also called a left eigenvector of A) (see Definition 2.11).

Definition 13.14 ([44])**.** We choose a Jordan basis of linear forms with notations as before. A subset $J \subseteq \{1, \ldots, s\}$ is called a monogenic block of indices, when J is of the form $J = \{m, m+1, \ldots, m+r\}$ ($r \geq 0$, $m \geq 1$, $m + r \leq s$) with $\varepsilon_m = 0$, $\varepsilon_k = 1$ for every $k \in \{m+1, \ldots, m+r\}$ and J is maximal for this property. Any monogenic block of indices J is associated with a unique eigenvalue of A that is denoted by $\lambda(J)$.

Remark 13.15 ([44])**.** Said differently, J is monogenic when the subspace $\mathrm{Span}(u_j, j \in J)$ is A-stable and when the matrix of the endomorphism of $\mathrm{Span}(u_j, j \in J)$ induced by A^T in the Jordan basis is one of the Jordan blocks mentioned above with number $\lambda(J)$ on its diagonal.

Definition 13.16 ([44])**.** A monogenic block of indices J is called a principal block when $\Re(\lambda(J)) = \sigma_2$ and J has maximal size among the monogenic blocks J' such that $\Re(\lambda(J')) = \sigma_2$.

Definition 13.17 ([44])**.** We consider the Pólya process with notations as before and with a Jordan basis $(u_k)_{1\leq k\leq s}$ of linear forms. We denote by $(v_k)_{1\leq k\leq s}$ its dual basis, meaning that $v_k \in V_{\mathbb{C}}$ for all $1 \leq k \leq s$ such that $u_k(v_l) = \delta_{k,l}$ (Kronecker notation) for all k and l. We define the s-tuple $\lambda = (\lambda_1, \ldots, \lambda_s)$ containing the eigenvalues (distinct or not) respectively associated with $u_1, \ldots, u_s$ (or $v_1, \ldots, v_s$). In particular, $\lambda_1 = 1$ for any Jordan basis of linear forms. We call the eigenvalues $\lambda_1, \ldots, \lambda_s$ the roots of the process. For any k, we denote by π_k the projection on the line $\mathbb{C}v_k$ relative to the decomposition $V_{\mathbb{C}} = \bigoplus_{1\leq k\leq s} \mathbb{C}v_k$.

Remark 13.18 ([44]). The projections satisfy $\mathrm{Id} = \sum_{1\leq k\leq s} \pi_k$ and $\pi_k = u_k v_k$. The projections commute with each other ($\pi_k \pi_l = \delta_{k,l}\pi_k$) but do not commute with A. However, A does commute with $\sum_{j\in J} \pi_j$, the sum being extended to any monogenic block of indices J (these sums are polynomials in A).

13.3 Semisimplicity, large and small projections

Definition 13.19 ([44]). We consider the Pólya process with notations as before and with a Jordan basis $(u_k)_{1\leq k\leq s}$ of linear forms. For $\alpha = (\alpha_k)_{1\leq k\leq s} \in \mathbb{Z}^s$, we use the notations

$$|\alpha| = \sum_{1\leq k\leq s} \alpha_k \text{ and } \langle \alpha, \lambda\rangle = \sum_{1\leq k\leq s} \alpha_k \lambda_k.$$

When all α_k are non-negative, we define the homogeneous polynomial function of degree $|\alpha|$ as

$$u^\alpha = \prod_{1\leq k\leq s} u_k^{\alpha_k}.$$

Definition 13.20 ([44]). We say that a Pólya process is semisimple if its replacement endomorphism A is semisimple, i.e. if A admits a basis of eigenvectors in $V_{\mathbb{C}}$ (meaning that all the u_k are real or complex eigenforms of A). The process is said to be principally semisimple if all principle blocks have size one (for any choice of Jordan basis).

Definition 13.21 ([44]). A root of the Pólya process is said to be small if its real part is $\leq \frac{1}{2}$; otherwise, we call it large. We say that the process is small if $\sigma_2 \leq \frac{1}{2}$, which means that 1 is a simple root and all other roots are small; otherwise, we call it a large process.

Definition 13.22 ([44]). We consider the Pólya process with notations as before and with a Jordan basis $(u_k)_{1\leq k\leq s}$ of linear forms. Let $\alpha = (\alpha_k)_{1\leq k\leq s} \in \mathbb{N}_0^s$.

(i) We say that α is a power of large projections if u^α is a product of linear forms associated with large roots, i.e. if for every $k \in \{1, \ldots, s\}$, $\alpha_k \neq 0 \Longrightarrow \Re(\lambda_k) > \frac{1}{2}$.

(ii) We say that α is a power of small projections if u^α is a product of linear forms associated with small roots, i.e. if for every $k \in \{1, \ldots, s\}$, $\alpha_k \neq 0 \Longrightarrow \Re(\lambda_k) \leq \frac{1}{2}$.

(iii) We say that α is a semisimple power if u^α is a product of eigenforms, i.e. if for every $k \in \{1, \ldots, s\}$, $\alpha_k \neq 0 \Longrightarrow u_k$ is an eigenform of the process.

(iv) We say that α is a monogenic power if its support is contained in a monogenic block of indices.

Notation 13.23 ([44]). We recall that $(e_k)_{1\leq k\leq s}$ is the canonical basis of $\mathbb{Z}^s$ (or $\mathbb{R}^s$) (see Notation 2.8).

Notation 13.24 ([44]). We denote by $\alpha \leq \beta$ for $\alpha, \beta \in \mathbb{N}_0^s$ the degree-antialphabetical (total) order, defined by $\alpha = (\alpha_1, \ldots, \alpha_s) < \beta = (\beta_1, \ldots, \beta_s)$ when $|\alpha| < |\beta|$ or $|\alpha| = |\beta|$ and $\exists r \in \{1, \ldots, s\}$ such that $\alpha_r < \beta_r$ and $\alpha_t = \beta_t$ for any $t > r$.

We have $e_1 < e_2 < \cdots < e_s < 2e_1 < e_1 + e_2 \ldots$.
When $\alpha = (\alpha_1, \ldots, \alpha_s) \in \mathbb{R}^s$, the inequality $\alpha \geq 0$ means that all α_k are non-negative.

Definition 13.25 ([44])**.** The transition operator Φ associated to a Pólya process with notations as before is the map defined on the space of all functions $f\colon V \to \mathbb{R}$ (or more generally on the space of all functions $f\colon V \to W$, where W is any vector space over $\mathbb{R}$, by

$$\Phi(f)(v) = \sum_{1 \leq k \leq s} l_k(v)(f(v + w_k) - f(v)),\ v \in V.$$

13.4 Asymptotics of a Pólya process

Proposition 13.26 ([44] Proposition 3.1)**.** *For any choice of Jordan basis $(u_k)_{1 \leq k \leq s}$ of linear forms of a Pólya process and for every $\alpha \in \mathbb{N}_0^s$,*

$$\Phi(u^\alpha) - \langle \alpha, \lambda \rangle u^\alpha \in \mathrm{Span}\{u^\beta \mid \beta < \alpha\}.$$

Sketch of the proof. By definition of a Pólya process, all vectors w_k are non-zero. Therefore we can consider the differential operator Φ_∂ defined by

$$\Phi_\partial = \sum_{1 \leq k \leq s} l_k(v) \frac{\partial f}{\partial w_k}(v),$$

for every function $f\colon V \to W$ (W being any $\mathbb{R}$-vector space) and derivable at each point $v \in V$ along the directions carried by the vectors w_k, $1 \leq k \leq s$. We can prove the same statement for this operator, i.e. for every $\alpha \in \mathbb{N}_0^s$,

$$\Phi_\partial(u^\alpha) - \langle \alpha, \lambda \rangle u^\alpha \in \mathrm{Span}\{u^\beta \mid \beta < \alpha\},$$

which then, together with the Taylor formula, completes the proof. All details can be found in [44] in the proof of Proposition 3.1. □

Remark 13.27 ([44])**.** It follows from this proposition that the finite-dimensional polynomial subspace $S_\alpha = \mathrm{Span}\{u^\beta \mid \beta \leq \alpha\}$ for any $\alpha \in \mathbb{N}_0^s$ is Φ-stable. These subspaces form an increasing sequence whose union is the space $S(V)$ of all polynomial functions on V, so that Proposition 13.26 asserts that the eigenvalues of Φ on $S(V)$ are exactly all numbers $\langle \alpha, \lambda \rangle$, $\alpha \in \mathbb{N}_0^s$ (in the (ordered) basis $(u^\beta)_{\beta \leq \alpha}$ of any S_α, the matrix of Φ is triangular).

Notation 13.28 ([44])**.** If Ψ is an endomorphism of a vector space, we denote by $\ker(\Psi^\infty)$ the characteristic space of Ψ associated with zero, given by

$$\ker(\Psi^\infty) = \bigcup_{p \geq 0} \ker(\Psi^p).$$

Notation 13.29 ([44])**.** We write Φ for the operator as well as for the endomorphism induced by Φ on $S(V)$ or on some stable subspace.

Remark 13.30 ([44])**.** We can decompose all S_α as direct sums of characteristic subspaces of Φ, giving us the splitting

$$S(V) = \sum_{z \in \mathbb{C}} \ker((\Phi - z)^\infty).$$

Definition 13.31 ([44])**.** We consider the Pólya process with notations as before and with a Jordan basis $(u_k)_{1 \le k \le s}$ of linear forms. For any $\alpha \in \mathbb{N}_0^s$, the reduced polynomial of rank α is the projection of u^α on $\ker((\Phi - \langle \alpha, \lambda \rangle)^\infty)$ parallel to $\sum_{z \neq \langle \alpha, \lambda \rangle} \ker((\Phi - z)^\infty)$. We denote it by Q_α.

Definition 13.32 ([44])**.** For every $\alpha \in \mathbb{N}_0^s$, the non-negative integer ν_α is defined by

$$\nu_\alpha = \max\{p \ge 0, (\Phi - \langle \alpha, \lambda \rangle)^p (Q_\alpha) \neq 0\}.$$

Theorem 13.33 ([44] Theorem 3.4)**.** *Let $(X_n)_{n \in \mathbb{N}_0}$ be a Pólya process with a Jordan basis $(u_k)_{1 \le k \le s}$ of linear forms. Let $\alpha \in \mathbb{N}_0^s$.*

(i) If α is a power of small projections, then there exists some $\nu \in \mathbb{N}_0$ such that

$$\mathbb{E}(u^\alpha(X_n)) \in O\left(n^{\frac{|\alpha|}{2}} \log^\nu(n)\right)$$

as $n \to \infty$.

(ii) If α is a power of small projections, then there exists a $c \in \mathbb{C}$ such that

$$\mathbb{E}(u^\alpha(X_n)) = c n^{\langle \alpha, \lambda \rangle} \log^{\nu_\alpha}(n) + o\left(n^{\Re(\langle \alpha, \lambda \rangle)} \log^{\nu_\alpha}(n)\right)$$

as $n \to \infty$.

(iii) If α is a semisimple power of large projections, then

$$\mathbb{E}(u^\alpha(X_n)) = n^{\langle \alpha, \lambda \rangle} \frac{\Gamma(\tau_0)}{\Gamma(\tau_0 + \langle \alpha, \lambda \rangle)} Q_\alpha(X_0) + o\left(n^{\Re(\langle \alpha, \lambda \rangle)}\right)$$

as $n \to \infty$, where Q_α is the reduced polynomial of rank α relative to the Jordan basis $(u_k)_{1 \le k \le s}$.

Sketch of the proof. The proof of Theorem 13.33 is done in several steps. First of all, the transition operator is studied, as well as how we can compute the moments of $f(X_n)$ for a function $f \colon V \to W$, where W is a vector space over $\mathbb{R}$. Moreover, we consider the asymptotic behaviour of these moments, especially in the case where f is non-zero and belongs to a Φ-stable subspace of measurable functions $V \to W$. Then, the operator on this subspace can be written as $zI + \Phi_N$, where I is the identity operator on this subspace, $z \in \mathbb{C}$ and Φ_N is a non-zero nilpotent operator on this subspace.
In the next step, we look at the action of Φ on polynomials, especially on the reduced polynomials. For this, the differential operator is used. We introduce cones and polyhedrons to study the action of Φ on the reduced polynomials even further. Finally, we can use the asymptotics of the reduced polynomials. All details can be found in [44] in the proof of Theorem 3.4. □

We finally state and prove the theorem about the asymptotic behaviour of a large Pólya process.

Theorem 13.34 ([44] Theorem 3.5)**.** *Suppose that a Pólya process $(X_n)_{n\in\mathbb{N}_0}$ is large and principally semisimple. Fix a Jordan basis $(u_k)_{1\le k\le s}$ of linear forms such that $u_1,\dots,u_r$ $(2\le r\le s)$ are all the eigenforms of the basis that are associated with roots $\lambda_1=1,\lambda_2,\dots,\lambda_r$ having a real part $\ge\frac{1}{2}$ (meaning that if 1 is a multiple root, $\lambda_1=\dots=\lambda_r=1$, otherwise $\frac{1}{2}<\Re(\lambda_2)=\dots=\Re(\lambda)=\sigma_2<1$).*
Then, there exist unique (complex-valued) random variables $W_2,\dots,W_r$ such that

$$X_n = nv_1 + \sum_{2\le k\le r} n^{\lambda_k} W_k v_k + o(n^{\sigma_2})$$

as $n\to\infty$, where the small o is almost surely and in L^p for every $p\ge 1$. Furthermore, if one denotes by $(Q_\alpha)_{\alpha\in\mathbb{N}_0^s}$ the reduced polynomials relative to the Jordan basis $(u_k)_{1\le k\le s}$, all joint moments of the random variables $W_2,\dots,W_r$ exist and are given by

$$\mathbb{E}\left(\prod_{2\le k\le r} W_k^{\alpha_k}\right) = \frac{\Gamma(\tau_0)}{\Gamma(\tau_0+\langle\alpha,\lambda\rangle)} Q_\alpha(X_0)$$

for all $\alpha_2,\dots,\alpha_r\in\mathbb{N}_0$ and with $\alpha=\sum_{2\le k\le r}\alpha_k e_k=(0,\alpha_2,\dots,\alpha_r,\dots)$.

Proof. We define $\pi=\sum_{2\le k\le r}\pi_k$ and $\pi'=\sum_{k\ge r+1}\pi_k$, where π_k are the projections defined in Definition 13.17. We decompose the vector X_n into the sum

$$X=\pi_1 X_n + Y_n + Z_n,$$

where $Y_n=\pi X_n$ and $Z_n=\pi' X_n$. We consider these terms separately.

(i) We consider the first term $\pi_1 X_n$.
By the definition of the Jordan basis $(u_k)_{1\le k\le s}$ of linear forms and of its dual basis $(v_k)_{1\le k\le s}$ of vectors, we have $\pi_1(v)=u_1(v)v_1$ for all $v\in V$. Equation 13.6 leads to $\pi_1 X_n = nv_1+O(1)$ as $n\to\infty$.

(ii) We consider the second term Y_n.
By definition of the Jordan basis of linear forms and of its dual vectors, we have $Y_n=\sum_{2\le k\le r}u_k(X_n)v_k$. We fix $k\in\{2,\dots,r\}$ and compute the expectation of $u_k(X_{n+1})$ conditionally to the state at time n. This gives us

$$\mathbb{E}(u_k(X_{n+1})\mid\mathcal{F}_n)=\left(1+\frac{\lambda_k}{n+\tau_0}\right)u_k(X_n)$$

for all $n\in\mathbb{N}_0$ (see proof of Lemma 13.7 for a similar argument). Therefore,

$$(\gamma_{\tau_0,n}(\lambda_k)^{-1}u_k(X_n))_{n\in\mathbb{N}_0}$$

is a martingale (we can divide by $\gamma_{\tau_0,n}(\lambda_k)$ since λ_k is not a negative integer).

We know that $\overline{u_k}$ is an eigenform associated to the eigenvalue $\overline{\lambda_k}$. Thus, it is a linear combination of eigenforms u_l's, all associated with $\overline{\lambda_k}$. Hence, if $q \in \mathbb{N}$, then $|u_k^{2q}|$ is a linear combination of polynomials u^α's for some suitable semisimple powers of large projections α's such that $\langle \alpha, \lambda \rangle = 2q\sigma_2$. Then, part (iii) in Theorem 13.33 leads to

$$\mathbb{E}(|u_k^{2q}(X_n)|) = O(n^{2q\sigma_2}).$$

This is also valid if $\lambda_k \in \mathbb{R}$. It follows now that the martingales $(\gamma_{\tau_0,n}(\lambda)^{-1} u_k(X_n))_{n\in\mathbb{N}_0}$ are all convergent in L^p for $p \geq 1$.
For $k \in \{2, \dots, r\}$, let W_k be the (complex-valued) random variables defined by

$$W_k = \lim_{n\to\infty} \frac{u_k(X_n)}{\gamma_{\tau_0,n}(\lambda_k)} \frac{\Gamma(\tau_0)}{\Gamma(\tau_0+\lambda_k)} = \lim_{n\to\infty} u_k\left(\frac{X_n}{n^{\lambda_k}}\right),$$

where we used Stirling's asymptotics as $n \to \infty$ in the second equality:

$$\gamma_{\tau_0,n}(\lambda)\Gamma(\tau_0+\lambda) = \Gamma(\tau_0) n^\lambda (1+o(1)),$$

for every $\lambda \notin \{-\tau_0 - q \mid q \in \mathbb{N}_0\}$ (see Lemma 8.5 for a similar computation). This proves that

$$Y_n = \sum_{2\leq k\leq r} n^{\lambda_k} W_k v_k + o(n^{\sigma_2}),$$

the small o being almost sure and in L^P for $p \geq 1$.
We compute the limit of the joint moments.
If $\alpha = (0, \alpha_2, \dots, \alpha_r, 0, \dots) \in \mathbb{N}_0^s$, α is a semisimple power of large projections and if we write $W^\alpha = \prod_k W_k^{\alpha_k}$, then by Theorem 13.33,

$$\mathbb{E}(W^\alpha) = \lim_{n\to\infty} \frac{1}{n^{\langle\alpha,\lambda\rangle}} \mathbb{E}(u^\alpha(X_n)) = \frac{\Gamma(\tau_0)}{\Gamma(\tau_0+\langle\alpha,\lambda\rangle)} Q_\alpha(X_0).$$

(iii) We consider the third term Z_n.
We have $Z_n = \sum_{k>r} u_k(X_n) v_k$. Our goal is to show that $n^{-\sigma_2} u_k(X_n)$ converges to 0 almost surely and in L^p, $p \geq 1$, for all $k > r$. We fix a $k > r$ and an integer $q \geq 1$. As in the second case, $\overline{u_k}$ is a linear combination of u_l's, all associated with the root $\overline{\lambda_k}$ (even if u_k and the u_l's are not necessarily eigenforms). Therefore, $|u_k^{2q}|$ is a linear combination of polynomials u^α's for some suitable α's that are powers of large (resp. small) projections if $\Re(\lambda_k) > \frac{1}{2}$ (resp. if $\Re(\lambda_k) \leq \frac{1}{2}$), such that $\langle \alpha, \lambda \rangle = 2q\Re(\lambda_k)$. Now, Theorem 13.33 implies in both cases that

$$\mathbb{E}(|u_k^{2q}(X_n)|) \in o(n^{2q\sigma_2}),$$

which gives us the L^q-convergence. Moreover, let $p \in \mathbb{N}$ such that $\frac{1}{p} < 2(\sigma_2 - \Re(\lambda_k))$ if $\Re(\lambda_k) > \frac{1}{2}$ or such that $\frac{1}{p} < 2\sigma_2 - 1$ if $\Re(\lambda_k) \leq \frac{1}{2}$. For such a p, the series

$$\sum_{n\in\mathbb{N}} \mathbb{E}\left(\left|\frac{1}{n^{\sigma_2}} u_k(X_n)\right|^{2p}\right)$$

converges. The almost sure convergence to 0 of $n^{-\sigma_2}u_k(X_n)$ follows from the almost sure convergence of the series of non-negative random variables

$$\sum_{n\in\mathbb{N}}\left|\frac{1}{n^{\sigma_2}}u_k(X_n)\right|^{2p}.$$

This completes the proof. □

The following results about reduced polynomials are used later in the thesis.

Proposition 13.35 ([44] Proposition 4.8)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a Pólya process with a Jordan basis $(u_k)_{1\le k\le s}$ of linear forms. Let $\alpha\in\mathbb{N}_0^s$.*

(i) $Q_0 = 1$ *and* $Q_\alpha = u^\alpha$ *if* $|\alpha| = 1$*;*

(ii) $\{Q_\beta \mid \beta\le\alpha\}$ *is a basis of* S_α*;*

(iii) for every $z\in\mathbb{C}$, $\{Q_\alpha \mid \langle\alpha,\lambda\rangle = z\}$ *is a basis of* $\ker((\Phi - z)^\infty)$*;*

(iv) $Q_\alpha - u^\alpha \in \mathrm{Span}\{Q_\beta \mid \beta<\alpha, \langle\beta,\lambda\rangle\neq\langle\alpha,\lambda\rangle\}$*;*

(v) $\Phi(Q_\alpha) - \langle\alpha,\lambda\rangle Q_\alpha \in \mathrm{Span}\{Q_\beta \mid \beta<\alpha, \langle\beta,\lambda\rangle = \langle\alpha,\lambda\rangle\}$.

Sketch of the proof. The statements follow from the definition of the reduced polynomials, the Jordan basis and as a consequence of Proposition 13.26. More details can be found in [44] in the proof of Proposition 4.8. □

Remark 13.36 ([44])**.** A consequence of this proposition is that for $p\in\mathbb{N}_0$, $Q_{pe_1} = u_1(u_1+1)\dots(u_1+p-1)$. Moreover, the Q_α can be inductively computed for any $\alpha\in\mathbb{N}_0^s$. This means that the expectation of the W_k in Theorem 13.34 can be inductively computed.

13.5 The asymptotics of a large Pólya urn process in discrete time

Construction 13.37. Let $(U_n)_{n\in\mathbb{N}_0}$ be the discrete time Pólya urn process with all notations as in section 11. We want to construct a Pólya process $(U_n^{PP})_{n\in\mathbb{N}_0}$ from $(U_n)_{n\in\mathbb{N}_0}$. The balance of the Pólya urn process is $S = a+b = c+d$, so we take our replacement matrix R and divide it by S, then the sums of the row entries are equal to 1 in every row. Moreover, we divide the vector U_n by S.
Let

$$U_n^{PP} := \frac{1}{S}U_n,\ R^{PP} := \frac{1}{S}R.$$

The vectors w_k from the definition of a Pólya process (see Definition 13.2) are now the k-th row of matrix R^{PP} and the linear forms are the coordinate forms $l_k(x_1,x_2) = x_k$, where $1\le k\le 2$.
With this setting, $(U_n^{PP})_{n\in\mathbb{N}_0}$ is a Pólya process. The matrix of the replacement endomorphism of $(U_n^{PP})_{n\in\mathbb{N}_0}$ is denoted by A^{PP} and satisfies $A^{PP} = (R^{PP})^T$. We have

the following relation between A^{PP} and the transpose of the replacement matrix R, which is denoted by A as in section 11:

$$A^{PP} \coloneqq \frac{1}{S}A.$$

It follows that A^{PP} has the eigenvalues 1 and $\sigma_2 = \frac{m}{S}$. We find eigenvectors v_1^{PP} and v_2^{PP} of A^{PP}, which correspond to the eigenvectors v_1 and v_2 of A, by defining them in the following way:

$$v_1^{PP} \coloneqq \frac{1}{S}v_1,\ v_2^{PP} \coloneqq \frac{1}{S}v_2.$$

Similarly, we can define the corresponding linear forms in the dual basis of v_1^{PP} and v_2^{PP}, which are the eigenforms of A^{PP}:

$$u_1^{PP} \coloneqq Su_1,\ u_2^{PP} \coloneqq Su_2.$$

We finally state and prove the theorem about the asymptotic behaviour of a large discrete time Pólya urn process.

Theorem 13.38 ([19] Theorem 2.1)**.** *Let $(U_n)_{n\in\mathbb{N}_0}$ be the discrete time Pólya urn process with all notations as in section 11. We assume that the process is large, i.e. $\sigma_2 = \frac{m}{S} > \frac{1}{2}$. Then, there exists a unique random variable W^U such that, as $n \to \infty$,*

$$U_n = nv_1 + n^{\sigma_2}W^U v_2 + o(n^{\sigma_2}),$$

where the convergence is almost surely and in L^p for every $p \geq 1$. The moments of W^U can be inductively computed.

Proof. We use the Pólya process $(U_n^{PP})_{n\in\mathbb{N}_0}$ derived from the Pólya urn process as in Construction 13.37. The matrix A^{PP} of the Pólya process is diagonalizable since A is diagonalizable. All monogenic blocks of indices are of size 1, especially the principal block. Therefore, the Pólya process is principally semisimple and since we assume that the Pólya urn process is large, the corresponding Pólya process is large too. Therefore, we apply Theorem 13.34 to the process $(U_n^{PP})_{n\in\mathbb{N}_0}$.
There exists a random variable W^{PP} such that, as $n \to \infty$,

$$U_n^{PP} = nv_1^{PP} + n^{\sigma_2}W^{PP}v_2^{PP} + o(n^{\sigma_2}), \tag{13.8}$$

where the convergence happens almost surely and in L^p for every $p \geq 1$. The moments of W^{PP} can be inductively computed by Remark 13.36. From the proof of this theorem, we know that

$$W^{PP} = \lim_{n\to\infty} \frac{1}{n^{\sigma_2}} u_2^{PP}(U_n^{PP}).$$

Now, we use the developed relations between the Pólya process and the Pólya urn process in Equation 13.8, which leads to

$$\frac{1}{S}U_n = \frac{1}{S}nv_1 + n^{\sigma_2}\left(\lim_{n\to\infty}\frac{1}{n^{\sigma_2}}Su_2\left(\frac{1}{S}U_n\right)\right)\frac{1}{S}v_2 + o(n^{\sigma_2})$$

respectively

$$U_n = nv_1 + n^{\sigma_2}\left(\lim_{n\to\infty}\frac{1}{n^{\sigma_2}}u_2(U_n)\right)v_2 + o(n^{\sigma_2}),$$

where

$$\begin{aligned} W^{PP} &= \lim_{n\to\infty}\frac{1}{n^{\sigma_2}}u_2^{PP}(U_n^{PP}) = \lim_{n\to\infty}\frac{1}{n^{\sigma_2}}Su_2\left(\frac{1}{S}U_n\right) \\ &= \lim_{n\to\infty}\frac{1}{n^{\sigma_2}}u_2(U_n) =: W^U. \end{aligned}$$

This completes the proof. □

Remark 13.39 ([15])**.** We proved in section 11 that $\gamma_{\frac{u}{S},n}\left(\frac{A}{S}\right)^{-1}U_n$ is a martingale and thus its projections onto the eigenlines of A are also martingales. Hence,

$$M^U(n) \coloneqq \gamma_{\frac{u}{S},n}\left(\frac{m}{S}\right)^{-1}u_2(U_n)$$

is a martingale, and from the proofs of Theorem 13.34 and Theorem 13.38, we know that this martingale converges almost surely and in L^p for every $p \geq 1$. From these two theorems, we also have the limit random variable

$$W \coloneqq \lim_{n\to\infty}\frac{1}{n^{\sigma_2}}u_2(U_n).$$

13.6 The asymptotics of a large m-ary search tree process in discrete time

Remark 13.40. Let $(X_n)_{n\in\mathbb{N}_0}$ be an m-ary search tree process with notations as in section 12. We have already seen that we can assign a Pólya urn process $(U_n^{PP})_{n\in\mathbb{N}_0}$ to this process (see subsection 12.3) and that this Pólya urn process has balance $S = 1$. We have also seen that a Pólya urn process is a Pólya process when we divide the composition vector and its replacement matrix by S, but since we already have $S = 1$, the Pólya urn process associated to an m-ary search tree process is a Pólya process. We have seen that the processes satisfy $P \cdot X_n = U_n$ where we defined the matrix P in Remark 12.20. When A denotes the composition matrix of the m-ary search tree process and R is the replacement matrix of the corresponding Pólya urn process, we have $A = P^{-1}R^TP$. We define $A^{PP} = R^T$, which is the matrix of the replacement endomorphism of the corresponding Pólya process. We notice that A and A^{PP} have the same eigenvalues. Let $\lambda_k \in \mathrm{Sp}(A)$, $1 \leq k \leq m-1$, be the eigenvalues of A, enumerated such that $\lambda_1 = 1 > \Re(\lambda_2) \geq \Re(\lambda_3) \geq \cdots \geq \Re(\lambda_{m-1})$, and we set $\Re(\lambda_2) = \sigma_2$. We recall that $\overline{\lambda_2}$ is also an eigenvalue of A with equal real part as λ_2. The corresponding eigenvector and eigenform are $\overline{v_2}$ and $\overline{u_2}$ respectively. Let $(v_k)_{1\leq k\leq m-1}$ be a basis of eigenvectors of A, where v_k corresponds to the eigenvalue λ_k, and such that v_1 and v_2 are defined as in Definition 12.15. We denote by $(u_k)_{1\leq k\leq m-1}$ the corresponding dual basis. Since $Av_k = \lambda_k v_k$, we have $P^{-1}A^{PP}Pv_k = \lambda_k v_k$ and thus $A^{PP}Pv_k = \lambda_k Pv_k$, i.e. $v_k^{PP} \coloneqq Pv_k$ is an eigenvector of A^{PP}. Similarly, $u_kA^{PP} = \lambda_k u_k$ and hence $u_kP^{-1}A = \lambda_k u_kP^{-1}$, i.e. $u_k^{PP} \coloneqq u_kP^{-1}$ is an eigenform of A^{PP}.

We finally state and prove the theorem about the asymptotic behaviour of large discrete time m-ary search tree processes.

Theorem 13.41 ([18] Theorem 1)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be an m-ary search tree process with all notations as in section 12. We assume that the process is large, i.e. $\sigma_2 = \Re(\lambda_2) > \frac{1}{2}$. Then, there exist unique complex-valued random variables W_{λ_2} and $W_{\overline{\lambda_2}}$, such that, as $n \to \infty$,*

$$X_n = nv_1 + n^{\lambda_2} W_{\lambda_2} v_2 + n^{\overline{\lambda_2}} W_{\overline{\lambda_2}} \overline{v_2} + o(n^{\sigma_2}), \tag{13.9}$$

respectively

$$X_n = nv_1 + 2\Re(n^{\lambda_2} W_{\lambda_2} v_2) + o(n^{\sigma_2}), \tag{13.10}$$

where the convergence is almost surely and in L^p for every $p \geq 1$. The moments of W_{λ_2} and $W_{\overline{\lambda_2}}$ can be computed.

Proof. We use the Pólya process $(U_n^{PP})_{n\in\mathbb{N}_0}$ we have introduced before, associated to the m-ary search tree process (notice that the Pólya process and the Pólya urn process are the same). Since the matrix A is diagonalizable, the matrix A^{PP} is diagonalizable too. All monogenic blocks of indices are of size 1, especiall the principal blocks. Therefore, the Pólya process is principally semisimple and since we assume that the m-ary search tree process is large, the Pólya process is large too. Hence, we can apply Theorem 13.34 to the process $(U_n^{PP})_{n\in\mathbb{N}_0}$. There exist random variables $W_{\lambda_2}^{PP}$ and $W_{\overline{\lambda_2}}^{PP}$ such that, as $n \to \infty$,

$$U_n^{PP} = nv_1^{PP} + n^{\lambda_2} W_{\lambda_2}^{PP} v_2^{PP} + n^{\overline{\lambda_2}} W_{\overline{\lambda_2}}^{PP} \overline{v_2^{PP}} + o(n^{\sigma_2}),$$

where the convergence is almost surely and in L^p for every $p \geq 1$. The moments of W_{λ_2} and $W_{\overline{\lambda_2}}$ can be computed. Now, with the developed relations between the Pólya process and the m-ary search tree process, we get

$$\begin{aligned} PX_n = nPv_1 &+ n^{\lambda_2} \left(\lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2 P^{-1} P X_n \right) Pv_2 \\ &+ n^{\overline{\lambda_2}} \left(\lim_{n\to\infty} \frac{1}{n^{\overline{\lambda_2}}} \overline{u_2} P^{-1} P X_n \right) P\overline{v_2} + o(n^{\sigma_2}). \end{aligned}$$

Multiplying by P^{-1} gives us

$$\begin{aligned} X_n = nv_1 &+ n^{\lambda_2} \left(\lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2(X_n) \right) v_2 \\ &+ n^{\overline{\lambda_2}} \left(\lim_{n\to\infty} \frac{1}{n^{\overline{\lambda_2}}} \overline{u_2}(X_n) \right) \overline{v_2} + o(n^{\sigma_2}), \end{aligned}$$

where

$$W_{\lambda_2}^{PP} = \lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2 P^{-1} P X_n = \lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2(X_n) =: W_{\lambda_2},$$

and similarly $W_{\overline{\lambda_2}}^{PP} =: W_{\overline{\lambda_2}}$. This gives us Equation 13.9. For the second equation, we notice that $\overline{n^{\lambda_2}} = n^{\overline{\lambda_2}}$ and $\overline{u_2}(X_n) = \overline{u_2(X_n)}$ and remember that $2\Re(z) = z + \overline{z}$ for all $z \in \mathbb{C}$, hence

$$\begin{aligned} X_n &= nv_1 + n^{\lambda_2} W_{\lambda_2} v_2 + \overline{n^{\lambda_2} W_{\lambda_2} v_2} + o(n^{\sigma_2}) \\ &= nv_1 + 2\Re(n^{\lambda_2} W_{\lambda_2} v_2) + o(n^{\sigma_2}). \end{aligned}$$

This completes the proof. □

Remark 13.42 ([15])**.** We proved in section 12 that $(\gamma_{1,n}(A)^{-1} X_n)_{n \in \mathbb{N}_0}$ is a martingale and thus its projections onto the eigenlines of A are also martingales. In particular,

$$(\gamma_{1,n}(\lambda_2)^{-1} u_2(X_n))_{n \in \mathbb{N}_0}$$

is a martingale and from the proofs of Theorem 13.34 and Theorem 13.41 we know that this martingale converges almost surely and in L^p for every $p \geq 1$. From these two theorems, we also have the limit random variable

$$W \coloneqq \lim_{n \to \infty} \frac{1}{n^{\lambda_2}} u_2(X_n) = W_{\lambda_2}.$$

14 Pólya urns in continuous time

We define the two-dimensional Pólya urn process in continuous time and show its relation to the discrete time Pólya urn process.

14.1 Definition of the process

Definition 14.1 ([15])**.** We define a continuous time random process in the following way. We consider a continuous time setting with the time denoted by t and we have $t \in \mathbb{R}_{\geq 0}$. We start at time $t = 0$ with an urn which contains balls of two different colours, let us say α white and β blue balls, where $(\alpha, \beta)^T \in \mathbb{N}_0^2 \setminus \{0\}$. Every ball has an Exp(1)-distributed clock attached to it, meaning that the ringing time is exponentially distributed with parameter 1.
When the clock of a white ball rings, the ball is removed and replaced by $a + 1$ white and b blue balls. If the clock of a blue ball rings, the ball is removed and replaced by c white and $d + 1$ balls. All the new balls have again an Exp(1)-distributed clock attached to them. All the clocks are independent of each other. Furthermore, we want $a, b, c, d \in \mathbb{N}_0$. Sometimes this process is also defined with a, b, c, d negative. Then it could go extinct, but we do not consider this case. We call this process a continuous time Pólya urn process.
The composition vector is $\mathscr{U}(t) \in \mathbb{N}_0^2$, with

$$\mathscr{U}(t) = \begin{pmatrix} \mathscr{U}(t)^{(1)} \\ \mathscr{U}(t)^{(2)} \end{pmatrix} = \begin{pmatrix} \text{card}\{\text{white balls at time t}\} \\ \text{card}\{\text{blue balls at time t}\} \end{pmatrix},$$

and the replacement matrix is

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

We only consider a balanced urn, meaning the total number of balls added stays the same, i.e. $S \coloneqq a + b = c + d$ is a constant, called the balance. Let

$$\mathscr{U}(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

be the initial composition vector. Then, after n clocks ringed, the total number of balls in the urn is $\alpha + \beta + n \cdot S = u + n \cdot S$, where $\alpha + \beta = u$ is the total number of initial balls in the urn.
We only treat the case where $S > m \coloneqq a - c = d - b$.

Remark 14.2 ([19])**.** We see immediately that the replacement rules are the same as in discrete time. Moreover, this process is a continuous time Markov chain, since the state of the process only depends on the last state and the lifetime of the clocks has an exponential distribution. It is a continuous time branching process.

14.2 The embedding principle

Definition 14.3 ([19])**.** We consider the continuous time Pólya urn process with composition vectors $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$. The jump times of this process are denoted by $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_n < \ldots$, where the n-th jump time is the time of the n-th dislocation of the branching process. The process is constant on every interval $[\tau_n, \tau_{n+1})$.

Remark 14.4 ([19])**.** We notice that for a discrete time Pólya urn process $(U_n)_{n\in\mathbb{N}_0}$ and a continuous time Pólya urn process $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$, both with initial composition vector $(\alpha, \beta)^T$, we have

$$\mathscr{U}(t) = U_{a(t)},$$

where $a(t) \coloneqq \inf\{n \in \mathbb{N}_0 \mid \tau_n \geq t\}$.

Lemma 14.5 ([19] Lemma 3.1)**.** *The discrete time Pólya urn process $(U_n)_{n\in\mathbb{N}_0}$ and the continuous time Pólya urn process $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$ considered at the jump times $(\tau_n)_{n\in\mathbb{N}_0}$, both with initial composition vector $(\alpha, \beta)^T$, satisfy the following properties.*

(i) For $n \in \mathbb{N}_0$, $\tau_{n+1} - \tau_n$ is $\mathrm{Exp}(u + Sn)$-distributed, with u the total number of balls at time 0.

(ii) The processes $(\tau_n)_{n\in\mathbb{N}_0}$ and $(\mathscr{U}(\tau_n))_{n\in\mathbb{N}_0}$ are independent.

(iii) The processes $(U(n))_{n\in\mathbb{N}_0}$ and $(\mathscr{U}(\tau_n))_{n\in\mathbb{N}_0}$ have the same distribution.

Proof.

(i) The total number of balls at time $t \in [\tau_n, \tau_{n+1})$ is $u + Sn$. τ_{n+1} is the first jump time after τ_n, so $\tau_{n+1} - \tau_n$ is the minimum of $u + Sn$ independent $\mathrm{Exp}(1)$-distributed random variables, which means it is $\mathrm{Exp}(u + Sn)$-distributed (see Remark I.35).

(ii) The process $(\tau_n)_{n\in\mathbb{N}_0}$ describes when the continuous time Pólya urn process changes, but $(\mathscr{U}(\tau_n))_{n\in\mathbb{N}_0}$ describes how it changes, so they are independent.

(iii) The two considered processes have the same initial state and their evolution behaves the same as we just noticed in the remark before.

□

Remark 14.6 ([19])**.** From now on, we consider the discrete and continuous time Pólya urn processes in the same probability space. Then we have that

$$(U_n)_{n\in\mathbb{N}_0} = (\mathscr{U}(\tau_n))_{n\in\mathbb{N}_0} \ a.s. \tag{14.1}$$

We say that the discrete process is embedded in the continuous process and call it the embedding principle.

Remark 14.7 ([19])**.** Since the continuous time Pólya urn process has the same replacement matrix R as the discrete time version, it follows that the corresponding eigenvalues, eigenvectors and eigenforms of $A = R^T$ we introduced in Remark 11.4 are the same for the continuous time process.

15 A continuous time multitype branching process with $m-1$ types

We define the $(m-1)$-dimensional multitype branching process with $m-1$ types in continuous and discrete time and study their connections and the relation to the m-ary search tree process.

15.1 Definition of the process in continuous and discrete time

Definition 15.1 ([15])**.** We define a continuous time stochastic process in the following way. We consider a continuous time setting with the time denoted by t and we have $t \in \mathbb{R}_{\geq 0}$. The dimension of the process is $m-1$. We consider particles of different types, more precisely of $m-1$ types. Let

$$\mathscr{X}(t) = \begin{pmatrix} \mathscr{X}(t)^{(1)} \\ \vdots \\ \mathscr{X}(t)^{(m-1)} \end{pmatrix}$$

be the composition vector of this process at time t, where $\mathscr{X}(t)^{(j)}$ is the number of particles of type j alive at time t. Every particle of type j has an $\mathrm{Exp}(j)$-distributed clock attached to it. When the clock rings, the particle of type j dies and gives birth to new particles according to the following rule. If $1 \leq j < m-1$, then it gives birth to one particle of type $j+1$. If $j = m-1$, the particle gives birth to m particles of type 1. The clocks are independent of each other. The initial composition vector is $\mathscr{X}(0) \in \mathbb{N}_0^{m-1} \setminus \{0\}$.
This is a continuous time multitype branching process with $m-1$ types. We just consider this one version of a process, so we refer to it as the continuous time multitype branching process with $m-1$ types.

Remark 15.2 ([17])**.** We can see that the replacement rules of this process are the same as in the m-ary search tree process. The process is a continuous time Markov chain, since the state of the process only depends on the last state and the lifetime of the particles is exponentially distributed.

Definition 15.3 ([17])**.** We consider the continuous time multitype branching process with $m-1$ types as defined before. The jump times of this process are denoted by $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \ldots$, where the n-th jump time is the time of the n-th dislocation of the branching process. The process is constant on the intervals $[\tau_n, \tau_{n+1})$.

Definition 15.4 ([17])**.** We define a discrete time multitype branching process with $m-1$ types. We consider particles of $m-1$ different types and denote by

$$X_n = \begin{pmatrix} X_n^{(1)} \\ \vdots \\ X_n^{(m-1)} \end{pmatrix}$$

the composition vector of this process for $n \in \mathbb{N}_0$. $X_n^{(j)}$ is the number of particles of type j for $n \in \mathbb{N}_0$. The initial composition vector is $X_0 \in N_0^{m-1} \setminus \{0\}$. The composition vector changes according to the following rules. We have for all $n \in \mathbb{N}_0$ and for all $j = 1, \ldots, m-1$,

$$\mathbb{P}(X_{n+1} + \Delta_j \mid X_n) = \frac{jX_n^{(j)}}{n+1},$$

where

$$\begin{aligned}
\Delta_1 &= \big(-1, 1, 0, \ldots\big)^T \\
\Delta_2 &= \big(0, -1, 1, 0, \ldots\big)^T \\
&\vdots \\
\Delta_{m-2} &= \big(0, \ldots, 0, -1, 1\big)^T \\
\Delta_{m-1} &= \big(m, 0, \ldots, 0, -1\big)^T .
\end{aligned}$$

Remark 15.5 ([17])**.** This discrete process is a slight generalization of the m-ary search tree process, when we set the initial composition vector to

$$X_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

we obtain an m-ary search tree process. This process is by definition a discrete time Markov chain and it has the same evolution as the continuous time multitype branching process with $m-1$ types we defined before.

Remark 15.6 ([17])**.** Since the stochastic behaviour of the discrete time multitype branching process with $m-1$ types is equal to the evolution of an m-ary search tree process (only the initial composition vector is different), Theorem 13.41 and Remark 13.42 also hold for the large discrete time multitype branching process with $m-1$ types.

15.2 The embedding principle

Lemma 15.7 ([17] Lemma 3.1)**.** *The discrete time multitype branching process with $m-1$ types $(X_n)_{n\in\mathbb{N}_0}$ and the continuous time multitype branching process with $m-1$ types $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ considered at the jump times $(\tau_n)_{n\in\mathbb{N}_0}$, both with initial composition vector $X_0 = \mathscr{X}(0)$, satisfy the following properties.*

(i) For $n \in \mathbb{N}_0$, $\tau_{n+1} - \tau_n$ is $\mathrm{Exp}(n - 1 + N_0)$-distributed, with $N_0 = \sum_{j=1}^{m-1} j\mathscr{X}(0)^{(j)}$ being the number of free places in $\mathscr{X}(0)$.

(ii) The processes $(\tau_n)_{n\in\mathbb{N}_0}$ and $(\mathscr{X}(\tau_n))_{n\in\mathbb{N}_0}$ are independent.

(iii) The processes $(X_n)_{n\in\mathbb{N}_0}$ and $(\mathscr{X}(\tau_n))_{n\in\mathbb{N}_0}$ have the same distribution.

Proof.

(i) We consider first $\tau_1 - \tau_0 = \tau_1$. When a particle of type j is $\text{Exp}(j)$-distributed, the first time one of k such particles dies is the minimum of k independent $\text{Exp}(j)$-distributed random variables, so it is $\text{Exp}(k \cdot j)$-distributed and thus has the same distribution as if the first one of $k \cdot j$ particles of type 1 dies (see Remark I.35). If we consider a particle of type j and a particle of type k, the probability of the first one of them to die is $\text{Exp}(j+k)$-distributed, which is the same distribution as the one of the minimum of $j+k$ independent $\text{Exp}(1)$-distributed random variables. Therefore, at time τ_1, we can see it as the time where the first one of $N_0 = \sum_{j=1}^{m-1} j\mathscr{X}(0)^{(j)}$ independent $\text{Exp}(1)$-distributed particles dies. So it is $\text{Exp}(N_0)$-distributed. Now we consider $\tau_2 - \tau_1$. If a particle of type j, $j < m-1$, dies, a particle of type $j+1$ is born, meaning there is one particle of type j less and one particle of type $j+1$ more. Therefore the probability for the next particle to die has the same distribution as the minimum of $N_0 - j + (j+1) = N_0 + 1$ independent $\text{Exp}(1)$-distributed random variables. In the case where $j = m-1$, it would be the minimum of $N_0 - j + m = N_0 + 1$ independent $\text{Exp}(1)$-distributed random variables. Repeating this process inductively shows that $\tau_{n+1} - \tau_n$ is indeed $\text{Exp}(N_0 + n - 1)$-distributed.

(ii) The process $(\tau_n)_{n\in\mathbb{N}_0}$ describes when the continuous time branching process changes, but $(\mathscr{X}(\tau_n))_{n\in\mathbb{N}_0}$ describes how it changes, so they are independent.

(iii) The two considered processes have the same initial state and their evolution behaves the same as we just mentioned in the remark before.

□

Remark 15.8 ([17])**.** From now on, we consider the discrete and continuous time multitype branching processes with $m-1$ types in the same probability space. Then we have that

$$(X_n)_{n\in\mathbb{N}_0} = (\mathscr{X}(\tau_n))_{n\in\mathbb{N}_0} \ a.s. \tag{15.1}$$

We say that the discrete process is embedded in the continuous process and call it the embedding principle.

Definition 15.9 ([17])**.** The composition matrix of the process $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ is

$$A \coloneqq \begin{pmatrix} -1 & & & & & m(m-1) \\ 1 & -2 & & & & \\ & 2 & -3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & -(m-2) & \\ & & & & m-2 & -(m-1) \end{pmatrix}.$$

16 Martingales in the Pólya urn process and the multitype branching process with $m-1$ types in continuous time

We find a martingale in the continuous time Pólya urn process (see section 14) as well as in the continuous time multitype branching process with $m-1$ types (see section 15).

The method used in the proofs of the following results are inspired by the proof of Theorem 8.1.1 in [46] and by [5].

Lemma 16.1 ([31] Lemma 9.2)**.** *Let $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$ be the continuous time Pólya urn process with replacement matrix R. Let $A = R^T$ be its transpose. Then, we have*

$$\frac{d}{dt}\mathbb{E}(\mathscr{U}(t)) = A \cdot \mathbb{E}(\mathscr{U}(t)).$$

Proof. Let $\varepsilon > 0$ and $t \in \mathbb{R}_{\geq 0}$. We want to study the behaviour of $\mathbb{E}(\mathscr{U}(t+\varepsilon))$ when $\varepsilon \to 0$.
First of all, we notice that the following formula holds

$$\mathscr{U}(t+\varepsilon) = \sum_{i=1}^{2} \sum_{j=1}^{\mathscr{U}(t)^{(i)}} \mathscr{U}(\varepsilon)^{ij},$$

where $\mathscr{U}(\varepsilon)^{ij}$ is the vector denoting the line of descent from the j-th ball of colour white ($i=1$), resp. blue ($i=2$), living at time t. The random variables $\mathscr{U}(\varepsilon)^{ij}$ are independent of $\mathscr{U}(t)$. Therefore we get the following equation

$$\begin{aligned}\mathbb{E}(\mathscr{U}(t+\varepsilon)) &= \mathbb{E}\left(\sum_{i=1}^{2} \sum_{j=1}^{\mathscr{U}(t)^{(i)}} \mathscr{U}(\varepsilon)^{ij}\right) \\ &= \mathbb{E}\left(\mathscr{U}(t)^{(1)}\right) \cdot \mathbb{E}\left(\mathscr{U}(\varepsilon)^{1j}\right) + \mathbb{E}\left(\mathscr{U}(t)^{(2)}\right) \cdot \mathbb{E}\left(\mathscr{U}(\varepsilon)^{2j}\right).\end{aligned}$$

Now we notice that the expectation of $\mathscr{U}(\varepsilon)^{1j}$ is the expectation of a continuous time Pólya urn process at time ε that started with a single white ball, i.e. it started with an initial composition vector $(1,0)^T$. Therefore we get

$$\mathbb{E}(\mathscr{U}(\varepsilon)^{1j}) = \mathbb{E}\left(\mathscr{U}(\varepsilon) \mid \mathscr{U}(0) = \begin{pmatrix}1\\0\end{pmatrix}\right)$$

and since ε is taken to be small, we have further

$$\mathbb{E}\left(\mathscr{U}(\varepsilon) \mid \mathscr{U}(0) = \begin{pmatrix}1\\0\end{pmatrix}\right) = (1-\varepsilon)\cdot\begin{pmatrix}1\\0\end{pmatrix} + \varepsilon \cdot \begin{pmatrix}a+1\\b\end{pmatrix} + O(\varepsilon^2)\cdot I$$

and thus

$$\mathbb{E}\left(\mathscr{U}(\varepsilon) \mid \mathscr{U}(0) = \begin{pmatrix}0\\1\end{pmatrix}\right) = (1-\varepsilon)\cdot\begin{pmatrix}0\\1\end{pmatrix} + \varepsilon \cdot \begin{pmatrix}c\\d+1\end{pmatrix} + O(\varepsilon^2)\cdot I.$$

This gives us

$$\begin{aligned}\mathbb{E}(\mathscr{U}(t+\varepsilon)) &= \mathbb{E}\left(\mathscr{U}(t)^{(1)}\right)\cdot\mathbb{E}\left(\mathscr{U}(\varepsilon)^{1j}\right)+\mathbb{E}\left(\mathscr{U}(t)^{(2)}\right)\cdot\mathbb{E}\left(\mathscr{U}(\varepsilon)^{2j}\right)\\ &= \mathbb{E}\left(\mathscr{U}(t)^{(1)}\right)\cdot\begin{pmatrix}1-\varepsilon+\varepsilon\cdot(a+1)\\ \varepsilon\cdot b\end{pmatrix}\\ &+\mathbb{E}\left(\mathscr{U}(t)^{(2)}\right)\cdot\begin{pmatrix}\varepsilon\cdot c\\ 1-\varepsilon+\varepsilon\cdot(d+1)\end{pmatrix}+O(\varepsilon^2)\cdot I\\ &= \mathbb{E}(\mathscr{U}(t))+\begin{pmatrix}\varepsilon\cdot a\cdot\mathbb{E}\left(\mathscr{U}(t)^{(1)}\right)+\varepsilon\cdot c\cdot\mathbb{E}\left(\mathscr{U}(t)^{(2)}\right)\\ \varepsilon\cdot b\cdot\mathbb{E}\left(\mathscr{U}(t)^{(1)}\right)+\varepsilon\cdot d\cdot\mathbb{E}\left(\mathscr{U}(t)^{(2)}\right)\end{pmatrix}+O(\varepsilon^2)\cdot I\\ &= \mathbb{E}(\mathscr{U}(t))+\varepsilon\cdot A\cdot\mathbb{E}(\mathscr{U}(t))+O(\varepsilon^2)\cdot I.\end{aligned}$$

We rearrange the equation and obtain

$$\frac{\mathbb{E}(\mathscr{U}(t+\varepsilon))-\mathbb{E}(\mathscr{U}(t))}{\varepsilon}=A\cdot\mathbb{E}(\mathscr{U}(t))+O(\varepsilon)\cdot I.$$

Now we take the limit $\varepsilon\to 0$, which finally leads to

$$\frac{d}{dt}\mathbb{E}(\mathscr{U}(t))=A\cdot\mathbb{E}(\mathscr{U}(t)).$$

□

Lemma 16.2 ([31] Lemma 9.2)**.** *Let $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ be the continuous time multitype branching process with $m-1$ types. Let A be its composition matrix. Then we have*

$$\frac{d}{dt}\mathbb{E}(\mathscr{X}(t))=A\cdot\mathbb{E}(\mathscr{X}(t)).$$

Proof. We prove this statement similar to the case of the continuous time Pólya urn process. Let $\varepsilon>0$ and $t\in\mathbb{R}_{\geq 0}$. We want to study the behaviour of $\mathbb{E}(\mathscr{X}(t+\varepsilon))$ when $\varepsilon\to 0$.
First of all, we notice that the following formula holds

$$\mathscr{X}(t+\varepsilon)=\sum_{i=1}^{m-1}\sum_{j=1}^{\mathscr{X}(t)^{(i)}}\mathscr{X}(\varepsilon)^{ij},$$

where $\mathscr{X}(\varepsilon)^{ij}$ is the vector denoting the line of descent from the j-th particle type i living at time t. The random variables $\mathscr{X}(\varepsilon)^{ij}$ are independent of $\mathscr{X}(t)$. Therefore, we get the following equation:

$$\mathbb{E}(\mathscr{X}(t+\varepsilon))=\mathbb{E}\left(\sum_{i=1}^{m-1}\sum_{j=1}^{\mathscr{X}(t)^{(i)}}\mathscr{X}(\varepsilon)^{ij}\right)=\sum_{i=1}^{m-1}\mathbb{E}\left(\mathscr{X}(t)^{(i)}\right)\cdot\mathbb{E}\left(\mathscr{X}(\varepsilon)^{ij}\right).$$

The expectation of $\mathscr{X}(\varepsilon)^{ij}$ is the expectation of a continuous time multitype branching process with $m-1$ types at time ε that started with only one particle of type i and no other particles. Let $(e_i)_{1\leq i\leq m-1}$ be the canonical basis of $\mathbb{R}^{m-1}$ (see Notation 2.8). Then

$$\mathbb{E}(\mathscr{X}(\varepsilon)^{ij}) = \mathbb{E}\left(\mathscr{X}(\varepsilon) \mid \mathscr{X}(0) = e_i\right)$$

and since ε is taken to be small, we have further

$$\begin{aligned}
\mathbb{E}\left(\mathscr{X}(\varepsilon) \mid \mathscr{X}(0) = e_1\right) &= (1-\varepsilon\cdot 1)\cdot e_1 + \varepsilon\cdot 1\cdot e_2 + O(\varepsilon^2)\cdot I;\\
\mathbb{E}\left(\mathscr{X}(\varepsilon) \mid \mathscr{X}(0) = e_2\right) &= (1-\varepsilon\cdot 2)\cdot e_2 + \varepsilon\cdot 2\cdot e_3 + O(\varepsilon^2)\cdot I;\\
&\vdots\\
\mathbb{E}\left(\mathscr{X}(\varepsilon) \mid \mathscr{X}(0) = e_i\right) &= (1-\varepsilon\cdot i)\cdot e_i + \varepsilon\cdot i\cdot e_{i+1} + O(\varepsilon^2)\cdot I;\\
&\vdots\\
\mathbb{E}\left(\mathscr{X}(\varepsilon) \mid \mathscr{X}(0) = e_{m-1}\right) &= (1-\varepsilon\cdot(m-1))\cdot e_{m-1} + \varepsilon\cdot(m-1)\cdot m\cdot e_1 + O(\varepsilon^2)\cdot I.
\end{aligned}$$

This gives us

$$\begin{aligned}
\mathbb{E}(\mathscr{X}(t+\varepsilon)) &= \sum_{i=1}^{m-1}\mathbb{E}\left(\mathscr{X}(t)^{(i)}\right)\cdot\mathbb{E}\left(\mathscr{X}(\varepsilon)^{ij}\right)\\
&= \mathbb{E}(\mathscr{X}(t)) + \varepsilon\cdot\begin{pmatrix}
(m-1)m\mathbb{E}\left(\mathscr{X}(t)^{(m-1)}\right) - \mathbb{E}\left(\mathscr{X}(t)^{(1)}\right)\\
\mathbb{E}\left(\mathscr{X}(t)^{(1)}\right) - 2\mathbb{E}\left(\mathscr{X}(t)^{(2)}\right)\\
2\mathbb{E}\left(\mathscr{X}(t)^{(2)}\right) - 3\mathbb{E}\left(\mathscr{X}(t)^{(3)}\right)\\
\vdots\\
(i-1)\mathbb{E}\left(\mathscr{X}(t)^{(i-1)}\right) - i\mathbb{E}\left(\mathscr{X}(t)^{(i)}\right)\\
\vdots\\
(m-2)\mathbb{E}\left(\mathscr{X}(t)^{(m-2)}\right) - (m-1)\mathbb{E}\left(\mathscr{X}(t)^{(m-1)}\right)
\end{pmatrix}\\
&\quad + O(\varepsilon^2)\cdot I\\
&= \mathbb{E}(\mathscr{X}(t)) + \varepsilon\cdot A\cdot\mathbb{E}(\mathscr{X}(t)) + O(\varepsilon^2)\cdot I.
\end{aligned}$$

We rearrange the equation and obtain

$$\frac{\mathbb{E}(\mathscr{X}(t+\varepsilon)) - \mathbb{E}(\mathscr{X}(t))}{\varepsilon} = A\cdot\mathbb{E}(\mathscr{X}(t)) + O(\varepsilon)\cdot I.$$

Now we take the limit $\varepsilon\to 0$, which finally leads to

$$\frac{d}{dt}\mathbb{E}(\mathscr{X}(t)) = A\cdot\mathbb{E}(\mathscr{X}(t)).$$

□

Theorem 16.3 ([31] Lemma 9.2)**.** *Let $m \in \mathbb{N}$, $m \geq 3$. We denote by $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ the continuous time Pólya urn process or the continuous time multitype branching process with $m-1$ types. Let A be the transpose of the corresponding replacement matrix or let A be the corresponding composition matrix. Let $(\mathcal{F}_t)_{t\in\mathbb{R}_{\geq 0}}$ be the corresponding filtration of the process, defined by $\mathcal{F}_t = \sigma(\mathscr{X}(s) \mid s \leq t,\ s \in \mathbb{R}_{\geq 0})$.*
Then

$$e^{-tA} \cdot \mathscr{X}(t),\ t \in \mathbb{R}_{\geq 0},$$

is a vectorial $\mathcal{F}_t$-martingale.

Proof. We know that both processes satisfy

$$\frac{d}{dt}\mathbb{E}(\mathscr{X}(t)) = A \cdot \mathbb{E}(\mathscr{X}(t)).$$

We solve this differential equation:

$$\begin{aligned}
&\frac{d}{dt}\mathbb{E}(\mathscr{X}(t)) = A \cdot \mathbb{E}(\mathscr{X}(t)) \\
\iff &\frac{1}{\mathbb{E}(\mathscr{X}(t))}\, d(\mathbb{E}(\mathscr{X}(t))) = A\, dt \\
\iff &\log(\mathbb{E}(\mathscr{X}(t))) = tA + c \mid c \in \mathbb{R} \\
\iff &\mathbb{E}(\mathscr{X}(t)) = e^{tA+c}.
\end{aligned}$$

Now taking $\mathbb{E}(\mathscr{X}(0)) = e^c$, we obtain

$$\mathbb{E}(\mathscr{X}(t)) = e^{tA} \cdot \mathbb{E}(\mathscr{X}(0)).$$

Let $d \in \mathbb{N}$ be the dimension of the process. Let $t, s \in \mathbb{R}_{\geq 0}$, then we have by the Markov property

$$\begin{aligned}
\mathbb{E}(\mathscr{X}(t+s) \mid \mathcal{F}_s) &= \mathbb{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{\mathscr{X}(s)^{(i)}} \mathscr{X}(t)^{ij} \mid \mathcal{F}_s\right) \\
&= \sum_{i=1}^{d} \mathbb{E}\left(\mathscr{X}(s)^{(i)} \mid \mathscr{X}(s)^{(i)}\right) \cdot \mathbb{E}\left(\mathscr{X}(t)^{ij} \mid \mathcal{F}_s\right) \\
&= \sum_{i=1}^{d} \mathscr{X}(s)^{(i)} \cdot \mathbb{E}\left(\mathscr{X}(t) \mid \mathscr{X}(0) = e_i\right) \\
&= \sum_{i=1}^{d} \mathscr{X}(s)^{(i)} \cdot e^{tA} \cdot \mathbb{E}\left(\mathscr{X}(0) \mid \mathscr{X}(0) = e_i\right) \\
&= e^{tA} \cdot \sum_{i=1}^{d} \mathscr{X}(s)^{(i)} \cdot e_i \\
&= e^{tA} \cdot \mathscr{X}(s).
\end{aligned}$$

This implies finally that for $t, s \in \mathbb{R}_{\geq 0}$, $t \geq s \geq 0$,

$$\mathbb{E}(e^{-tA} \cdot \mathscr{X}(t) \mid \mathcal{F}_s) = e^{-sA} \cdot \mathscr{X}(s).$$

The $\mathcal{F}_t$-adaptation is clear. We notice that $\mathscr{X}(t)$ has only non-negative entries. It was proven in [31] using the definition of the matrix exponential, projections into the generalized eigenspaces of A, quotient spaces and inequalities about eigenvalues of matrices that $\|e^{tA}\| \leq Ce^{\lambda_1 t}$ (the choice of norm $\|\cdot\|$ is irrelevant), with C a unspecified constant depending on the process and λ_1 being the largest real eigenvector of A. This implies $\mathbb{E}(\mathscr{X}(t)) = O(e^{t\lambda_1})$ and we conclude that every component of $e^{-tA} \cdot \mathscr{X}(t)$ lies in L^1. This completes the proof. □

Remark 16.4 ([31])**.** Let $\lambda \in \mathbb{C}$ be an eigenvalue of A and denote by λ_1 the largest real eigenvector of A. Janson proved in [31] (Lemma 10.2) that if $\Re(\lambda) > \frac{1}{2}\lambda_1$, then the projection of $(e^{-tA}\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ to the corresponding eigenspace of the eigenvalue λ is bounded in L^p for every $p \geq 1$. This was done using that $(\mathbb{E}(|\mathscr{X}(t)|^p))^{\frac{1}{p}} = O(e^t)$ for $p = 2$, implying the statement holds for $p \leq 2$ and then using induction on p.

17 Asymptotic behaviour of the continuous time Pólya urn process

We prove a result about the asymptotic behaviour of a large continuous time Pólya urn process and derive a relation to the discrete time Pólya urn process. Moreover, we show a result involving dislocation equations.

Notation 17.1. We use the same notations as in sections 11, 13 and 14. We write $(U_n)_{n\in\mathbb{N}_0}$ resp. $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$ for the discrete resp. continuous time Pólya urn process. The corresponding replacement matrix is

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

and we set $A = R^T$. We denote by v_1 and v_2 the eigenvectors of A with corresponding eigenforms u_1 and u_2, defined in Remark 11.4. Its eigenvalues are $S = a + b$ and $m = a - c$. The process is called large if $\sigma_2 = \frac{m}{S} > \frac{1}{2}$. The initial composition vector of the process is $\mathscr{U}(0) = (\alpha, \beta)^T$ and we set $u = \alpha + \beta$.
Furthermore, we recall the limit random variable

$$W = \lim_{n\to\infty} \frac{1}{n^{\sigma_2}} u_2(U_n),$$

defined in Remark 13.39.

17.1 Asymptotics of a large Pólya urn process in continuous time

Remark 17.2 ([19])**.** We recall the transition operator we used in section 13.5 for Pólya processes, which in the case of Pólya urn processes is defined on the space of all functions $f\colon \mathbb{R}^2 \to W$, where W is any real vector space. We have

$$\Phi(f)(x,y) = x(f(x+a, y+b) - f(x,y)) + y(f(x+c, y+d) - f(x,y)),\ x, y \in \mathbb{R}^2.$$

Proposition 13.35 states several results about the reduced polynomials. The following lemma restates some of these results, as well as some others, for the case of a 2-dimensional process. Further details can be found in [44].

Lemma 17.3 ([19] Lemma 3.4)**.** *There exists a unique family of polynomials (called reduced polynomials) $Q_{p,q} \in \mathbb{R}[x, y]$, where $p, q \in \mathbb{N}_0$, such that:*

(i) $Q_{0,0} = 1$, $Q_{1,0} = u_1$ *and* $Q_{0,1} = u_2$*;*

(ii) $\Phi(Q_{p,q}) = (pS + qm)Q_{p,q}$ *for all* $p, q \in \mathbb{N}_0$*;*

(iii) $u_1^p u_2^p - Q_{p,q} \in \mathrm{Span}\{Q_{p',q'} \mid p'S + q' < pS + qm\}$ *for all* $p, q \in \mathbb{N}_0$*;*

(iv) for a polynomial $P \in \mathbb{R}[x,y]$ and $t \in \mathbb{R}_{\geq 0}$, we have

$$\mathbb{E}(P(\mathscr{U}(t))) = e^{t\Phi}(P)(\mathscr{U}(0)),$$

where Φ is restricted to any finite-dimensional polynomial space containing P.

We finally state the theorem about the asymptotic behaviour of a large continuous time Pólya urn process.

Theorem 17.4 ([19] Theorem 3.2)**.** *Let $(\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$ be the large Pólya urn process with notations as described before. Then, as $t \to \infty$, we have*

$$\mathscr{U}(t) = e^{St}\xi v_1(1+o(1)) + e^{mt}\mathscr{W} v_2(1+o(1)), \tag{17.1}$$

where ξ and $\mathscr{W}$ are real-valued random variables, the small o being almost surely and in L^p for $p \geq 1$. Moreover, ξ is Gamma $\left(\frac{u}{S}, 1\right)$*-distributed, with $u = \alpha + \beta$ being the total number of balls at time $t = 0$ (see Definition 2.20).*

Proof. We have seen in section 16, that the process

$$(e^{-tA}\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$$

is a vectorial martingale and we know by Remark 16.4, that the projections of $(e^{-tA}\mathscr{U}(t))_{t\in\mathbb{R}_{\geq 0}}$ to the corresponding eigenspace of the eigenvalue λ of A is bounded in L^p for every $p \geq 1$ if $\Re(\lambda) > \frac{1}{2}\lambda_1$, where λ_1 is the largest real eigenvalue of A. In our case, this means that $u_1(e^{-tA}\mathscr{U}(t))$ and $u_2(e^{-tA}\mathscr{U}(t))$ converge in L^p for $p \geq 1$, since the process is large and $S,\ m > \frac{1}{2}S$. However, we want to use a different technique here, which leads easily to the rest of the statement. We just use the fact that the vectorial martingale is bounded in L^2 and therefore converges in L^2. Consequently, the projections on the eigenlines $\mathbb{R}v_1$ and $\mathbb{R}v_2$, given by $e^{-t}u_1(\mathscr{U}(t))$ and $e^{-t\lambda_2}u_2(\mathscr{U}(t))$ respectively, are also L^2-convergent real valued martingales, thus converge almost surely. We define the random variables ξ and $\mathscr{W}$ by

$$\xi = \lim_{t\to\infty} u_1(e^{-tA}\mathscr{U}(t)) = \lim_{t\to\infty} e^{-tS}u_1(\mathscr{U}(t)), \tag{17.2}$$

$$\mathscr{W} = \lim_{t\to\infty} u_2(e^{-tA}\mathscr{U}(t)) = \lim_{t\to\infty} e^{-tm}u_2(\mathscr{U}(t)). \tag{17.3}$$

Now, part (iv) from Lemma 17.3 implies together with part (ii) that for all $p, q \in \mathbb{N}_0$

$$\mathbb{E}(Q_{p,q}(\mathscr{U}(t))) = e^{t(pS+qm)}Q_{p,q}(\mathscr{U}(0)).$$

Part (iii) of Lemma 17.3 leads to

$$\mathbb{E}(u_1^p u_2^q(\mathscr{U}(t))) = e^{t(pS+qm)}Q_{p,q}(\mathscr{U}(0)) + o(e^{t(pS+qm)})$$

for all $p, q \in \mathbb{N}_0$. Hence, we have

$$\mathbb{E}(e^{-tpS}u_1^p(\mathscr{U}(t))) = Q_{p,0}(\mathscr{U}(0)) + o(1)$$

and

$$\mathbb{E}(e^{-tqm}u_2^q(\mathscr{U}(t))) = Q_{0,q}(\mathscr{U}(0)) + o(1)$$

for all $p, q \in \mathbb{N}_0$ Therefore, the martingales $e^{-tS}u_1(\mathscr{U}(t))$ and $e^{-tm}u_2(\mathscr{U}(t))$ are bounded in L^p for any $p \geq 1$ and their limits ξ and $\mathscr{W}$ satisfy

$$\mathbb{E}(\xi^p) = Q_{p,0}(\mathscr{U}(0)) \text{ and } \mathbb{E}(\mathscr{W}^p) = Q_{0,p}(\mathscr{U}(0)) \tag{17.4}$$

for all $p \in \mathbb{N}_0$. Due to the spectral decomposition of A, we have

$$\mathscr{U}(t) = u_1(\mathscr{U}(t)) \cdot v_1 + u_2(\mathscr{U}(t)) \cdot v_2.$$

Thus, we get

$$\mathscr{U}(t) = e^{tS}e^{-tS}u_1(\mathscr{U}(t)) \cdot v_1 + e^{tm}e^{-tm}u_2(\mathscr{U}(t)) \cdot v_2,$$

which is equal to

$$\mathscr{U}(t) = e^{St}\xi v_1(1 + o(1)) + e^{mt}\mathscr{W}v_2(1 + o(1)).$$

We know from Remark 13.36 that $Q_{p,0}$ is of the form

$$Q_{p,0} = u_1(u_1 + 1)\dots(u_1 + p - 1)$$

for $p \in \mathbb{N}_0$, which, together with Equation 17.4, shows that the p-th moment of ξ for any $p \in \mathbb{N}_0$ is

$$\mathbb{E}(\xi^p) = \frac{u}{S}\left(\frac{u}{S} + 1\right)\dots\left(\frac{u}{S} + p - 1\right) = \frac{\Gamma(\frac{u}{S} + p)}{\Gamma(\frac{u}{S})}.$$

We remember that u is the total number of balls at time $t = 0$ and u_1 was defined as

$$u_1\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{S}(x + y)$$

in Remark 11.4. It is a well-known fact that the gamma distribution is characterized by its moments, which proves that ξ is Gamma $\left(\frac{u}{S}, 1\right)$-distributed (see Definition 2.20). □

17.2 A martingale connection

Theorem 17.5 (Martingale connection, [19] Theorem 3.10)**.** *The following equation holds almost surely*

$$\mathscr{W} = \xi^{\sigma_2}W,$$

with ξ and W being independent.

Proof. We apply the first projection to the embedding principle (Equation 14.1):

$$u_1(\mathscr{U}(\tau_n)) = u_1(U_n) \text{ a.s.}$$

By definition, u_1 gives us the number of balls in the urn at time n divided by S, which also equals $\frac{1}{S}(u + Sn) = n(1 + o(1))$. The jump times τ_n tend to infinity as $n \to \infty$. Then,

$$\lim_{n\to\infty} ne^{-S\tau_n} = \lim_{n\to\infty} e^{-S\tau_n}u_1(U_n) = \lim_{n\to\infty} e^{-S\tau_n}u_1(\mathscr{U}(\tau_n)) = \xi.$$

Furthermore,

$$\xi^{\sigma_2} = \lim_{n\to\infty} (ne^{-S\tau_n})^{\frac{m}{S}} = \lim_{n\to\infty} (n^{\sigma_2} e^{-m\tau_n}).$$

We apply the second projection to the embedding principle (Equation 14.1):

$$u_2(\mathscr{U}(\tau_n)) = u_2(U_n) \ a.s.$$

This leads to

$$\mathscr{W} = \lim_{n\to\infty} e^{-m\tau_n} u_2(\mathscr{U}(\tau_n)) = \lim_{n\to\infty} e^{-m\tau_n} n^{\sigma_2} \frac{1}{n^{\sigma_2}} u_2(U_n) = \xi^{\sigma_2} W \ a.s.,$$

which is a martingale connection in finite time due to the definition of W (see Remark 13.39). Part (ii) of Lemma 14.5 implies that ξ and W are independent. □

17.3 Dislocation equations for continuous time Pólya urn processes

Notation 17.6 ([19])**.** We write $(\mathscr{U}_{(\alpha,\beta)}(t))_{t\in\mathbb{R}_{\geq 0}}$ for the continuous time Pólya urn process started with composition vector $(\alpha,\beta)^T \in \mathbb{N}_0^2 \setminus \{0\}$.

Remark 17.7 ([19])**.** Since the clocks attached to the balls in the continuous time Pólya urn process are all independent of each other, a continuous time Pólya urn process $(\mathscr{U}_{(\alpha,\beta)}(t))_{t\in\mathbb{R}_{\geq 0}}$ is equal to the sum of α copies of $(\mathscr{U}_{(1,0)}(t))_{t\in\mathbb{R}_{\geq 0}}$ (started from one white ball) and β copies of $(\mathscr{U}_{(0,1)}(t))_{t\in\mathbb{R}_{\geq 0}}$ (started from one blue ball). We want to study these two processes.
Let $\tau \in \mathbb{R}_{\geq 0}$ be the first splitting time of the processes, then we get the following finite time dislocation equations using the Markov property:

$$\forall t > \tau, \begin{cases} \mathscr{U}_{(1,0)}(t) \overset{\mathcal{L}}{=} [a+1]\mathscr{U}_{(1,0)}(t-\tau) + [b]\mathscr{U}_{(0,1)}(t-\tau) \ a.s., \\ \mathscr{U}_{(0,1)}(t) \overset{\mathcal{L}}{=} [c]\mathscr{U}_{(1,0)}(t-\tau) + [d+1]\mathscr{U}_{(0,1)}(t-\tau) \ a.s., \end{cases} \tag{17.5}$$

where $[n]X$ denotes the sum of n copies of the random variable X.

Remark 17.8 ([19])**.** Thanks to Equations 17.4, the martingales $(e^{-tm}u_2(\mathscr{U}_{(1,0)}(t)))_{t\in\mathbb{R}_{\geq 0}}$ and $(e^{-tm}u_2(\mathscr{U}_{(0,1)}(t)))_{t\in\mathbb{R}_{\geq 0}}$ have expectations

$$u_2(\mathscr{U}_{(1,0)}(0)) = \frac{b}{S} \text{ and } u_2(\mathscr{U}_{(0,1)}(0)) = -\frac{c}{S}$$

respectively (see Remark 11.4). We remember that these martingales converge in L^p for every $p \geq 1$. We define

$$X = \lim_{t\to\infty} e^{-tm} u_2(\mathscr{U}_{(1,0)}(t)) \text{ and } Y = \lim_{t\to\infty} e^{-tm} u_2(\mathscr{U}_{(0,1)}(t)).$$

Proposition 17.9 ([19] Proposition 4.2)**.** *The limit random variables X and Y are solutions of the following limit dislocation equations:*

$$\begin{cases} X \stackrel{\mathcal{L}}{=} e^{-\tau m}([a+1]X + [b]Y), \\ Y \stackrel{\mathcal{L}}{=} e^{-\tau m}([c]X + [d+1]Y), \end{cases}$$

with

$$\mathbb{E}(X) = \frac{b}{S},\ \mathbb{E}(Y) = -\frac{c}{S},$$

where all mentioned random variables are independent.

Proof. We take the finite time dislocation equations from Equations 17.5 and project onto the second eigenline and scale correctly, this gives us

$$\begin{cases} e^{-tm}u_2(\mathscr{U}_{(1,0)}(t)) \stackrel{\mathcal{L}}{=} & [a+1]e^{-(t-\tau)m}u_2(\mathscr{U}_{(1,0)}(t-\tau)) \\ & +[b]e^{-(t-\tau)m}u_2(\mathscr{U}_{(0,1)}(t-\tau))\ a.s.; \\ e^{-tm}u_2(\mathscr{U}_{(0,1)}(t)) \stackrel{\mathcal{L}}{=} & [c]e^{-(t-\tau)m}u_2(\mathscr{U}_{(1,0)}(t-\tau)) \\ & +[d+1]e^{-(t-\tau)m}u_2(\mathscr{U}_{(0,1)}(t-\tau))\ a.s. \end{cases}$$

Now, taking the limit $t \to \infty$ leads to

$$\begin{cases} X \stackrel{\mathcal{L}}{=} e^{-\tau m}([a+1]X + [b]Y), \\ Y \stackrel{\mathcal{L}}{=} e^{-\tau m}([c]X + [d+1]Y). \end{cases}$$

We have

$$\mathbb{E}(X) = \mathbb{E}(e^{-tm}u_2(\mathscr{U}_{(1,0)}(t))) = \frac{b}{S} \text{ and } \mathbb{E}(Y) = \mathbb{E}(e^{-tm}u_2(\mathscr{U}_{(0,1)}(t))) = -\frac{c}{S}$$

since we are dealing with the limits of L^p-convergent non-negative martingales (see Theorem I.41). As we already noticed before, the random variables are independent due to the properties of a continuous time Pólya urn process. □

18 Asymptotic behaviour of the continuous time multitype branching process with $m-1$ types

We derive the asymptotic behaviour of a large continuous time multitype branching process with $m-1$ types and prove a relation to the discrete time multitype branching process with $m-1$ types.

Notation 18.1. We use the same notations as in sections 12, 13 and 15. We write $(X_n)_{n\in\mathbb{N}_0}$ resp. $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ for the discrete resp. continuous time multitype branching process with $m-1$ types. The jump times of $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ are denoted by $(\tau_n)_{n\in\mathbb{N}_0}$ (see Definition 15.3). The corresponding composition matrix is A. We denote by v_1 and v_2 the eigenvectors of A with corresponding eigenforms u_1 and u_2, defined in Definition 12.15. The eigenvalues of A are denoted by $\lambda_1, \dots \lambda_{m-1}$. They are ordered with decreasing real parts, i.e. $1 = \lambda_1 > \Re(\lambda_2) \geq \Re(\lambda_3) \geq \cdots \geq \Re(\lambda_{m-1})$. Let $(v_\lambda)_{\lambda\in \mathrm{Sp}(A)}$ be a basis of eigenvectors, each v_λ being an eigenvector to the corresponding (complex) eigenvalue λ. Let $(u_\lambda)_{\lambda\in \mathrm{Sp}(A)}$ be the dual basis of $(v_\lambda)_{\lambda\in \mathrm{Sp}(A)}$, meaning that

$$u_\lambda(v_\mu) = \begin{cases} 1 & \text{if } \lambda = \mu; \\ 0 & \text{if } \lambda \neq \mu. \end{cases}$$

We choose $u_{\lambda_2} = u_2$ and $v_{\lambda_2} = v_2$. The process is called large if $\sigma_2 = \Re(\lambda_2) > \frac{1}{2}$. Furthermore, we recall the limit random variable

$$W = \lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2(X_n),$$

defined in Remark 13.42.

18.1 Asymptotics of a large multitype branching process with $m-1$ types in continuous time

Remark 18.2 ([17])**.** We recall the transition operator we used in section 13 and which in this case is defined on the space of all functions $f\colon \mathbb{C}^{m-1} \to W$, where W is any real or complex vector space. We have

$$\Phi(f)(v) = \sum_{k=1}^{m-1} k l_k(v)(f(v + w_k) - f(v)),$$

where $v \in \mathbb{C}^{m-1}$ and l_k are the coordinate forms $l_k(x_1, \dots, x_{m-1}) = x_k$.

Remark 18.3 ([17])**.** We recall from Remark 16.4, that the projections of $(e^{-tA}\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ to the corresponding eigenspaces of the eigenvalue λ of A are bounded in L^p for every $p \geq 1$ if $\Re(\lambda) > \frac{1}{2}\lambda_1$, where λ_1 is the largest real eigenvalue of A. In our case, this means that $u_1(e^{-tA}\mathscr{X}(t))$ and $u_2(e^{-tA}\mathscr{X}(t))$ converge in L^p for $p \geq 1$, if $\sigma_2 = \Re(\lambda_2) > \frac{1}{2}$ (see Remark 12.12).

Definition 18.4 ([17])**.** We define the random variables ξ and $\mathscr{W}$ by

$$\xi = \lim_{t\to\infty} u_1(e^{-tA}\mathscr{X}(t)) = \lim_{t\to\infty} e^{-t}u_1(\mathscr{X}(t)), \tag{18.1}$$

$$\mathscr{W} = \lim_{t\to\infty} u_2(e^{-tA}\mathscr{X}(t)) = \lim_{t\to\infty} e^{-t\lambda_2}u_2(\mathscr{X}(t)). \tag{18.2}$$

Remark 18.5 ([17])**.** For $t \in \mathbb{R}_{\geq 0}$, we write the vector $\mathscr{X}(t)$ in four terms according to the spectral decomposition of A,

$$\mathscr{X}(t) = \sum_{\lambda\in \mathrm{Sp}(A)} u_\lambda(\mathscr{X}(t))v_\lambda = \mathscr{X}_1(t) + \mathscr{X}_2(t) + \mathscr{X}_3(t) + \mathscr{X}_4(t),$$

where

$$\begin{cases} \mathscr{X}_1(t) = u_1(\mathscr{X}(t))v_1, \\ \mathscr{X}_2(t) = u_{\lambda_2}(\mathscr{X}(t))v_{\lambda_2} + u_{\overline{\lambda_2}}(\mathscr{X}(t))v_{\overline{\lambda_2}}, \\ \mathscr{X}_3(t) = \sum_{\frac{1}{2}<\Re(\lambda)<\Re(\lambda_2)} u_\lambda(\mathscr{X}(t))v_\lambda, \\ \mathscr{X}_4(t) = \sum_{\Re(\lambda)<\frac{1}{2}} u_\lambda(\mathscr{X}(t))v_\lambda. \end{cases}$$

This partition of $\mathrm{Sp}(A)$ is valid because $\frac{1}{2}$ can not be an eigenvalue of A, this follows from the characteristic polynomial of A (see Remark 12.12).

Lemma 18.6 ([17] Lemma 4.2)**.** *Suppose that λ is an eigenvalue such that $\Re(\lambda) < \frac{1}{2}$ and let $\eta > 0$. Then, $e^{-\left(\frac{1}{2}+\eta\right)t}u_\lambda(\mathscr{X}(t))$ is bounded almost surely and in L^p, $p \geq 1$.*

Sketch of the proof. The proof of the lemma uses that as long as t lies in the interval $[\tau_n, \tau_{n+1})$, the vector $\mathscr{X}(t)$ is equal to X_n. We make use of the moment bounds of the discrete time process, which we stated in part (i) of Theorem 13.33. When $\Re(\lambda) < \frac{1}{2}$,

$$\forall p \geq 1,\ \forall \varepsilon > 0,\ \mathbb{E}(|u_\lambda(X_n)|^p) = O\left(n^{p\left(\frac{1}{2}+\varepsilon\right)}\right) \text{ as } n \to \infty.$$

For the almost sure bound, we prove that

$$\lim_{C\to\infty} \mathbb{P}\left(\exists t > 0,\ e^{-\left(\frac{1}{2}+\varepsilon\right)}|u_\lambda(X_n)| > C\right) = 0,$$

for $C > 0$, $\eta > 0$ and an eigenvalue λ with $\Re(\lambda) < \frac{1}{2}$. This is done using that $\mathscr{X}(t)$ does not change for $t \in [\tau_n, \tau_{n+1})$, the Markov inequality, the fact that τ_n and $\mathscr{X}(\tau_n)$ are independent and the Laplace transform of τ_n.
For the boundedness in L^p, we let $p \geq 1$, $t \in \mathbb{R}_{>0}$ and use again the relation with the discrete time process $(X_n)_{n\in\mathbb{N}_0}$, the independence of τ_n and $\mathscr{X}(\tau_n)$, the density of τ_n and Stirling's approximation (see Theorem I.50). All details can be found in [17] in the proof of Lemma 4.2. □

We finally state the theorem about the asymptotic behaviour of a large continuous time multitype branching process.

Theorem 18.7 ([17] Theorem 4.1)**.** *Suppose that $m \geq 27$. Then, as $t \to \infty$,*

$$\mathscr{X}(t) = e^t \xi v_1(1+\varepsilon_1(t)) + 2\Re(e^{\lambda_2 t}\mathscr{W} v_2)(1+\varepsilon_2(t)) + e^{\lambda_2 t}\varepsilon_3(t), \tag{18.3}$$

where

1. *ξ is a positive* Gamma$(N_0, 1)$*-distributed random variable with expectation $N_0 = u_1(\mathscr{X}(0))$ (the total weighted number of particles at time 0);*
2. *$\mathscr{W}$ is a complex-valued random variable that admits moments of all orders $p \geq 1$ and whose expectation equals $u_2(\mathscr{X}(0))$;*
3. *the real-valued random variables $\varepsilon_1(t)$ and $\varepsilon_2(t)$ and the real random vector $\varepsilon_3(t)$ tend to 0 almost surely and in any L^p-space, $p \geq 1$, as $t \to \infty$.*

Proof. Let $t \in \mathbb{R}_{\geq 0}$. We split the vector $\mathscr{X}(t)$ into four terms as in Remark 18.5,

$$\mathscr{X}(t) = \sum_{\lambda \in \mathrm{Sp}(A)} u_\lambda(\mathscr{X}(t))v_\lambda = \mathscr{X}_1(t) + \mathscr{X}_2(t) + \mathscr{X}_3(t) + \mathscr{X}_4(t).$$

Now, we deal with each term separately. We define $\varepsilon_3(t) = \mathscr{X}_3(t) + \mathscr{X}_4(t)$ for all $t \in \mathbb{R}_{\geq 0}$. The Equations 18.1 and 18.2 give us directly the aymptotics of $\mathscr{X}_1(t)$ and $\mathscr{X}_2(t)$, i.e.

$$\begin{aligned}
\mathscr{X}_1(t) &= (e^t\xi + o(e^t))v_1, \\
\mathscr{X}_2(t) &= (e^{\lambda_2 t}W + o(e^{\lambda_2 t}))v_2 + \overline{(e^{\lambda_2 t}W + o(e^{\lambda_2 t}))v_2} \\
&= 2\Re((e^{\lambda_2 t}W + o(e^{\lambda_2 t}))v_2).
\end{aligned}$$

This gives us the first two terms in Equation 18.3.
Let λ be an eigenvalue of A such that $\frac{1}{2} < \Re(\lambda) < \Re(\lambda_2)$. Then,

$$(u_\lambda(e^{-tA}\mathscr{X}(t)))_{t\in\mathbb{R}_{\geq 0}} = (e^{-t\lambda}u_\lambda(\mathscr{X}(t)))_{t\in\mathbb{R}_{\geq 0}}$$

is a convergent martingale, bounded in L^p for $p \geq 1$ (see Remark 16.4). When $t \to \infty$, we have almost surely and in L^p, $p \geq 1$, that $u_\lambda(\mathscr{X}(t)) = o(e^{t\lambda_2})$. Hence we have $\mathscr{X}_3(t) = o(e^{t\lambda})$ as $t \to \infty$.
For the case where $\Re(\lambda) < \frac{1}{2}$, we use Lemma 18.6. It follows that if $\Re(\lambda) < \frac{1}{2}$, then

$$e^{-\lambda_2 t}u_\lambda(\mathscr{X}(t)) = e^{\left(\frac{1}{2}+\eta-\lambda_2\right)t}\left(e^{-\left(\frac{1}{2}+\eta\right)t}u_\lambda(\mathscr{X}(t))\right) \to 0 \text{ as } t \to \infty,$$

almost surely as soon as $0 < \eta < \Re(\lambda_2) - \frac{1}{2}$. Such η exist because $\Re(\lambda_2) > \frac{1}{2}$. Therefore we have $\mathscr{X}_4(t) = o(e^{\lambda_2 t})$ for $t \to \infty$. The same argument holds for the L^p convergence. Using the transition operator (see Remark 18.2), the developed theory about Pólya processes (see section 13, especially Proposition 13.35) and the same arguments as for the proof of Theorem 17.4 (see also Lemma 17.3), we find a reduced polynomial Q_{pe_1} such that $\mathbb{E}(\xi^p) = Q_{pe_1}(\mathscr{X}(0))$ for all $p \in \mathbb{N}_0$. Moreover, this polynomial has the form

$$Q_{pe_1}(\xi^p) = u_1(u_1+1)\dots(u_1+p-1)$$

for $p \in \mathbb{N}_0$ (see Remark 13.36), which implies that (similar to the proof of Theorem 17.4)

$$\mathbb{E}(\xi^p) = \frac{\Gamma(N_0 + p)}{\Gamma(N_0)},$$

where $N_0 = u_1(\mathscr{X}(0))$ is the weighted number of particles at time 0. It is a well-known fact that the gamma distribution is characterized by its moments, which shows that ξ is Gamma$(N_0, 1)$-distributed (see Definition 2.20). This completes the proof. □

18.2 A martingale connection

Proposition 18.8 ([17] Proposition 4.4)**.** *The following two assertions hold.*

$$\lim_{n\to\infty} ne^{-\tau_n} = \xi \text{ a.s. and in } L^p, \forall p \geq 1, \tag{18.4}$$

$$\mathscr{W} = \xi^{\lambda_2} W \text{ a.s. with } \xi \text{ and } W \text{ independent.} \tag{18.5}$$

Remark 18.9 ([17])**.** Equation 18.5 is also called martingale connection and establishes a link between $\mathscr{W}$ and W.

Proof. We first show that Equation 18.4 holds. We apply the first projection to the embedding principle (Equation 15.1) and obtain

$$u_1(\mathscr{X}(\tau_n)) = u_1(X_n) \text{ a.s.}$$

By definition of u_1 (see Definition 12.15), u_1 gives us the total number of free places at time τ_n and is equal to $N_0 + n - 1 = n(1 + o(1))$. By Equation 18.1 and the fact that $\tau_n \to \infty$ almost surely when $n \to \infty$, we get

$$\xi = \lim_{t\to\infty} e^{-t} u_1(\mathscr{X}(t)) = \lim_{n\to\infty} ne^{-\tau_n} \text{ a.s.},$$

which proves Equation 18.4.
To prove Equation 18.5, we apply the second projection to the embedding principle (Equation 15.1) and obtain

$$u_2(\mathscr{X}(\tau_n)) = u_2(X_n) \text{ a.s.},$$

with u_2 defined as in Definition 12.15. We use Equation 18.2 and the fact that $\tau_n \to \infty$ almost surely when $n \to \infty$. Then we get

$$\mathscr{W} = \lim_{t\to\infty} e^{-\lambda_2 t} u_2(\mathscr{X}(t)) = \lim_{n\to\infty} e^{-\lambda_2 \tau_n} u_2(X_n).$$

Equation 18.5 follows now from Equation 18.4 and the asymptotics in discrete time we proved in Theorem 13.41 resp. Remark 13.42. More precisely, we have by linearity of u_2,

$$\begin{aligned}\mathscr{W} &= \lim_{n\to\infty} e^{-\lambda_2 \tau_n} u_2(X_n) \\ &= \lim_{n\to\infty} n^{\lambda_2} e^{-\lambda_2 \tau_n} \cdot n^{-\lambda_2} \cdot u_2(X_n) = \xi^{\lambda_2} \cdot W,\end{aligned}$$

which is a martingale connection in finite time due to the definition of W (see Remark 13.42). Since $\xi = \lim_{n\to\infty} ne^{-\tau_n}$, $W = \lim_{n\to\infty} n^{-\lambda_2} u_2(X_n)$ and $(\tau_n)_{n\in\mathbb{N}_0}$ and $(X_n)_{n\in\mathbb{N}_0}$ are independent, the variables ξ and W are independent as well. □

19 Asymptotic behaviour of Pólya urns and m-ary search trees in discrete time part II

We prove a convergence result for small discrete time Pólya urn and m-ary search tree processes.

We consider the processes in a more general setting, which allows us to prove the results simultaneously for both processes. Moreover, we make use of the embedding principle and the continuous time processes.

19.1 Notations

We use the following notations.

Notation 19.1 ([31])**.** We write $(X_n)_{n\in\mathbb{N}_0}$ for the composition vectors of either the discrete time 2-dimensional Pólya urn process defined in section 11 or for the discrete time m-ary search tree process defined in section 12, with $m \in \mathbb{N}$, $m \geq 3$. In the case of a Pólya urn process, we denote by A the transpose of the corresponding replacement matrix, whereas for an m-ary search tree, A is the corresponding composition matrix. We denote the dimension of the composition vector by $s \geq 2$.

Notation 19.2 ([31])**.** If we draw a ball of type i, $i = 1, 2$, in the Pólya urn process, we add w_{ij} balls of type j. This gives us the vectors $w_i = (w_{ij})_{1\leq j\leq 2}$ for $i = 1, 2$.
If we insert a new key in a node of type i in an m-ary search tree, we add w_{ij} nodes of type j. This gives us the vectors $w_i = (w_{ij})_{1\leq j\leq m-1}$ for $i = 1, \ldots, m-1$.

Notation 19.3 ([31])**.** The corresponding continuous time process is written $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ (i.e. for the continuous time Pólya urn process defined in section 14 or for the continuous time multitype branching process with $m-1$ types defined in section 15). We denote by τ_n the n-th jump time of the process for $n \in \mathbb{N}$ and $\tau_0 = 0$. The lifetime of a ball or particle of type i is $\mathrm{Exp}(a_i)$-distributed, so in terms of a Pólya urn process, we have $a_i = 1$ for all $i = 1, 2$ and in the case of a multitype branching process with $m-1$ types, we have $a_i = i$ for $i = 1, \ldots, m-1$. We set $a = (a_i)_{1\leq i\leq s}$.

Notation 19.4 ([31])**.** The eigenvalues of A are denoted by $\lambda_1, \ldots \lambda_s$. They are ordered with decreasing real parts, i.e. $\lambda_1 > \Re(\lambda_2) \geq \Re(\lambda_3) \geq \cdots \geq \Re(\lambda_s)$. We write u_1, u_2 (resp. v_1, v_2) for the eigenforms (resp. eigenvectors) of the corresponding eigenvalue λ_1 (resp. λ_2), which we defined in Remark 11.4 for Pólya urns and in Definition 12.15 for m-ary search trees, they satisfy $u_1(v_1) = u_2(v_2) = 1$, $u_2(v_1) = u_1(v_2) = 0$.
We have a decomposition into eigenspaces of the eigenvalues and denote by π_λ the projection into the eigenspace corresponding to the eigenvalue λ.

Notation 19.5. We remember from sections 17 and 18 the martingale limit

$$\xi = \lim_{t\to\infty} u_1(e^{-tA}\mathscr{X}(t)) = \lim_{t\to\infty} e^{-t\lambda_1}u_1(\mathscr{X}(t)).$$

Notation 19.6 ([31])**.** We define the set

$$\Lambda_I = \left\{ \lambda \in \mathrm{Sp}(A) \mid \Re(\lambda) < \frac{\lambda_1}{2} \right\}.$$

We denote by π_I the projection onto the sum of eigenspaces with $\Re(\lambda) < \frac{\lambda_1}{2}$, $\pi_I = \sum_{\lambda \in \Lambda_I} \pi_\lambda$.

Notation 19.7 ([31])**.** The space of right-continuous functions on $\mathbb{R}$ with left-hand limits is denoted by $D(\mathbb{R})$.

Remark 19.8. We remember that we defined the processes such that they can not become extinct.

19.2 First results

We state and prove several results to for the main statement of this section.

Theorem 19.9 ([31] Theorem 3.1)**.** *For $t \to \infty$, as $e^{-\lambda_1 t}\mathscr{X}(t) \to \xi v_1$ almost surely and, with joint convergence in distribution,*

$$e^{-\frac{\lambda_1(t+x)}{2}} \pi_I(\mathscr{X}(t+x)) \xrightarrow{\mathcal{L}} \xi^{\frac{1}{2}} U_I(x) \text{ in } D(\mathbb{R});$$

equivalently,

$$e^{-\frac{\lambda_1 t}{2}} \pi_I(\mathscr{X}(t+x)) \xrightarrow{\mathcal{L}} \xi^{\frac{1}{2}} e^{\frac{\lambda_1 x}{2}} U_I(x) \text{ in } D(\mathbb{R}).$$

Here, U_I is a real Gaussian vector-valued stochastic process with $U_I(x)$ defined for $x \in \mathbb{R}$. The process U_I is characterized by expectation 0 and a computable covariances.

Sketch of the proof. The proof of this statement is done in several steps. The main idea is to consider the quadratic variation, resp. covariation of martingales, a convergence in probability to a matrix-valued function and a convergence in distribution to a Gaussian process with expectation 0 and computable covariances. It uses the martingales we have found in Theorem 16.3. Their asymptotic behaviour is studied, especially boundedness in L^p for $p \geq 1$. Moreover, the projection into eigenspaces is used on the processes and the martingales. All details can be found in the proof of Theorem 3.1 in [31]. □

Remark 19.10 ([31])**.** The explicit form of the covariances of U_I is mentioned in Theorem 3.1. in [31]. However, we are not interested in the exact result. It is enough to know that the covariances can be computed.

Definition 19.11 ([31])**.** Let $b \in \mathbb{R}^s$ be a fixed vector and $z \in \mathbb{R}_{\geq 0}$. We define

$$\tau_b(z) := \min\{t \in \mathbb{R}_{\geq 0} \mid \langle b, \mathscr{X}(t)\rangle \geq z\}, \text{ (with } \min(\emptyset) = +\infty).$$

Lemma 19.12 ([31] Lemma 3.14)**.** *Assume $\langle b, v_1\rangle > 0$. Conditioned on non-extinction, we have almost surely $\langle b, \mathscr{X}(t)\rangle \to \infty$ as $t \to \infty$ and thus $0 \leq \tau_b(z) < \infty$. Moreover, $\tau_b(z) \to \infty$ as $z \to \infty$.*

Lemma 19.13 ([31] Lemma 9.7)**.** *We have $\xi \geq 0$ almost surely, $\mathbb{P}(\xi > 0) > 0$ and $\mathbb{P}(\xi = w) = 0$ for any $w \neq 0$. Furthermore,*

(i) as $t \to \infty$, $u_1(e^{-tA}\mathscr{X}(t)) = e^{-t\lambda_1}u_1(\mathscr{X}(t)) \to \xi$ a.s.

(ii) if $u_1(e^{-tA}\mathscr{X}(t)) = 0$ for some t then $\xi = 0$ almost surely.

(iii) Conversely, $\mathbb{P}(u_1(e^{-tA}\mathscr{X}(t)) > 0$ and $\xi = 0) \to 0$ as $t \to \infty$. Consequently, $\xi = 0$ almost surely if and only if $\mathscr{X}$ becomes extinct.

Sketch of the proof. The statements are proved using the projections onto eigenspaces. All details can be found in the proof of Lemma 9.7 in [31]. □

Proof (of Lemma 19.12). Theorem 19.9 implies

$$e^{-\lambda_1 t}\langle b, \mathscr{X}(t)\rangle \to \langle b, v_1\rangle \xi \text{ a.s. as } t \to \infty, \tag{19.1}$$

which proves that $\langle b, \mathscr{X}(t)\rangle \to \infty$ almost surely when $\xi > 0$. The convergence $\tau_b(z) \to \infty$ when $z \to \infty$ follows directly. We notice that

$$\sup_{s \leq t}(\langle b, \mathscr{X}(s)\rangle) < \infty$$

for any $t \in \mathbb{R}_{\leq 0}$, which implies that $0 \leq \tau_b(z) < \infty$. □

Lemma 19.14 ([31] Lemma 11.1)**.** *Let $b \in \mathbb{R}^s$ with $\langle b, v_1\rangle > 0$. As $z \to \infty$,*

$$\tau_b(z) - \frac{1}{\lambda_1}\log(z) \to \frac{1}{\lambda_1}(\log(\xi) + \log(\langle b, v_1\rangle))$$

almost surely (with the convention that $\log(0) = -\infty$).

Proof. Let E be the event where $\langle b, \mathscr{X}(t)\rangle \to \infty$. On E, $\tau_b(z) < \infty$ for all z and $\tau_b(z) \to \infty$ when $z \to \infty$. So by Equation 19.1, we have on E

$$e^{-\lambda_1 \tau_b(z)}\langle b, \mathscr{X}(\tau_b(z))\rangle \to \langle b, v_1\rangle \xi \text{ a.s. as } z \to \infty. \tag{19.2}$$

Since the process $\mathscr{X}(t)$ is right-continuous, $\langle b, \mathscr{X}(\tau_b(z))\rangle \geq z$. If, without loss of generality, $\varepsilon(z) = \frac{1}{z}$, and z is large enough such that $\tau_b(z) > \varepsilon(z)$, then $\langle b, \mathscr{X}(\tau_b(z) - \varepsilon(z))\rangle < z$ and by Equation 19.1, on E,

$$e^{-\lambda_1 \tau_b(z)}\langle b, \mathscr{X}(\tau_b(z) - \varepsilon(z))\rangle \to \langle b, v_1\rangle \xi \text{ a.s. as } z \to \infty. \tag{19.3}$$

We combine Equations 19.2 and 19.3 to get that on E,

$$e^{-\lambda_1 \tau_b(z)} z \to \langle b, v_1\rangle \xi \text{ a.s. as } z \to \infty. \tag{19.4}$$

On the complement of E, we have $\xi = 0$ almost surely by Lemma 19.12 and $\tau_b(z) = \infty$ for large z. Therefore Equation 19.4 holds. Altogether, Equation 19.4 holds almost surely. We take the logarithm and divide by $-\lambda_1$ to obtain the desired equation. □

Lemma 19.15 ([31] Lemma 11.2)**.** *As $t \to \infty$, $e^{-\frac{\lambda_1 t}{2}} \sup_{s \le t} |\Delta \mathscr{X}(s)| \to 0$ almost surely, where $\Delta \mathscr{X}(s) = \mathscr{X}(s) - \mathscr{X}(s-)$ is the jump of $\mathscr{X}$ at s.*

Sketch of the proof. The statement is proven by showing first that $\sum_{n \in \mathbb{N}_0} \mathbb{P}(E_n) < \infty$, where E_n is the event $\sup_{t \in [n,n+1]} e^{-\frac{\lambda_1 t}{2}} \sup_{s \le t} |\Delta \mathscr{X}(s)| > \varepsilon$ for $\varepsilon > 0$. Applying the Borel-Cantelli lemma proves the statement. More details can be found in the proof of Lemma 11.2 of [31]. □

Lemma 19.16 ([31] Lemma 11.3)**.** *Conditioned on $\xi > 0$, $z^{-\frac{1}{2}} \sup_{s \le \tau_b(z)} |\Delta \mathscr{X}(s)| \to 0$ almost surely as $z \to \infty$.*

Proof. If we condition on $\xi > 0$, it follows that $\tau_b(z) \to \infty$ almost surely by Lemma 19.12 and therefore Lemma 19.15 implies $e^{-\frac{\lambda_1 \tau_b(z)}{2}} \sup_{s \le \tau_b(z)} |\Delta \mathscr{X}(s)| \to 0$ almost surely. Additionally, Equation 19.4 from Lemma 19.14 says that $z^{-\frac{1}{2}} e^{\frac{\lambda_1 \tau_b(z)}{2}} \to \langle b, v_1 \rangle^{-\frac{1}{2}} \xi^{-\frac{1}{2}}$ almost surely. We multiply these two terms and get as $z \to \infty$,

$$z^{-\frac{1}{2}} e^{\frac{\lambda_1 \tau_b(z)}{2}} e^{-\frac{\lambda_1 \tau_b(z)}{2}} \sup_{s \le \tau_b(z)} |\Delta \mathscr{X}(s)| = z^{-\frac{1}{2}} \sup_{s \le \tau_b(z)} |\Delta \mathscr{X}(s)| \to 0 \cdot \langle b, v_1 \rangle^{-\frac{1}{2}} \xi^{-\frac{1}{2}} = 0 \ a.s.$$

□

Theorem 19.17 ([31] Theorem 3.15)**.** *Let $b \in \mathbb{R}^s$ with $\langle b, v_1 \rangle > 0$. Conditioned on non-extinction, we have as $z \to \infty$,*

$$z^{-1} \mathscr{X}(\tau_b(z)) \to \langle b, v_1 \rangle^{-1} v_1 \ a.s., \tag{19.5}$$

and, with joint convergence in distribution of all random vectors,

(i) $z^{-\frac{1}{2}} \pi_I(\mathscr{X}(\tau_b(z))) \xrightarrow{\mathcal{L}} \langle b, v_1 \rangle^{-\frac{1}{2}} V_I$;

(ii)

$$z^{-\frac{1}{2}} \left(\pi_{\lambda_1}(\mathscr{X}(\tau_b(z))) - \frac{z}{\langle b, v_1 \rangle} v_1 + \sum_{\lambda \in \mathrm{Sp}(A) \setminus \{\lambda_1\}} \frac{\langle b, \pi_\lambda(\mathscr{X}(\tau_b(z))) \rangle}{\langle b, v_1 \rangle} v_1 \right) \to 0 \ a.s.$$

(iii)

$$z^{-\frac{1}{2}} \left(\mathscr{X}(\tau_b(z)) - \frac{z}{\langle b, v_1 \rangle} v_1 - \left(I - \frac{v_1 \cdot b^T}{\langle b, v_1 \rangle} \right) \sum_{\lambda \in \mathrm{Sp}(A) \setminus \{\lambda_1\}} \pi_\lambda(\mathscr{X}(\tau_b(z))) \right) \to 0 \ a.s.$$

Here, V_I is a real, vector-valued random variable with Gaussian distribution. It is characterized by mean 0 and computable covariances.
The same result holds for $X_{\tau_b(z)}$.

Proof. Since our processes do not go extinct, we can condition on non-extinction, which is by Lemma 19.13 the same as $\xi > 0$. Now, let $z \to \infty$. We have that $\tau_b(z) \to \infty$ and by Theorem 19.9, $e^{-\lambda_1 \tau_b(z)} \mathscr{X}(\tau_b(z)) \to \xi v_1$ almost surely. We divide by Equation 19.4 and get

$$z^{-1} \mathscr{X}(\tau_b(z)) \to \langle b, v_1 \rangle^{-1} v_1 \ a.s.,$$

which is exactly the desired Equation 19.5.
We set $t = \lambda_1^{-1} \log(z)$ and $x = \tau_b(z) - t$. Then $t \to \infty$ and by Lemma 19.14, $x \to x_0 \coloneqq -\lambda_1^{-1}(\log(\xi) + \log(\langle b, v_1 \rangle))$ almost surely. Hence Theorem 19.9 leads to

$$e^{-\frac{\lambda_1 t}{2}} \pi_I(\mathscr{X}(t+x)) \xrightarrow{\mathcal{L}} e^{\frac{\lambda_1 x_0}{2}} \xi^{\frac{1}{2}} U_I(x_0) = \langle b, v_1 \rangle^{-\frac{1}{2}} U_I(x_0).$$

Here, x_0 is random, but U_I is independent of ξ and of x_0, which gives us $U_I(x_0) \stackrel{\mathcal{L}}{=} U_I(0)$. Therefore, part (i) of our theorem follows with $V_I = U_I(0)$.
For part (ii), we have

$$|\langle b, \mathscr{X}(\tau_b(z)) \rangle - z| \leq |b| |\Delta \mathscr{X}(\tau_b(z))|$$

and so

$$z^{-\frac{1}{2}}(\langle b, \mathscr{X}(\tau_b(z)) \rangle - z) \to 0$$

almost surely by Lemma 19.16. We multiply this by $\langle b, v_1 \rangle^{-1} v_1$, write $\mathscr{X} = \sum_{\lambda \in \mathrm{Sp}(A)} \pi_\lambda(\mathscr{X})$ and notice that for every $v \in \mathrm{Im}(\pi_{\lambda_1})$, $\frac{\langle b, v \rangle}{\langle b, v_1 \rangle} v_1 = v$. Then, the equation becomes

$$z^{-\frac{1}{2}} \left(\pi_{\lambda_1}(\mathscr{X}(\tau_b(z))) - \frac{z}{\langle b, v_1 \rangle} v_1 + \frac{\langle b, \sum_{\lambda \in \mathrm{Sp}(A) \setminus \{\lambda_1\}} \pi_\lambda(\mathscr{X}(\tau_b(z))) \rangle}{\langle b, v_1 \rangle} v_1 \right) \to 0 \ a.s.,$$

which is what we wanted to show, using linearity of the scalar product.
For part (iii), we use the identity $\mathscr{X} = \sum_{\lambda \in \mathrm{Sp}(A)} \pi_\lambda(\mathscr{X})$ respectively

$$\pi_{\lambda_1}(\mathscr{X}) = \sum_{\lambda \in \mathrm{Sp}(A)} \pi_\lambda(\mathscr{X}) - \sum_{\lambda \in \mathrm{Sp}(A) \setminus \{\lambda_1\}} \pi_\lambda(\mathscr{X}),$$

which gives us with part (ii)

$$z^{-\frac{1}{2}} \left(\mathscr{X}(\tau_b(z)) - \frac{z}{\langle b, v_1 \rangle} v_1 - \left(I - \frac{v_1 \cdot b^T}{\langle b, v_1 \rangle} \right) \sum_{\lambda \in \mathrm{Sp}(A) \setminus \{\lambda_1\}} \pi_\lambda(\mathscr{X}(\tau_b(z))) \right) \to 0 \ a.s.$$

This completes the proof. □

Corollary 19.18 ([31] Corollary 3.16)**.** *Let $b \in \mathbb{R}^s$ with $\langle b, v_1 \rangle > 0$. We assume that $\Re(\lambda_2) < \frac{1}{2}\lambda_1$. Conditioned on non-extinction, we have as $z \to \infty$,*

$$z^{-\frac{1}{2}} \left(\mathscr{X}(\tau_b(z)) - \frac{z}{\langle b, v_1 \rangle} v_1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

where Σ can be computed.
The same result holds for $X_{\tau_b(z)}$.

Proof. Let R be a function such that $R(z) \to 0$ almost surely for $z \to \infty$. Since $\mathrm{Sp}(A) = \Lambda_I \cup \{\lambda_1\}$, we have by part (iii) of Theorem 19.17 that

$$z^{-\frac{1}{2}} \left(\mathscr{X}(\tau_b(z)) - \frac{z}{\langle b, v_1 \rangle} v_1 \right) = z^{-\frac{1}{2}} \left(I - \frac{v_1 \cdot b^T}{\langle b, v_1 \rangle} \right) \pi_I(\mathscr{X}(\tau_b(z))) + R(z).$$

With part (i) of Theorem 19.17, we get a convergence to the Gaussian limit

$$\langle b, v_1 \rangle^{-\frac{1}{2}} \left(I - \frac{v_1 \cdot b^T}{\langle b, v_1 \rangle} \right) V_I,$$

where the covariances can be computed. □

19.3 Asymptotics of small discrete time Pólya urn and m-ary search tree processes

Theorem 19.19 ([31] Theorem 3.21)**.** *Conditioned on non-extinction, we have*

$$n^{-1} X_n \to \lambda_1 v_1 \ a.s.$$

as $n \to \infty$.

Proof. We extend our composition vector by a dummy variable of a new type $s+1$. Whenever the process splits resp. changes, we add $+1$ to the dummy variable, such that it counts the number of splits. We have $a_{s+1} = 0$ and $w_{s+1} = 0$. This does not affect the rest of the process. We write $\tilde{}$ over the symbols to denote the modified process and the quantities defined for it. If we write the vectors and matrices in block form, corresponding to a split $\mathbb{R}^s \times \mathbb{R}$, we have for $i = 1, \ldots, s$,

$$\tilde{w}_i = \begin{pmatrix} w_i \\ 1 \end{pmatrix}, \ \tilde{a} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \ \tilde{A} = \begin{pmatrix} A & 0 \\ a^T & 0 \end{pmatrix}.$$

The eigenvalues are now $\mathrm{Sp}(\tilde{A}) = \mathrm{Sp}(A) \cup \{0\}$, and thus $\tilde{\lambda}_1 = \lambda_1$. The matrix $\tilde{A}$ has the following eigenvector with corresponding eigenform

$$\tilde{v}_1 = \begin{pmatrix} v_1 \\ \lambda_1^{-1} \end{pmatrix}, \ \tilde{u}_1 = \begin{pmatrix} u_1 \\ 0 \end{pmatrix}. \tag{19.6}$$

If we choose $b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, then $\tau_b(n)$ is the first time we have n splits, i.e. this is the n-th splitting time τ_n. Therefore $\tilde{X}_n = \tilde{\mathscr{X}}(\tau_b(n))$. By definition of the eigenvector (Equation 19.6), we have $\langle b, \tilde{v}_1 \rangle = \lambda_1^{-1}$. We apply Theorem 19.17 to $\tilde{X}_n = \tilde{\mathscr{X}}(\tau_b(n))$, and Equation 19.5 can then be written as

$$n^{-1} \tilde{X}_n \to \lambda_1 \tilde{v}_1 \ a.s.,$$

resp.

$$n^{-1} X_n \to \lambda_1 v_1 \ a.s.,$$

which proves the statement. □

We finally state the theorem about the asymptotic behaviour of small discrete time Pólya urn and m-ary search tree processes.

Theorem 19.20 ([31] Theorem 3.22)**.** *We assume that* $\Re(\lambda_2) < \frac{1}{2}\lambda_1$*. Conditioned on non-extinction, we have as* $n \to \infty$*,*

$$n^{-\frac{1}{2}}(X_n - n\lambda_1 v_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

where Σ *can be computed.*

Proof. We continue in the setting of the proof of Theorem 19.19.
By construction, $\Re(\tilde{\lambda}_2) = \max(\Re(\lambda_2), 0) < \frac{1}{2}\tilde{\lambda}_1$, so we use Corollary 19.18 with $\tilde{X}_n = \tilde{\mathscr{X}}(\tau_b(n))$. This leads to

$$n^{-\frac{1}{2}}\left(\tilde{X}_n - \frac{n}{\langle b, \tilde{v}_1\rangle}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tilde{\Sigma}),$$

resp.

$$n^{-\frac{1}{2}}(\tilde{X}_n - n\lambda_1 \tilde{v}_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tilde{\Sigma}),$$

where $\tilde{\Sigma}$ can be computed. Thus the limit, denoted by $\tilde{V}$, is Gaussian. We ignore now the dummy variable and get $V = \langle \begin{pmatrix} I \\ 0 \end{pmatrix}, \tilde{V} \rangle$. Then

$$n^{-\frac{1}{2}}(X_n - n\lambda_1 v_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

where Σ can be computed. □

20 Smoothing transformations

We continue in the setting of section 18. We study a distributional equation arising from a martingale limit. The unicity of a solution is proven and we characterize the distribution of the solution of this equation.

Notation 20.1. We use the same notations as in sections 15 and 18. We write $(\mathscr{X}(t))_{t\in\mathbb{R}_{\geq 0}}$ for the continuous time multitype branching process with $m-1$ types. The corresponding composition matrix is A. We denote by $\lambda_1 = 1$ and λ_2 the eigenvalues of A as defined in Remark 12.12 and Notation 12.13. The corresponding eigenvectors are v_1 and v_2 with eigenforms u_1 and u_2, defined in Definition 12.15. We assume $\Re(\lambda_2) > \frac{1}{2}$, i.e. we consider a large process.
We recall the martingale limit random variable

$$\mathscr{W} = \lim_{t\to\infty} u_2(e^{-tA}\mathscr{X}(t)) = \lim_{t\to\infty} e^{-t\lambda_2}u_2(\mathscr{X}(t))$$

defined in Equation 18.2.

20.1 A distributional equation

We derive the distributional equation.

Remark 20.2 ([17])**.** We denote by $\mathscr{X}_j(t)$ the process $\mathscr{X}(t)$ started from $\mathscr{X}(0) = e_j$, with e_j being the j-th vector of the canonical basis of $\mathbb{R}^{m-1}$ (see Notation 2.8). So the process starts with one particle of type j.
We let $\tau_{(j)}$, $j = 1, \ldots, m-1$, be the first splitting time when the process starts from $\mathscr{X}(0) = e_j$, which means that $\tau_{(j)}$ is Exp(j)-distributed. When we consider the branching property at the first splitting time, we get

$$\forall t > \tau_1, \begin{cases} \mathscr{X}_1(t) \stackrel{\mathcal{L}}{=} \mathscr{X}_2(t-\tau_{(1)}), \\ \mathscr{X}_2(t) \stackrel{\mathcal{L}}{=} \mathscr{X}_3(t-\tau_{(2)}), \\ \vdots \\ \mathscr{X}_{m-2}(t) \stackrel{\mathcal{L}}{=} \mathscr{X}_{m-1}(t-\tau_{(m-2)}), \\ \mathscr{X}_{m-1}(t) \stackrel{\mathcal{L}}{=} [m]\mathscr{X}_1(t-\tau_{(m-1)}), \end{cases} \tag{20.1}$$

where $[m]X$ denotes the sum of m independent copies of a random variable X.

Definition 20.3 ([17], [24])**.** The Beta distribution Beta(p,q) has probability density function

$$f_{\mathrm{Beta}(p,q)}(u) = \frac{1}{B(p,q)}u^{p-1}(1-u)^{q-1},$$

where B is Euler's Beta function defined as

$$B(p,q) := \int_0^1 u^{p-1}(1-u)^{q-1}\,du = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},\ p,q>0.$$

Proposition 20.4 ([17]). *Let $n \in \mathbb{N}$, $n \geq 2$. Let $s_1, \ldots, s_n$ be independent random variables, where s_j is* Exp(j)*-distributed. Let $S = s_1 + \cdots + s_n$. Then S has density*

$$f_S(u) = ne^{-u}(1-e^{-u})^{n-1}\mathbb{1}_{\mathbb{R}_{>0}}(u),\ u \in \mathbb{R}.$$

Proof. We use the Laplace transform of the random variables (see Definition 2.23). We denote by $\mathcal{L}(X)(u) = \mathbb{E}(e^{-uX})$ the Laplace transform of a random variable X. When X is Exp(λ)-distributed, it has density $f_X(u) = \lambda e^{-\lambda u}\mathbb{1}_{\mathbb{R}_{>0}}(u)$. Therefore its Laplace transform is

$$\begin{aligned}\mathcal{L}(X)(u) = \mathbb{E}(e^{-uX}) &= \int_0^\infty e^{-ux}\lambda e^{-\lambda x}\,dx = \int_0^\infty \lambda e^{-(u+\lambda)x}\,dx \\ &= \frac{\lambda}{u+\lambda}.\end{aligned}$$

Now, using $S = s_1 + \cdots + s_n$, s_j being Exp(j)-distributed and all the s_j being independent of each other, we have

$$\mathcal{L}(S)(u) = \mathbb{E}(e^{-uS}) = \prod_{j=1}^n \mathbb{E}(e^{-us_j}) = \prod_{j=1}^n \frac{j}{u+j} = \frac{n!}{\prod_{j=1}^n (u+j)}.$$

Let Y be a random variable with density $f_Y(u) = ne^{-u}(1-e^{-u})^{n-1}\mathbb{1}_{\mathbb{R}_{>0}}(u)$, $u \in \mathbb{R}$. We compute its Laplace transform

$$\begin{aligned}\mathcal{L}(Y)(u) = \mathbb{E}(e^{-uY}) &= \int_0^\infty e^{-uy} ne^{-y}(1-e^{-y})^{n-1}\,dy \mid v = e^{-y} \\ &= \int_0^1 v^u n(1-v)^{n-1}\,dv \\ &= nB(u+1,n) = n\frac{\Gamma(u+1)\Gamma(n)}{\Gamma(u+n+1)} = n!\cdot\left(\frac{\Gamma(u+n+1)}{\Gamma(u+1)}\right)^{-1} \\ &= n!\cdot\left(\prod_{j=1}^n (u+j)\right)^{-1},\end{aligned}$$

where B denotes Euler's Beta function. Therefore, the Laplace transforms of S and Y are equal, which implies that the density of the two random variables have to be the same, i.e. S has density $f_S(u) = ne^{-u}(1-e^{-u})^{n-1}\mathbb{1}_{\mathbb{R}_{>0}}(u)$, $u \in \mathbb{R}$, as we wanted. □

Corollary 20.5 ([17]). *Let $T = \tau_{(1)} + \cdots + \tau_{(m-1)}$, where the $\tau_{(j)}$ are independent of each other and $\tau_{(j)}$ is* Exp(j)*-distributed. Then T has density function*

$$f_T(u) = (m-1)e^{-u}(1-e^{-u})^{m-2}\mathbb{1}_{\mathbb{R}_{>0}}(u),\ u \in \mathbb{R}. \tag{20.2}$$

Moreover, for any complex number λ such that $\Re(\lambda) > 0$,

$$\mathbb{E}\left(e^{-\lambda T}\right) = \frac{(m-1)!}{\prod_{j=1}^{m-1}(\lambda+j)}, \tag{20.3}$$

in particular,

$$m\mathbb{E}\left(|e^{-\lambda T}|\right)\begin{cases}<1 \text{ if } \Re(\lambda)>1;\\ =1 \text{ if } \Re(\lambda)=1;\\ >1 \text{ if } \Re(\lambda)<1.\end{cases} \tag{20.4}$$

Proof. The reason for the last result is that for a complex number $a+bi\in\mathbb{C}$, we have $|e^{a+bi}|=|e^a|\cdot|e^{bi}|=e^a\cdot 1=e^{\Re(a+bi)}$. Then

$$m\mathbb{E}\left(|e^{-\lambda T}|\right)=m\mathbb{E}\left(e^{-\Re(\lambda)T}\right)=\frac{m!}{\prod_{j=1}^{m-1}(\Re(\lambda)+j)},$$

which behaves exactly as stated. □

Remark 20.6 ([17])**.** We take the random variables $\mathscr{X}_j(t)$ and project with u_2, scale with $e^{-\lambda_2 t}$ and let $t\to\infty$. Then we get the random variables

$$\mathscr{W}_j \coloneqq \lim_{t\to\infty} e^{-\lambda_2 t}u_2(\mathscr{X}_j(t)).$$

As a consequence, the system of Equations 20.1 on $\mathscr{X}_j(t)$ becomes the following new system of distributional equations on $\mathscr{W}_j$

$$\forall t>\tau_1,\begin{cases}\mathscr{W}_1 \stackrel{\mathcal{L}}{=} e^{-\lambda_2\tau_{(1)}}\mathscr{W}_2,\\ \mathscr{W}_2 \stackrel{\mathcal{L}}{=} e^{-\lambda_2\tau_{(2)}}\mathscr{W}_3,\\ \vdots\\ \mathscr{W}_{m-2} \stackrel{\mathcal{L}}{=} e^{-\lambda_2\tau_{(m-2)}}\mathscr{W}_{m-1},\\ \mathscr{W}_{m-1} \stackrel{\mathcal{L}}{=} e^{-\lambda_2\tau_{(m-1)}}[m]\mathscr{W}_1.\end{cases} \tag{20.5}$$

Therefore, $\mathscr{W}_1$ is a solution of the fixed point equation

$$Z \stackrel{\mathcal{L}}{=} e^{-\lambda_2 T}\left(Z^{(1)}+\cdots+Z^{(m)}\right), \tag{20.6}$$

where $Z^{(j)}$ are independent copies of Z, which are also independent of T.

Notation 20.7 ([17])**.** The Fourier transform of a random variable Z is

$$\phi(Z)(t)=\mathbb{E}\left(e^{i\langle t,Z\rangle}\right)=\mathbb{E}\left(e^{i\Re(\bar{t}Z)}\right),\ t\in\mathbb{C},$$

with $\langle\cdot,\cdot\rangle$ being the canonical scalar product (see Definition 2.12). When Z is clear from the context, we just write $\phi(t)$.

Construction 20.8 ([17])**.** We want to derive a version of Equation 20.6 in terms of Fourier transforms. Let $\phi(Z)(t)$ be the corresponding Fourier transform of the random

variable Z. We remember that T, Z and the $Z^{(j)}$ are independent of each other and Z and $Z^{(j)}$ are identically distributed. Then

$$\begin{aligned}\phi\left(e^{-\lambda_2 T}\left(Z^{(1)}+\cdots+Z^{(m)}\right)\right)(t) &= \mathbb{E}\left(e^{i\left\langle t,e^{-\lambda_2 T}\left(Z^{(1)}+\cdots+Z^{(m)}\right)\right\rangle}\right)\\ &= \mathbb{E}\left(e^{i\left\langle t,e^{-\lambda_2 T}Z^{(1)}\right\rangle}\right)^m\\ &= \mathbb{E}\left(e^{i\left\langle te^{-\overline{\lambda_2}T},Z^{(1)}\right\rangle}\right)^m\\ &= \phi^m(Z)(te^{-\overline{\lambda_2}T}).\end{aligned}$$

Now Equation 20.6 implies

$$\phi(Z)(t)=\int_0^\infty \phi^m(Z)(te^{-\overline{\lambda_2}T})f_T(u)\,du,\ t\in\mathbb{C}.$$

Remark 20.9 ([17])**.** To study the equations we derived, we state the following slightly more general versions. Let $\lambda\in\mathbb{C}$, then we consider the equation

$$Z\stackrel{\mathcal{L}}{=}e^{-\lambda T}\left(Z^{(1)}+\cdots+Z^{(m)}\right), \tag{20.7}$$

where $Z^{(j)}$ are independent copies of Z, which are also independent of T. In terms of Fourier transforms, we consider

$$\phi(Z)(t)=\int_0^\infty \phi^m(Z)(te^{-\overline{\lambda}T})f_T(u)\,du,\ t\in\mathbb{C}, \tag{20.8}$$

with f_T as defined in Equation 20.2.

20.2 The smoothing transformation

We show two approaches for a proof of the unicity of a solution to the distributional equation.

Notation 20.10 ([17])**.** Let $C\in\mathbb{C}$ be a complex number. We denote by $\mathcal{M}_2(C)$ the space of probability distributions on $\mathbb{C}$ admitting a second absolute moment and having C as expectation.

Definition 20.11 ([17])**.** Let $\lambda\in\mathbb{C}$ be a complex number. Let T be as defined in Equation 20.2. Let μ be a probability measure on $\mathbb{C}$. We define

$$K(\mu):=\mathcal{L}\left(e^{-\lambda T}\left(Z^{(1)}+\cdots+Z^{(m)}\right)\right), \tag{20.9}$$

with $Z^{(j)}$ random variables of law μ that are independent of each other and of T. We call K the smoothing transformation.

Lemma 20.12 ([17] Lemma 6.1)**.** *If λ is a root of the characteristic polynomial (Equation 12.2) such that $\Re(\lambda)>-\frac{1}{2}$ and if $C\in\mathbb{C}$, then K maps $\mathcal{M}_2(C)$ into itself.*

Proof. We have $\Re(\lambda) > -1$, this implies that $e^{-\lambda T}$ has an expectation. Since λ is a root of the Equation 12.2, we have

$$\prod_{j=1}^{m-1} (\lambda + j) = m!,$$

so by Equation 20.3, $m\mathbb{E}\left(e^{-\lambda T}\right) = 1$. Therefore,

$$\mathbb{E}\left(e^{-\lambda T}\left(Z^{(1)} + \cdots + Z^{(m)}\right)\right) = \mathbb{E}\left(e^{-\lambda T}\right) \cdot m \cdot \mathbb{E}\left(Z^{(1)}\right) = \mathbb{E}\left(Z^{(1)}\right).$$

We have furthermore $\Re(\lambda) > -\frac{1}{2}$, which ensures $\mathbb{E}\left(|e^{-\lambda T}|^2\right) < \infty$. This implies that $K(\mu)$ admits a second absolute moment whenever μ does. □

Definition 20.13 ([17])**.** Let $\mu, \nu \in \mathcal{M}_2(C)$. We define the Wasserstein distance on $\mathcal{M}_2(C)$ by

$$d_2(\mu, \nu) = \left(\min_{(X,Y)} \left(\mathbb{E}\left(|X - Y|^2\right)\right)\right)^{\frac{1}{2}},$$

where we take the minimum over couples of random variables (X, Y) where X has distribution μ and Y has distribution ν. The Kantorovich-Rubinstein Theorem (see [24] Theorem 11.8.2) ensures that the minimum is attained. Together with the distance d_2, $\mathcal{M}_2(C)$ is a complete metric space.

We finally state and prove the first theorem concerning the unicity of a solution of the distributional equation.

Theorem 20.14 ([17] Theorem 6.2)**.** *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (Equation 12.2) such that $\Re(\lambda) > \frac{1}{2}$, and let $C \in \mathbb{C}$. Then K is a contraction on the complete metric space $(\mathcal{M}_2(C), d_2)$, and the fixed point equation (Equation 20.7) has a unique solution Z in $\mathcal{M}_2(C)$.*

Corollary 20.15 ([17] Corollary 6.3)**.** *The distribution of the limit complex random variable $\mathscr{W}_1$ is the unique solution in the space $\mathcal{M}_2(1)$ of the fixed point equation (Equation 20.7)*

Proof (of Corollary 20.15). We have $\mathbb{E}(\mathscr{W}_1) = 1$ and $\mathbb{E}(|\mathscr{W}_1|^2) < \infty$, now we take $\lambda = \lambda_2$ and apply Theorem 20.14. □

Proof (of Theorem 20.14). We use the Banach fixed point theorem (see Theorem I.53), so it is enough to show that K is a contraction. Let $\mu, \nu \in \mathcal{M}_2(C)$. Let (X, Y) be a couple of complex-valued random variables such that $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$ and $d_2(\mu, \nu) = \sqrt{\mathbb{E}(|X - Y|^2)}$. Let (X_j, Y_j), $j = 1, \ldots, m$ be m independent copies of the d_2-optimal couple (X, Y), and T be a real random variable with density f_T as defined in Equation 20.2, independent of all (X_j, Y_j). Then

$$K(\mu) = \mathcal{L}\left(e^{-\lambda T} \sum_{j=1}^{m} X_j\right) \text{ and } K(\nu) = \mathcal{L}\left(e^{-\lambda T} \sum_{j=1}^{m} Y_j\right).$$

$$
\begin{aligned}
d_2(K(\mu),K(\nu))^2 &\leq \mathbb{E}\left(\left|\left(e^{-\lambda T}\sum_{j=1}^{m} X_j\right)-\left(e^{-\lambda T}\sum_{j=1}^{m} Y_j\right)\right|^2\right)\\
&= \mathbb{E}\left(\left|e^{-\lambda T}\sum_{j=1}^{m}(X_j-Y_j)\right|^2\right)\\
&= \mathbb{E}\left(\left|e^{-\lambda T}\right|^2\right)\cdot\mathbb{E}\left(\left|\sum_{j=1}^{m}(X_j-Y_j)\right|^2\right)\\
&= \mathbb{E}\left(\left|e^{-\lambda T}\right|^2\right)\cdot\left(\sum_{j=1}^{m}\mathbb{E}\left(|(X_j-Y_j)|^2\right)+\sum_{i\neq j}\mathbb{E}\left((X_i-Y_i)(\overline{X_j}-\overline{Y_j})\right)\right)\\
&= m\mathbb{E}\left(\left|e^{-2\lambda T}\right|\right)\cdot d_2(\mu,\nu)^2.
\end{aligned}
$$

Here we used that $|z|^2 = z\cdot\overline{z}$ for a complex number $z\in\mathbb{C}$, which implies that

$$
\begin{aligned}
\left|\sum_{j=1}^{m}(X_j-Y_j)\right|^2 &= \sum_{j=1}^{m}(X_j-Y_j)\cdot\sum_{j=1}^{m}(\overline{X_j}-\overline{Y_j})\\
&= \sum_{j=1}^{m}(X_j-Y_j)(\overline{X_j}-\overline{Y_j})+\sum_{i\neq j}(X_i-Y_i)(\overline{X_j}-\overline{Y_j})\\
&= \sum_{j=1}^{m}|X_j-Y_j|^2+\sum_{i\neq j}(X_i-Y_i)(\overline{X_j}-\overline{Y_j}).
\end{aligned}
$$

Since $\mu,\nu\in\mathcal{M}_2(C)$, we have $\mathbb{E}(X_j)=\mathbb{E}(Y_j)=C$, i.e. $\mathbb{E}(X_j-Y_j)=0$ and thus

$$
\sum_{j=1}^{m}\mathbb{E}\left(|(X_j-Y_j)|^2\right)+\sum_{i\neq j}\mathbb{E}\left((X_i-Y_i)(\overline{X_j}-\overline{Y_j})\right)=\sum_{j=1}^{m}\mathbb{E}\left(|(X_j-Y_j)|^2\right).
$$

By Equation 20.4 and since $2\Re(\lambda)>1$, we have $m\mathbb{E}\left(\left|e^{-2\lambda T}\right|\right)<1$, which finally implies that K is a contraction on $\mathcal{M}_2(C)$ and we are done. □

We give a second approach for the unicity of a solution of the distributional equation using Fourier analysis.

Remark 20.16 ([17])**.** Let $\mu\in\mathcal{M}_2(C)$ and let ϕ be its characteristic function. A random variable X with law μ satisfies $\mathbb{E}(|X|^2)<\infty$, so we can compute its moments, i.e. $\phi'(0)=\mathbb{E}(X)i=Ci$ and $\phi(0)=\mathbb{E}(e^{i\langle 0,X\rangle})=1$, so ϕ admits the expansion $\phi(t)=1+i\langle t,C\rangle+O(|t|^2)$ as $t\to 0$.

Definition 20.17 ([17])**.** Let $\mu, \nu \in \mathcal{M}_2(C)$ and let ϕ and ψ be their respective characteristic functions. We define the distance $d_2^*(\mu, \nu)$ on $\mathcal{M}_2(C)$ by

$$d_2^*(\mu, \nu) = \sup_{t \in \mathbb{C} \setminus \{0\}} \frac{|\phi(t) - \psi(t)|}{|t|^2}.$$

Remark 20.18 ([17])**.** We have $d_2^*(\mu, \nu) < \infty$, and it is actually a distance on $\mathcal{M}_2(C)$. Moreover, $(\mathcal{M}_2(C), d_2^*)$ is a complete metric space.

We state and prove the second theorem concerning the unicity of a solution of the distributional equation.

Theorem 20.19 ([17] Theorem 6.4)**.** *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (Equation 12.2) such that $\Re(\lambda) > \frac{1}{2}$, and let $C \in \mathbb{C}$. Then K is a contraction on the complete metric space $(\mathcal{M}_2(C), d_2^*)$, and the fixed point equation (Equation 20.7) has a unique solution Z in $\mathcal{M}_2(C)$.*

Proof. By the Banach fixed point theorem (see Theorem I.53), it is enough to show that K is a contraction on $\mathcal{M}_2(C)$ with respect to d_2^*. Let $\mu, \nu \in \mathcal{M}_2(C)$ and let ϕ and ψ be their respective characteristic functions. We have already shown in Construction 20.8 that the Fourier transform of $K(\mu)$ is $t \mapsto \mathbb{E}\left(\phi^m(te^{-\overline{\lambda}T})\right)$. For ν, we have a similar corresponding formula. It is well-known, that for $a, b \in \mathbb{R}$, the formula

$$(a^m - b^m) = (a - b)(a^{m-1} + a^{m-2}b + \cdots + ab^{m-2} + b^{m-1})$$

holds, and if $|a|, |b| \leq 1$, then $|a^i b^{m-1-i}| \leq 1$ for $i = 0, \ldots, m-1$. It follows that $|a^m - b^m| \leq m|a - b|$. Now, we know that $|\phi| \leq 1$ and $|\psi| \leq 1$, which implies

$$\mathbb{E}\left(\left|\phi^m(te^{-\overline{\lambda}T}) - \psi^m(te^{-\overline{\lambda}T})\right|\right) \leq m\mathbb{E}\left(\left|\phi(te^{-\overline{\lambda}T}) - \psi(te^{-\overline{\lambda}T})\right|\right).$$

From the definition of d_2^*, it follows that $|\phi(z) - \psi(z)| \leq |z|^2 d_2^*(\mu, \nu)$, and we apply it to the equation we had before with $z = te^{-\overline{\lambda}T}$,

$$\begin{aligned} m\mathbb{E}\left(\left|\phi(te^{-\overline{\lambda}T}) - \psi(te^{-\overline{\lambda}T})\right|\right) &\leq m\mathbb{E}\left(d_2^*(\mu, \nu)\left|te^{-\overline{\lambda}T}\right|^2\right) \\ &= d_2^*(\mu, \nu) m\mathbb{E}\left(\left|e^{-2\overline{\lambda}T}\right|\right)|t|^2 \\ &= d_2^*(\mu, \nu) m\mathbb{E}\left(\left|e^{-2\Re(\lambda)T}\right|\right)|t|^2. \end{aligned}$$

Now using that

$$\begin{aligned} d_2^*(K(\mu), K(\nu)) &= \sup_{t \in \mathbb{C} \setminus \{0\}} \frac{\left|\mathbb{E}\left(\phi^m(te^{-\overline{\lambda}T}) - \psi^m(te^{-\overline{\lambda}T})\right)\right|}{|t|^2} \\ &\leq \sup_{t \in \mathbb{C} \setminus \{0\}} \frac{\mathbb{E}\left(\left|\phi^m(te^{-\overline{\lambda}T}) - \psi^m(te^{-\overline{\lambda}T})\right|\right)}{|t|^2}, \end{aligned}$$

together with the above results, we get

$$d_2^*(K(\mu), K(\nu)) \leq \frac{d_2^*(\mu,\nu)|t|^2 m\mathbb{E}\left(\left|e^{-2\Re(\lambda)T}\right|\right)}{|t|^2} = d_2^*(\mu,\nu)m\mathbb{E}\left(\left|e^{-2\Re(\lambda)T}\right|\right).$$

We assumed $2\Re(\lambda) > 1$, so Equation 20.4 implies $\mathbb{E}\left(\left|e^{-2\Re(\lambda)T}\right|\right) < 1$, which proves that K is a contraction on $(\mathcal{M}_2(C), d_2^*)$. □

20.3 Support and density

We prove a result about the support and density of a solution of the distributional equation.

Definition 20.20 ([24])**.** The support of a random variable Z in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is the smallest closed set G in Ω such that $\mathbb{P}(Z \in \Omega \setminus G) = 0$.

We state the theorem about the support and density of a solution of the distributional equation.

Theorem 20.21 ([17] Theorem 7.1)**.** *Let $\lambda \in \mathbb{C}$ be a complex number such that $\lambda \neq 1$ and $\Re(\lambda) > 0$. Let the complex-valued random variable Z be a solution of the distributional equation (Equation 20.7):*

$$Z \stackrel{\mathcal{L}}{=} e^{-\lambda T}\left(Z^{(1)} + \cdots + Z^{(m)}\right),$$

with $\mathbb{E}(|Z|) < \infty$ and $\mathbb{E}(Z) \neq 0$. Then the following assertions hold.

(i) The support of Z is the whole complex plane $\mathbb{C}$.

(ii) As $|t| \to \infty$, $\mathbb{E}(e^{i\langle t, Z\rangle}) = O(|t|^{-a})$, for any $a \in \left(0, \frac{1}{\Re(\lambda)}\right)$.

(iii) The distribution of Z has a density with respect to the Lebesgue measure on $\mathbb{C}$.

Remark 20.22 ([17])**.** For $\lambda = 1$, the distributional equation (Equation 20.7) becomes

$$Z \stackrel{\mathcal{L}}{=} e^{-T}\left(Z^{(1)} + \cdots + Z^{(m)}\right). \tag{20.10}$$

We have already proven that there exists a unique solution in the space $\mathcal{M}_2(C)$ of probability measures on $\mathbb{C}$ with a given mean C. From the dislocation equations, a similar argument shows that

$$\xi = \lim_{t\to\infty} e^{-t} u_1(\mathscr{X}(t))$$

is a solution of this equation. As shown in Theorem 18.7, ξ is Gamma-distributed. Hence, the unique solution of Equation 20.10 in $\mathcal{M}_2(C)$ is $C\gamma$, where γ is Gamma$(1,1)$-distributed, and its support is the half line $C\mathbb{R}_{>0}$.

Corollary 20.23 ([17] Corollary 7.3)**.** *The distribution of $\mathscr{W}_1$ admits a density with respect to the Lebesgue measure on $\mathbb{C}$ and its support is the whole complex plane $\mathbb{C}$. Moreover, as $|t| \to \infty$, $\mathbb{E}(e^{i\langle t, \mathscr{W}_1\rangle}) = O(|t|^{-a})$, for any $a \in \left(0, \frac{1}{\Re(\lambda_2)}\right)$.*

Proof. The statement follows directly from Theorem 20.21 because of $\mathbb{E}(\mathscr{W}_1) = 1$. □

Notation 20.24 ([17])**.** We denote by Z a solution of the distributional equation (Equation 20.7) and by $\phi(t)$ its corresponding Fourier transform. Let $\mathcal{A} = e^{-\lambda T}$.

Definition 20.25 ([17])**.** We define the function $\psi(r) = \max_{|t|=r}(|\phi(t)|)$ for $r \in \mathbb{R}_{\geq 0}$.

Remark 20.26 ([17])**.** Let Z be a complex-valued random variable and $z \in \mathbb{C}$. Then $z \in \text{Supp}(Z)$ if and only if for all $\varepsilon > 0$, $\mathbb{P}(|Z - z| \leq \varepsilon) > 0$.

We state the following lemmas in order to prove Theorem 20.21.

Lemma 20.27 ([17] Lemma 7.4)**.** *Let $z \in \mathbb{C}$. If $z \in \text{Supp}(Z)$, then $D(0, |z|) \subseteq \text{Supp}(Z)$, where we denote by $D(0, |z|)$ the open disc with center 0 and radius $|z|$.*

Sketch of the proof. The proof of this lemma is done in two steps. First, we show that the implication

$$a \in \text{Supp}(\mathcal{A}) \text{ and } z \in \text{Supp}(z) \Longrightarrow maz \in \text{Supp}(Z)$$

holds. This is done with help of the alternative definition of support, given in the remark before. Then we can prove that maz lies in the support of $\mathcal{A}(Z^{(1)} + \cdots + Z^{(m)})$ which is equal to the support of Z by Equation 20.7.
In the second step, we notice that $\text{Supp}(T) = \mathbb{R}_{>0}$ (see Equation 20.2) and by the first step we have $me^{-\lambda t}z \in \text{Supp}(Z)$ for $t \in R_{>0}$ and $z \in \text{Supp}(Z)$. More precisely, we have

$$\{m^n e^{-\lambda t} z \mid n \in \mathbb{N}, t \in \mathbb{R}_{>0}\} \subseteq \text{Supp}(Z). \tag{20.11}$$

Since the support of a probability measure is a closed set, it is enough to show that $\{m^n e^{-\lambda t} \mid n \in \mathbb{N}, t \in \mathbb{R}_{>0}\}$ is everywhere dense on the unit disc to prove that $D(0, |z|) \subseteq \text{Supp}(Z)$. This is done using theory on transcendental and algebraic numbers. More details can be found in [17] in the proof of Lemma 7.4. □

Lemma 20.28 ([17] Lemma 7.5)**.** *For any $r > 0$, $\psi(r) < 1$.*

Sketch of the proof. We know $\psi(0) = 1$ and $\psi(r) \leq 1$ for any $r \geq 0$. By contradiction, we assume that $\psi(r_0) = 1$ for $r_0 > 0$. Using the mean of a complex random variable and its values on the unit disc, a contradiction to Lemma 20.27 is created. The details can be found in [17] in the proof of Lemma 7.5. □

Lemma 20.29 ([17] Lemma 7.7)**.** *We have $\lim_{r\to\infty} \psi(r) = 0$.*

Sketch of the proof. First of all, we show that $\limsup_{r\to\infty}\psi(r) = 0$ or 1 using Fatou's lemma (see Theorem I.38).
In the second step, we argue by contradiction. We assume $\limsup_r \psi(r) = 1$. Lemma 20.28 says that $\psi(1) < 1$. Let $\varepsilon \in (0, 1-\psi(1))$ and define
$r_1(\varepsilon) = \max\{r \in (0,1), \psi(r) = 1-\varepsilon\}$ and $r_2 = \min\{r > 1, \psi(r) = 1-\varepsilon\}$. Since $\psi(0) = 1$ and ψ is continuous, r_1 and r_2 are well-defined. We have $\psi(r_1(\varepsilon)) = \psi(r_2(\varepsilon)) = 1-\varepsilon$ and for $r \in [r_1(\varepsilon), r_2(\varepsilon)]$, $\psi(r) \le 1-\varepsilon$. It is then proven that $r_1(\varepsilon)$ goes to 0 as $\varepsilon \to 0$. This result, together with some inequalities involving $|\mathcal{A}|$ and independent copies of $|\mathcal{A}|$ as well as the Markov inequality lead to a contradiction. All the details can be found in [17] in the proof of Lemma 7.7. □

Lemma 20.30 ([17] Lemma 7.8)**.** *Let $\psi\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be a bounded function and let B be a positive random variable such that for some constants $p \in (0,1)$, $a > 0$, $C \ge 0$ and for all $r > 0$,*

$$\psi(r) \le p\mathbb{E}(\phi(Br)) + Cr^{-a}.$$

If $p\mathbb{E}(B^{-a}) < 1$, then $\psi(r) = O(r^{-a})$ as $r \to \infty$.

Sketch of the proof. We take $(B_i)_{1\le i\le n}$ to be independent copies of B. By induction, for all $n \in \mathbb{N}$ and all $r > 0$,

$$\psi(r) \le p^n\mathbb{E}(\psi(B_1\cdots B_n r)) + Cr^{-a}(1 + p\mathbb{E}(B^{-a}) + \cdots + (p\mathbb{E}(B^{-a}))^{n-1}).$$

Now, we let $n \to \infty$ and we see that for $r > 0$,

$$\psi(r) \le \frac{Cr^{-a}}{1 - p\mathbb{E}(B^{-a})}.$$

More details on the proof can be found in [17] in the proof of Lemma 7.8. □

Lemma 20.31 ([17] Lemma 7.9)**.** *For all $a \in \left(0, \frac{1}{\Re(\lambda)}\right)$, as $r \to \infty$, $\psi(r) = O(r^{-a})$.*

Sketch of the proof. We use the distributional equation (Equation 20.8) in terms of $\psi(r)$, Lemma 20.29, the Markov inequality, some property of the density of T and Lemma 20.30. □

We can finally prove the theorem about the support and density of a solution of the distributional equation.

Proof (of Theorem 20.21).

(i) First of all, by Lemma 20.27, whenever $z \in \mathbb{C}$ is in the support of Z, we have $D(0,|z|) \subseteq \mathrm{Supp}(Z)$.
Second, the support of Z is unbounded. We have from Equation 20.11 that as soon as $z \in \mathbb{C}\setminus\{0\}$ lies in the support of Z, then for any $t > 0$ and for any $n \in \mathbb{N}$, $m^n e^{-\lambda t} z$ is in the support of Z.

(ii) By Lemma 20.31, for all $a \in \left(0, \frac{1}{\Re(\lambda)}\right)$,

$$\phi(t) = O(t^{-a}) \text{ as } |t| \to \infty. \tag{20.12}$$

(iii) Since $\mathbb{E}(Z) \neq 0$, we obtain by Equation 20.8 that $m\mathbb{E}(e^{-\lambda T}) = 1$, hence $m\mathbb{E}(|e^{-\lambda T}|) = m\mathbb{E}(e^{-\Re(\lambda)T}) > 1$ as soon as $\Im(\lambda) \neq 0$. However, if $\Im(\lambda) = 0$, then $\lambda = 1$ by the equation $m\mathbb{E}(e^{-\lambda T}) = 1$. Thus, the hypotheses $\lambda \neq 1$ and $\mathbb{E}(Z) \neq 0$ imply that $\Im(\lambda) \neq 0$ and $\Re(\lambda) < 1$. As a consequence, Equation 20.12 holds for some $a > 1$, so the Fourier transform ϕ of Z is in L^2. Notice that the Fourier transform of a random variable is the same as the Fourier transform of its distribution. The distribution of Z is a finite measure and we just deduced that its Fourier transform is integrable. By Theorem I.55, the distribution of Z has a density with respect to the Lebesgue measure on $\mathbb{C}$.

□

20.4 Exponential moments

We state and prove a theorem about the exponential moments of a solution of the distributional equation. We introduce a martingale and its limit in order to prove the theorem.

Theorem 20.32 ([17] Theorem 8.1)**.** *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (Equation 12.2) with $\Re(\lambda) > \frac{1}{2}$ and let Z be a solution of Equation 20.7. Then, there exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$,*

$$\mathbb{E}(e^{\langle t, Z\rangle}) \leq e^{\Re(t)+C|t|^2} \tag{20.13}$$

and

$$\mathbb{E}(e^{|tZ|}) \leq 4e^{|t|+2C|t|^2}. \tag{20.14}$$

Remark 20.33 ([17])**.** We use the notation $\mathcal{A} = e^{-\lambda T}$. Since λ is a root of the characteristic polynomial (Equation 12.2), we have $m\mathbb{E}(\mathcal{A}) = 1$ and because $\Re(\lambda) > \frac{1}{2}$, we have further $m\mathbb{E}(|\mathcal{A}|^2) < 1$ (see Equation 20.4).

Definition 20.34 ([17])**.** We define $\mathbb{U}_{\{1,\ldots,m\}} \coloneqq \bigcup_{n\in\mathbb{N}} \{1, \ldots, m\}^n$. For $u = u_1 \cdots u_n \in \mathbb{U}_{\{1,\ldots,m\}}$, let $\mathcal{A}_u$ be independent copies of $\mathcal{A}$, indexed by all finite sequences of integers u. We set $Y_0 = 1$, $Y_1 = m\mathcal{A}$ and for $n \in \mathbb{N}$, $n \geq 2$,

$$Y_n = \sum_{u_1\cdots u_{n-1} \in \{1,\ldots,m\}^{n-1}} m\mathcal{A}\mathcal{A}_{u_1}\mathcal{A}_{u_1u_2}\cdots\mathcal{A}_{u_1\cdots u_{n-1}}.$$

Remark 20.35 ([17])**.** It follows directly from the definition, that

$$Y_{n+1} = \mathcal{A}\sum_{i=1}^{m} Y_{n,i}, \tag{20.15}$$

where $Y_{n,i}$, for $1 \leq i \leq m$, are independent of each other, independent of $\mathcal{A}$ and all have the same distribution as Y_n.

Proposition 20.36 ([17] Lemma 8.2)**.** *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (Equation 12.2) with $\Re(\lambda) > \frac{1}{2}$. Let $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be the filtration defined by $\mathcal{F}_n = \sigma(\{\mathcal{A}_{u_1\cdots u_{n-1}} \mid u_1 \cdots u_{n-1} \in \{1,\ldots,m\}^{n-1}\})$. Then Y_n is an $\mathcal{F}_n$-martingale with mean $\mathbb{E}(Y_n) = 1$.*
Moreover, as $n \to \infty$,

$$Y_n \to Y_\infty \text{ a.s. and in } L^2,$$

where Y_∞ is a complex-valued random variable with variance

$$\operatorname{Var}(Y_\infty) = \frac{\mathbb{E}(|\mathcal{A}|^2)m^2 - 1}{1 - m\mathbb{E}(|\mathcal{A}|^2)}.$$

Proof. We start by proving that Y_n is an $\mathcal{F}_n$-martingale. We have

$$\begin{aligned}
\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left(\sum_{u_1\cdots u_n\in\{1,\ldots,m\}^n} m\mathcal{A}\mathcal{A}_{u_1}\cdots\mathcal{A}_{u_1\cdots u_n} \mid \mathcal{F}_n\right)\\
&= \mathbb{E}\left(\sum_{u_n\in\{1,\ldots,m\}}\ \sum_{u_1\cdots u_{n-1}\in\{1,\ldots,m\}^{n-1}} m\mathcal{A}\mathcal{A}_{u_1}\cdots\mathcal{A}_{u_1\cdots u_{n-1}}\mathcal{A}_{u_1\cdots u_n} \mid \mathcal{F}_n\right)\\
&= \sum_{u_1\cdots u_{n-1}\in\{1,\ldots,m\}^{n-1}} m\mathcal{A}\mathcal{A}_{u_1}\cdots\mathcal{A}_{u_1\cdots u_{n-1}} \cdot \mathbb{E}\left(\sum_{u_n\in\{1,\ldots,m\}} \mathcal{A}_{u_1\cdots u_n} \mid \mathcal{F}_n\right)\\
&= Y_n \cdot \mathbb{E}(m\mathcal{A} \mid \mathcal{F}_n) = Y_n \cdot m\mathbb{E}(\mathcal{A}).
\end{aligned}$$

We assumed that λ is a root of the characteristic polynomial (Equation 12.2), so $m\mathbb{E}(\mathcal{A}) = 1$, and the equation required for Y_n to be a martingale is proven. Since $\Re(\lambda) > \frac{1}{2}$, we have $m\mathbb{E}(|\mathcal{A}|^2) < 1$ by Equation 20.4, which implies $Y_n \in L^2$. The $\mathcal{F}_n$-adaptation is trivial. Therefore, Y_n is an $\mathcal{F}_n$-martingale. Additionally, $E(Y_n) = \mathbb{E}(Y_0) = 1$.
Since Y_n lies in L^2, Doob's martingale convergence theorem (see Theorem I.40) implies that Y_n converges to a limit $Y_\infty \coloneqq \lim_{n\to\infty} Y_n$ almost surely and in L^2. We compute its variance, using Equation 20.15, independence and that $\mathbb{E}(Y_n) = 1$ for all $n \in \mathbb{N}_0$.

$$\begin{aligned}
\operatorname{Var}(Y_{n+1}) &= \mathbb{E}(|Y_{n+1} - \mathbb{E}(Y_{n+1})|^2)\\
&= \mathbb{E}(Y_{n+1}\overline{Y_{n+1}} - Y_{n+1}\overline{\mathbb{E}(Y_{n+1})} - \mathbb{E}(Y_{n+1})\overline{Y_{n+1}} + \mathbb{E}(Y_{n+1})\overline{\mathbb{E}(Y_{n+1})})\\
&= \mathbb{E}\left(\left(\mathcal{A}\sum_{i=1}^m Y_{n,i}\right)\left(\overline{\mathcal{A}}\sum_{i=1}^m \overline{Y_{n,i}}\right)\right) - 1\\
&= \mathbb{E}(|\mathcal{A}|^2)\cdot\mathbb{E}\left(\sum_{i=1}^m Y_{n,i}\overline{Y_{n,i}} + \sum_{i\neq j} Y_{n,i}\overline{Y_{n,j}}\right) - 1\\
&= \mathbb{E}(|\mathcal{A}|^2)\cdot(m\mathbb{E}(Y_n\overline{Y_n}) + (m^2-m)\mathbb{E}(Y_n)\mathbb{E}(\overline{Y_n})) - 1\\
&= \mathbb{E}(|\mathcal{A}|^2)\cdot m(\mathbb{E}(Y_n\overline{Y_n}) - \mathbb{E}(Y_n)\mathbb{E}(\overline{Y_n})) + m^2\mathbb{E}(|\mathcal{A}|^2) - 1
\end{aligned}$$

$$= \mathbb{E}(|\mathcal{A}|^2) \cdot m\mathrm{Var}(Y_n) + m^2\mathbb{E}(|\mathcal{A}|^2) - 1.$$

Taking $n \to \infty$ on both sides of the equation gives us

$$\mathrm{Var}(Y_\infty) = \mathbb{E}(|\mathcal{A}|^2) \cdot m\mathrm{Var}(Y_\infty) + m^2\mathbb{E}(|\mathcal{A}|^2) - 1,$$

and now rearranging finally leads to

$$\mathrm{Var}(Y_\infty) = \frac{\mathbb{E}(|\mathcal{A}|^2)m^2 - 1}{1 - m\mathbb{E}(|\mathcal{A}|^2)}.$$

□

Remark 20.37 ([17])**.** When we take the limit in Equation 20.15, we have the same equation as Equation 20.7, which gives us a new proof for the existence of a solution Z to this equation with $\mathbb{E}(Z) = 1$ and finite second moment whenever $\Re(\lambda) > \frac{1}{2}$. We already saw that the solution of this equation is unique. Therefore, Theorem 20.32 is proven as soon as it holds for Y_∞.

Lemma 20.38 ([17] Lemma 8.3)**.** *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (Equation 12.2) with $\Re(\lambda) > \frac{1}{2}$. Then, there exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$, we have*

$$E(e^{\langle t, Y_\infty \rangle}) \leq e^{\Re(t) + C|t|^2}. \tag{20.16}$$

Sketch of the proof. Let $\phi_n(t) \coloneqq \mathbb{E}(e^{\langle t, Y_n \rangle})$ for $t \in \mathbb{C}$, $n \in \mathbb{N}_0$, we have by Equation 20.15,

$$\phi_{n+1}(t) = \mathbb{E}(\phi_n^m(\overline{\mathcal{A}}t)),\ t \in \mathbb{C}.$$

Now, we prove by induction, that there exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $n \in \mathbb{N}_0$ and all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$, we have

$$\phi_n(t) \leq e^{\Re(t) + C|t|^2}. \tag{20.17}$$

This uses some analytical argument involving the proof of a local maximum using the Hessian matrix. Then, by taking $n \to \infty$ in Equation 20.17, we see that Inequality 20.16 holds by Fatou's lemma (see Theorem I.38). More details can be found in [17] in the proof of Lemma 8.3. □

We finally prove the theorem about the exponential moments of a solution of the distributional equation.

Proof (of Theorem 20.32). As mentioned in Remark 20.37, the uniqueness of a solution to Equation 20.7 implies $\mathcal{L}(Z) = \mathcal{L}(Y_\infty)$. By Lemma 20.38, there exist constants $C > 0$ and $\varepsilon > 0$ such that Equation 20.13 in Theorem 20.32 is proven. For the other equation, we notice that

$$e^{|tZ|} \leq e^{|t|(|\Re(Z)| + |\Im(Z)|)},$$

and $|t|(|\Re(Z)| + |\Im(Z)|)$ takes one of the four values $|t|(\pm\Re(Z) \pm \Im(Z))$. This gives us

$$e^{|tZ|} \leq e^{|t|(\Re(Z)+\Im(Z))} + e^{|t|(-\Re(Z)+\Im(Z))} + e^{|t|(\Re(Z)-\Im(Z))} + e^{|t|(-\Re(Z)-\Im(Z))}.$$

Moreover, we notice that

$$|t|(\pm\Re(Z) \pm \Im(Z)) = \Re((\pm 1 \pm i)|t|)\Re(Z) + \Im((\pm 1 \pm i)|t|)\Im(Z) = \langle(\pm 1 \pm i)|t|, Z\rangle.$$

We now take expectation in the inequality and apply the first inequality of the theorem. This leads to

$$\mathbb{E}(e^{\langle(1\pm i)|t|,Z\rangle}) \leq e^{\Re((1\pm i)|t|)+C|((1\pm i)|t|)|^2} = e^{|t|+2C|t|^2},$$

respectively

$$\mathbb{E}(e^{\langle(-1\pm i)|t|,Z\rangle}) \leq e^{-|t|+2C|t|^2} \leq e^{|t|+2C|t|^2},$$

and finally everything combined gives us

$$\mathbb{E}(e^{|tZ|}) \leq 4e^{|t|+2C|t|^2}.$$

□

21 Galton-Watson processes

We give an introduction to Galton-Watson processes and Galton-Watson trees, find a martingale and prove several statements concerning the extinction probability of a Galton-Watson process.

21.1 Definition of the process and associated trees

Notation 21.1 ([16])**.** We are working with trees defined as subsets of the set $\mathbb{U}_{\mathbb{N}_0} = \{\mathcal{E}\} \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}_0^n$.

Remark 21.2. [16] We recall that for a node u of a tree $\mathcal{T} \subseteq \mathbb{U}_{\mathbb{N}_0}$, we denote by $M_u(\mathcal{T})$, the number of children of u. In the following, we consider the case where the $M_u(\mathcal{T})$ are random variables.

Definition 21.3 ([39])**.** Let X be a random variable with values in $\mathbb{N}_0$ and $\mathbb{P}(X = k) = p_k$ for $p_k \in [0, 1]$, $k \in \mathbb{N}_0$. We define a stochastic process $(Z_n)_{n\in\mathbb{N}_0}$ recursively as follows. For $n = 0$, we set $Z_0 = 1$, meaning that we start with one individual at time $n = 0$. Then, at time $n = 1$, this individual has k children with probability p_k, we let Z_1 be the number of these children. Each of these children (provided that there exist any) has also children at time $n = 2$ with the same distribution $(p_k)_{k\in\mathbb{N}_0}$, independent of the others and of its parent. We write Z_2 for the number of children of the second generation. We continue this procedure recursively until there exist no more children (possibly an infinitely long amount of time).
Said differently, let $(X_i^{(n)})_{n,i\in\mathbb{N}}$ be independent copies of X. Then the sizes of the generations of this branching process are given inductively by

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^{(n+1)},$$

for $n \in \mathbb{N}_0$. We call this process a Galton-Watson branching process or just Galton-Watson process. The random variable X is called the offspring random variable or offspring distribution.

Remark 21.4 ([39])**.** In a Galton-Watson process, the random variable Z_{n+1} depends only on the random variable Z_n for all $n \in \mathbb{N}_0$, so this process is a discrete time Markov chain on $\mathbb{N}_0$.

Definition 21.5 ([16])**.** Let X be a random variable with values in $\mathbb{N}_0$ and $\mathbb{P}(X = k) = p_k$ for $p_k \in [0, 1]$, $k \in \mathbb{N}_0$. Let $\mathcal{T} \in \mathbb{U}_{\mathbb{N}_0}$ be an ordered tree such that the number of children $M_u(\mathcal{T})$ of a node u is a random variable which is independent and has the same distribution as X. A tree constructed under the distribution $(p_k)_{k\in\mathbb{N}_0}$ is called a Galton-Watson tree with reproduction distribution $(p_k)_{k\in\mathbb{N}_0}$.
Let $z_n(\mathcal{T}) = \{u \in \mathcal{T} \mid |u| = n\}$ be the n-th generation of the tree $\mathcal{T}$. Then we set $Z_n(\mathcal{T}) = \mathrm{card}(z_n(\mathcal{T}))$, which is the number of nodes in the n-th generation. Then $(Z_n)_{n\in\mathbb{N}_0}$ is a Galton-Watson process.
In conclusion, we can see a Galton-Watson process as a growing tree process.

21.2 The Galton-Watson martingale

Definition 21.6 ([39])**.** We define by $m \coloneqq \mathbb{E}(X) = \mathbb{E}(Z_1) = \sum_{k\in\mathbb{N}_0} kp_k$ the expectation of the offspring. We call m the expectation of the branching process. We say that the process is subcritical if $m < 1$, critical if $m = 1$ and supercritical if $m > 1$.

Theorem 21.7 ([39] Proposition 5.5)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. Assume that $0 < m < \infty$. Then the process*

$$W_n \coloneqq \frac{Z_n}{m^n}$$

is an $\mathcal{F}_n$-martingale, where the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ is defined by $\mathcal{F}_n = \sigma(Z_n, \ldots, Z_0)$. Moreover, we have $\mathbb{E}(W_n) = 1$ and hence, $\mathbb{E}(Z_n) = m^n$.

Proof. Let $n \in \mathbb{N}_0$. Since $(Z_n)_{n\in\mathbb{N}_0}$ is a Markov chain, we have

$$\begin{aligned}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(Z_{n+1} \mid Z_n) &= \mathbb{E}\left(\sum_{i=1}^{Z_n} X_i^{(n+1)} \mid Z_n\right)\\ &= \mathbb{E}(X) \cdot \mathbb{E}(Z_n \mid Z_n) = mZ_n.\end{aligned}$$

The $\mathcal{F}_n$-adaptation is clear. We compute the expectation of Z_{n+1},

$$\mathbb{E}(Z_{n+1}) = \mathbb{E}(\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n)) = m\mathbb{E}(Z_n).$$

Inductively, we get $\mathbb{E}(Z_{n+1}) = m^{n+1}$, resp. $\mathbb{E}(Z_n) = m^n$. Moreover, Z_n is non-negative, therefore

$$\mathbb{E}\left(\left|\frac{Z_n}{m^n}\right|\right) = \mathbb{E}\left(\frac{Z_n}{m^n}\right) = 1.$$

By Theorem 4.1, W_n is an $\mathcal{F}_n$-martingale with expectation 1. □

Definition 21.8 ([39])**.** Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. The martingale $(W_n)_{n\in\mathbb{N}_0}$ is called the Galton-Watson martingale.

Theorem 21.9 ([39])**.** *Let $(W_n)_{n\in\mathbb{N}_0}$ be the Galton-Watson martingale. Then $(W_n)_{n\in\mathbb{N}_0}$ converges almost surely to a limit denoted by $W_\infty \coloneqq \lim_{n\to\infty} W_n$.*

Proof. We know that W_n is non-negative martingale. It is a well-known result in probability theory that the limit of a non-negative martingale exists almost surely (see Theorem I.39). □

21.3 The extinction probability

Definition 21.10 ([39])**.** Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process. The event

$$\{\exists n \in \mathbb{N}_0 \colon Z_n = 0\}$$

is called extinction. The extinction probability is $q \coloneqq \mathbb{P}(\{\exists n \in \mathbb{N}_0 \colon Z_n = 0\})$. Extinction is the same event as $\{Z_n \to 0 \text{ for } n \to \infty\}$.

Remark 21.11 ([2])**.** If we have $p_1 = 1$, then $Z_1 = X, Z_2 = X, \ldots, Z_n = X$, so $Z_n = Z_1$ for all $n \in \mathbb{N}$, i.e. this case is trivial. The same goes for $p_0 = 1$, because we have $Z_n = 0$ for all $n \in \mathbb{N}$. Therefore, we exclude these cases and assume that $p_0 + p_1 < 1$.

Definition 21.12 ([39])**.** Let X be a random variable with values in $\mathbb{N}_0$ and $\mathbb{P}(X = k) = p_k$ for $p_k \in [0,1]$, $k \in \mathbb{N}_0$. The probability generating function of X is

$$G_X \colon [0,1] \to [0,1],\ G_X(s) = \mathbb{E}(s^X) = \sum_{k \in \mathbb{N}_0} s^k p_k.$$

Proposition 21.13 ([39] Proposition 5.2)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X. Then, we have for $n \in \mathbb{N}_0$,*

$$G_{Z_n}(s) = G_{Z_1}^{\circ n}(s) = \underbrace{G_{Z_1} \circ \cdots \circ G_{Z_1}(s)}_{n \text{ times}},$$

where $G_{Z_1} = G_X$.

Proof. We have $G_{Z_1}(s) = \mathbb{E}(s^X) = G_X(s)$ by definition of Z_1. Using some properties of the conditional expectation, we get

$$\begin{aligned}
G_{Z_n}(s) = \mathbb{E}(s^{Z_n}) &= \mathbb{E}\left(\mathbb{E}\left(s^{\sum_{i=1}^{Z_{n-1}} X_i^{(n)}} \mid Z_{n-1}\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=1}^{Z_{n-1}} s^{X_i^{(n)}} \mid Z_{n-1}\right)\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{Z_{n-1}} \mathbb{E}(s^{X_i^{(n)}})\right) \\
&= \mathbb{E}(\mathbb{E}(s^X)^{Z_{n-1}}) = \mathbb{E}(G_X(s)^{Z_{n-1}}) = G_{Z_{n-1}} \circ G_X(s).
\end{aligned}$$

We iterate this procedure $n-1$ times, which gives us $G_{Z_n}(s) = G_X^{\circ n}(s)$. □

Proposition 21.14 ([39] Corollary 5.3)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. The extinction probability satisfies*

$$q = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} G_X^{\circ n}(0)$$

Proof. We notice that if $Z_n = 0$, then $\mathbb{Z}_{n+1} = 0$ for all $n \in \mathbb{N}_0$, implying that $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$. Therefore,

$$\begin{aligned}
q = \mathbb{P}(\{\exists n \in \mathbb{N}_0 \colon Z_n = 0\}) &= \mathbb{P}\left(\bigcup_{n\in\mathbb{N}_0} \{Z_n = 0\}\right) \\
&= \lim_{n\to\infty} \mathbb{P}(Z_n = 0).
\end{aligned}$$

Now, since $G_X^{\circ n}(0) = G_{Z_n}(0) = \mathbb{P}(Z_n = 0)$, we finally have

$$q = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} G_X^{\circ n}(0).$$

□

Remark 21.15 ([4])**.** Since $Z_n = 0$ implies $\mathbb{Z}_{n+1} = 0$ for all $n \in \mathbb{N}_0$, it follows that 0 is an absorbing state, i.e. $\mathbb{P}(Z_{n+1} = k) = 0$ for all $k \in \mathbb{N}$ if $Z_n = 0$.

Proposition 21.16 ([4] Proposition 1.1)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. Let $p_1 \neq 1$. Conditioned on non-extinction, we have that $Z_n \to \infty$ almost surely as $n \to \infty$.*

Proof. In terms of Markov chains, this means that all states are transient except 0, which is absorbing as we already noticed. This means that we have to show

$$\mathbb{P}(Z_n = k, \text{ for some } n \geq 2 \mid Z_1 = k) < 1$$

for $k \in \mathbb{N}$. So let $k \in \mathbb{N}$.
We first assume that $p_0 = 0$. Then

$$\mathbb{P}(Z_n = k, \text{ for some } n \geq 2 \mid Z_1 = k) = \mathbb{P}(Z_2 = k \mid Z_1 = k) = p_1^k < 1,$$

since in this case $Z_1 \leq Z_2 \leq Z_3 \leq \ldots$.
In the second case, we assume $p_0 > 0$. Then

$$\mathbb{P}(Z_n = k, \text{ for some } n \geq 2 \mid Z_1 = k) \leq \mathbb{P}(Z_2 > 0 \mid Z_1 = k) = 1 - p_0^k < 1.$$

Therefore, $Z_n \to \infty$ as $n \to \infty$. □

Theorem 21.17 ([4] Proposition 1.2)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. Let $p_1 \neq 1$. Then, $m \leq 1$ if and only if $q = 1$.*

Proof. We have the following line of equivalences:

$$\begin{aligned} q < 1 &\iff \mathbb{P}(Z_n \to \infty) > 0 \\ &\iff m^n = \mathbb{E}(Z_n) \to \infty \\ &\iff m > 1. \end{aligned}$$

This is the same as $m \leq 1 \iff q = 1$. □

Lemma 21.18 ([6] Lemma I.3.1)**.** *Let X be an offspring random variable. Then $G_X(s)$ satisfies the following properties:*

(i) $G_X(1) = 1$ and $G_X(0) = p_0$;

(ii) $G_X(s)$ is strictly convex and monotonically increasing on $[0, 1]$;

Proof.

(i) $G_X(1) = 1$ and $G_X(0) = p_0$. It follows directly from the definition that $G_X(1) = \sum_{k\in\mathbb{N}_0} p_k = 1$ and $G_X(0) = \mathbb{P}(X = 0) = p_0$ since $0^0 = 1$ by convention.

(ii) We notice that $G'_X(s) = \sum_{k=1}^{\infty} k s^{k-1} p_k > 0$ and $G''_X(s) = \sum_{k=2}^{\infty}(k-1)k s^{k-2} p_k > 0$, which implies the monotonicity and the convexity.

□

Theorem 21.19 ([2] Theorem 1.7)**.** *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. The extinction probability q is the smallest fixed point of G_X in $[0,1]$, i.e. the smallest solution of the equation $G_X(s) = s$ in this interval.*
If $0 < p_0 \leq p_0 + p_1 < 1$, then G_X has exactly one fixed point on $[0,1)$ if $m > 1$ and none if $m \leq 1$. Therefore, $p_0 < q < 1$ if $m > 1$ and $q = 1$ if $m \leq 1$.

Proof. We have that $q = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} G_X^{\circ n}(0)$. G_X is continuous on $[0,1]$, thus

$$G_X(q) = \lim_{n\to\infty} G_X \circ G_X^{\circ n}(0) = \lim_{n\to\infty} G_X^{\circ(n+1)}(0) = q.$$

This means that q is a fixed point of G_X. Now let α be some arbitrary fixed point of G_X. G_X as well as $G_X^{\circ(n+1)}$ are monotonically increasing on $[0,1]$, hence

$$\alpha = G_X(\alpha) = G_X^{\circ(n+1)}(\alpha) \geq G_X^{\circ(n+1)}(0)$$

for all $n \in \mathbb{N}$, which implies $\alpha \geq \lim_{n\to\infty} G_X^{\circ(n+1)}(0) = q$. Therefore, q is the smallest fixed point of G_X on $[0,1]$.
We assume now that $0 < p_0 \leq p_0 + p_1 < 1$, so G_X is strictly increasing and strictly convex on $[0,1]$. We assume further that $m \leq 1$ and we let $F_X(s) = G_X(s) - s$. Then $F_X(0) = p_0 > 0$ and $F_X(1) = 0$. Since $G'_X(s) < G'_X(1) = m$ for all $s \in [0,1)$, we have $F'_X(s) = G'_X(s) - 1 < 0$ for all $s \in [0,1)$ since $m \leq 1$. This means that $F_X(s) \neq 0$ for all $s \in [0,1)$, i.e. $G_X(s)$ has no fixed point on $[0,1)$.
We assume that $m > 1$, then $G_X(s)$ grows faster than s in a left neighbourhood of 1, so $F'_X(1) = m - 1 > 0$ and F'_X is continuous. In a right neighbourhood of 0, we have $G_X(s) > s$, hence $F_X(0) > 0$ and F_X is continuous. Using the intermediate value theorem, there exists at least one fixed point s_1 of G_X in $(0,1)$. If there would be a second fixed point s_2, without loss of generality $s_1 < s_2 < 1$, we would have $F_X(s_1) = F_X(s_2) = F_X(1) = 0$ and hence there exist $a, b \in (0,1)$ with $s_1 < a < s_2 < b < 1$, with $F'_X(a) = F'_X(b) = 0$, or equivalently, $G'_X(a) = G'_X(b)$ (Rolle's theorem), which is a contradiction to the strict convexity of G_X. Thus q is the unique fixed point in $(0,1)$. Furthermore it satisfies $q > p_0$ since

$$q \geq \mathbb{P}(Z_1 = 0) + \mathbb{P}(Z_1 > 0, Z_2 = 0) \geq p_0 + p_k p_0^k$$

for any $k \in \mathbb{N}$ such that $p_k > 0$. This completes the proof. □

Remark 21.20 ([16]). We have seen that the function $G_X(s)$ always has the trivial fixed point $s = 1$ and if $m > 1$, it has the additional fixed point $s = q$. If $p_0 = 0$, then $q = 0$ and G_X has the second fixed point $s = q = 0$. So the result from Theorem 21.19 can be extended to $p_0 = 0$.

Definition 21.21 ([39]). A property of trees is called inherited, if

(i) every finite tree has this property;

(ii) whenever a tree has this property, then all the subtrees of the children of the root have this property.

Proposition 21.22 ([16] Lemme 5.6). *Every inherited property A in a Galton-Watson tree $\mathcal{T}$ has probability either q or 1. In particular, conditioned on non-extinction, the inherited property has probability either 0 or 1.*

Proof. Let A be an inherited property and let $\mathcal{T}$ be a Galton-Watson tree. We write $\mathcal{T} \in A$ if the tree $\mathcal{T}$ has has the inherited property. We assume that the root of the tree $\mathcal{T}$ has Z_1 children and we write $\mathcal{T}^{(i)}$, $i = 0, \ldots, Z_1 - 1$ for the subtrees of the root, i.e. the trees whose roots start at the children of the root. Since the subtrees are i.i.d. given Z_1,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{E}(\mathbb{P}(\mathcal{T} \in A \mid Z_1)) \\ &\le \mathbb{E}(\mathbb{P}(\mathcal{T}^{(0)} \in A, \mathcal{T}^{(1)} \in A, \ldots, \mathcal{T}^{(Z_1-1)} \in A \mid Z_1)) \\ &= \mathbb{E}(\mathbb{P}(A)^{Z_1}) = G_X(\mathbb{P}(A)).\end{aligned}$$

Every Galton-Watson tree, where the corresponding Galton-Watson process goes extinct, is finite. Therefore, $\mathbb{P}(A) \ge q$. Since the probability generating function $G_X(s)$ of the Galton-Watson process is convex and increasing, we deduce that $\mathbb{P}(A) \in \{q, 1\}$ as these are the only possible points on $[0, 1]$ satisfying $G_X(s) \ge s$ and $s \ge q$.
It follows immediately that if the process does not go extinct, we have $q = 0$ and thus $\mathbb{P}(A) \in \{0, 1\}$. □

Theorem 21.23 ([16] Proposition 5.7). *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. Let W_∞ be the Galton-Watson martingale limit. Then $\mathbb{P}(W_\infty = 0) \in \{q, 1\}$.*

Proof. We notice that $\{W_\infty = 0\}$ is an inherited property, so we apply Proposition 21.22 and get $\mathbb{P}(W_\infty = 0) \in \{q, 1\}$. □

Theorem 21.24 (Kesten-Stigum theorem, [1] Satz 6.2). *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process with offspring random variable X and mean m. Let W_∞ be the Galton-Watson martingale limit. We assume that $1 < m < \infty$. Then the following are equivalent:*

(i) $\mathbb{P}(W_\infty = 0) = q$;

(ii) $\mathbb{E}(W_\infty) = 1$;

(iii) $W_n \to W_\infty$ *as* $n \to \infty$ *in* L^1*;*

(iv) $(W_n)_{n\in\mathbb{N}_0}$ *is uniformly integrable (see Definition I.42);*

(v) $\mathbb{E}(\sup_{n\in\mathbb{N}_0}(W_n)) < \infty$*.*

Proof. Some of these equivalences are well-known results in probability theory. More precisely, the equivalence between (ii) and (iii) (see Theorem I.41), between (iii) and (iv) (see Theorem I.43) as well as the implication (v) $\Longrightarrow$ (vi) (see Proposition I.44) follow directly from more general results in probability theory. We prove the remaining implications.
Let us assume that $\mathbb{E}(W_\infty) = 1$. By Theorem 21.23, we have $\mathbb{P}(W_\infty = 0) \in \{q, 1\}$, but having $\mathbb{P}(W_\infty = 0) = 1$ would be a contradiction to $\mathbb{E}(W_\infty) = 1$. Therefore it must hold that $\mathbb{P}(W_\infty = 0) = q$.
Let us now assume that $\mathbb{P}(W_\infty = 0) = q$. Since $m > 1$, this is equivalent to $\mathbb{P}(W_\infty > 0) > 0$. We want to prove $\mathbb{E}(\sup_{n\in\mathbb{N}_0}(W_n)) < \infty$. Let $W_\infty^{(1)}, W_\infty^{(2)}, \dots$ be independent copies of W_∞. Since $\mathbb{P}(W_\infty > 0) > 0$, we have $\mathbb{E}(W_\infty)$. Now let $a \in (0, \mathbb{E}(W_\infty)$. By the strong law of large numbers (the convergence holds almost surely), there exist $N \in \mathbb{N}$ and $b > 0$ such that

$$\mathbb{P}\left(\sum_{i=1}^{tm^n} W_\infty^{(i)} > atm^n\right) \geq b$$

for all $t \geq 1$ and $n \geq N$. Then, for all $t \geq 1$, we have

$$\begin{aligned}
\mathbb{P}(W_\infty > at) &\geq \mathbb{P}\left(W_\infty > at, \sup_{n\geq N}(W_n) > t\right) \\
&= \sum_{n\geq N} \mathbb{P}\left(W_\infty > at, W_n > t, \sup_{N\leq k\leq n}(W_k) \leq t\right) \\
&= \sum_{n\geq N} \sum_{j\geq tm^n} \mathbb{P}\left(\lim_{k\to\infty} m^{-(n+k)} Z_{n+k} > at \mid Z_n = j\right) \\
&\quad \cdot \mathbb{P}\left(Z_n = j, \sup_{N\leq k\leq n}(W_k) \leq t\right) \\
&= \sum_{n\geq N} \sum_{j\geq tm^n} \mathbb{P}\left(\lim_{k\to\infty} m^{-k} Z_k > atm^n \mid Z_n = j\right) \\
&\quad \cdot \mathbb{P}\left(Z_n = j, \sup_{N\leq k\leq n}(W_k) \leq t\right) \\
&\geq \sum_{n\geq N} \sum_{j\geq tm^n} \mathbb{P}\left(\sum_{i=1}^{tm^n} W_\infty^{(i)} > atm^n\right) \mathbb{P}\left(Z_n = j, \sup_{N\leq k\leq n}(W_k) \leq t\right)
\end{aligned}$$

$$\geq b \cdot \mathbb{P}\left(\sup_{n\geq N}(W_n) > t\right).$$

We can use this result in the following inequalities. Since $E(W_n) = 1$, we get

$$\begin{aligned}\mathbb{E}\left(\sup_{n\geq 0}(W_n)\right) &\leq N + \mathbb{E}\left(\sup_{n\geq N}(W_n)\right)\\ &= N + \int_0^\infty \mathbb{P}\left(\sup_{n\geq N}(W_n) > t\right) dt\\ &\leq (N+1) + \frac{1}{b}\int_1^\infty \mathbb{P}(W_\infty > at)\, dt \leq (N+1) + \frac{\mathbb{E}(W_\infty)}{ab} < \infty,\end{aligned}$$

which completes the proof. □

22 Conclusion

This thesis gave an introduction to trees, especially oriented to random binary search trees and random m-ary search trees. Binary search trees were considered under the uniform permutation model and we could find a martingale. We introduced the notion of one-dimensional branching processes in discrete and continuous time using point processes and found a martingale in both cases. We talked about the bisection problem and how it is a particular case of a discrete time branching random walk. The Yule tree, which can be seen as a continuous time version of the binary search tree, was developed as a particular case of a continuous time branching random walk. The asymptotic behaviour of the binary search tree martingale for positive real numbers was elaborated and we saw that its limit is either zero or positive on a specific interval. We established the embedding principle which allowed us to consider the binary search tree and Yule tree in the same probability space. This allowed us to find a martingale connection and a limit connection, which gave us the opportunity to apply the results about the limit of the binary search tree martingale to the limit of the Yule tree martingale. Said differently, the limit of the Yule tree martingale is either zero or positive on the same specific interval. Moreover, we could prove a convergence result about the profile of binary search trees using independence properties of the Yule tree. Such properties allowed us also to find a bisection martingale in binary search trees and to show that its limit is equal to the binary search tree martingale limit on the same specific interval.

Pólya urns and m-ary search trees were introduced, we found discrete time martingales for both processes and a relation between these processes. We defined Pólya processes and developed their asymptotic behaviour when they are large. By proving that large Pólya urns, and thus large m-ary search trees, can be generalized to large Pólya processes, we received their asymptotic behaviour. We then introduced the continuous time versions of these processes, the continuous time Pólya urn process and the continuous time multitype branching process with $m-1$ types. An embedding principle was elaborated, we found martingales and used the discrete time versions to establish the asymptotic behaviour for large processes. With the continuous time versions we developed the asymptotic behaviour of the small discrete time versions.

From the multitype continuous time branching process with $m-1$ types we received a distributional equation for a limit random variable and we characterized its distribution.

The discrete time Galton-Watson process and its corresponding tree were introduced. We elaborated the extinction probability, found a martingale and its limit. Then we established certain relations between this limit, the martingale and the extinction probability.

Outlook

We have seen different kinds of trees and branching processes in this thesis. However, there is more to explore. We limited the study of the binary search tree martingale to positive real numbers. The study can be extended to complex numbers under the condition that the martingale stays well-defined. Additionally, we only considered the case of the uniform permutation model for binary search trees and were only interested in the distribution of the leaves in a tree. Other trees and other distributions on trees can be studied, as well as other properties of trees such as path lengths. The bisection problem is part of random fragmentation processes which are a more general concept. Furthermore, we never considered the case of extinction in Pólya urns and we only concentrated on the two-dimensional case. Similarly to continuous time multitype branching processes with $m-1$ types, distributional equations can be established to characterize a limit random variable in continuous time Pólya urns. Galton-Watson processes can be extended to multi-dimensional processes and continuous time versions can be elaborated. Last but not least, these processes have real world applications that can be studied.

After this thesis, the reader should now have a solid background in the introduced topics and be able to study other papers of similar subjects.

Appendix

I.1 A graph-theoretical introduction to trees

We give an additional introduction to trees in terms of graph theory.

Definition I.1 ([40])**.** Let V and E be two sets, V is non-empty. We call the set $G = (V, E)$ a simple undirected graph, where V is the set of vertices and E is the set of edges. Every edge $e \in E$ is an unordered pair of distinct vertices, i.e. $e = \{x, y\}$ with $x, y \in V$. These two vertices x and y are the endpoints of the edge e. Sometimes we also say that e is incident with x and y or e joins x and y.
When V is finite, then G is said to be finite.

Notation I.2 ([40])**.** An edge which is incident with the vertices x and y can also be denoted by xy.

Remark I.3 ([40])**.** At the moment, the edges have no orientation, which means that xy and yx refer to the same edge.

Definition I.4 ([40])**.** A graph G is said to be empty if it has no edges, i.e. if $E = \emptyset$.

Definition I.5 ([40])**.** We call the number of edges incident with a vertex $x \in V$ the degree of x and denote it by $d(x)$.

Remark I.6 ([40])**.** Let $G = (V, E)$ be an non-empty simple undirected graph. Each edge joins two vertices, so the sum of degrees of all edges is $\sum_{x \in V} d(x) = 2 \cdot |E|$.

Definition I.7 ([40])**.** Let V and E be two sets, V is non-empty. We say that $G = (V, E)$ is a simple directed graph (sometimes also called a digraph), with V the set of vertices and E the set of edges. An edge $e \in E$ is an ordered pair of distinct vertices. Using the notation from before, xy and yx are two different edges, where xy is directed from x to y. Said differently, xy starts at x and ends at y.
When V is finite, we call G finite.

Definition I.8 ([40])**.** Let $G = (V, E)$ be a digraph. We call the indegree $d_{\text{in}}(x)$ of a vertex $x \in V$ the number of edges starting at x and $d_{\text{out}}(x)$ the outdegree of x, which corresponds to the number of edges ending at x.

Remark I.9 ([40])**.** It follows that $\sum_{x \in V} d_{\text{in}}(x) = \sum_{x \in V} d_{\text{out}}(x) = |E|$.

Remark I.10 ([40])**.** From now on, we call *graph* a finite simple undirected graph and *digraph* a finite simple directed graph.

Definition I.11 ([40])**.** Let $G = (V, E)$ be a graph. A subgraph of G is a graph $H = (V', E')$ if $V' \subseteq V$ and $E' \subseteq E$.

Notation I.12 ([40])**.** We write $H \subseteq G$ for a subgraph H of G.

Definition I.13 ([40])**.** If $H \subseteq G$ but $H \neq G$, then we say that H is a proper subgraph of G.

Definition I.14 ([40])**.** Let $G = (V, E)$ be a graph and $X \subseteq V$ be a non-empty subset. We denote the graph induced by X by $G[X]$ and define it as the subgraph (X, E') with $E' = \{e \in E \mid e = xy,\ x, y \in X\}$, i.e. E' contains all the edges in E whose corresponding vertices lie in X.
In the same manner, the subgraph obtained by deleting a subset of vertices X is $G[V \setminus X]$ and is denoted by $G \setminus X$ or $G \setminus x$ for a singleton $X = \{x\}$.

Remark I.15 ([40])**.** Removing a vertex in a graph implies removing all the edges incident with this vertex. But removing an edge in a graph does not change the set of vertices.

Definition I.16 ([40])**.** For a graph $G = (V, E)$ and a subset $E' \subseteq E$, we define the graph $G \setminus E'$ as the subgraph $(V, E \setminus E')$, the subgraph obtained by removing the edges lying in E'. We write $E \setminus e$ for $E \setminus \{e\}$.

Definition I.17 ([40])**.** Let $x_0, x_1, \ldots, x_k$ be a sequence of distinct vertices in a simple undirected (resp. directed) graph $G = (V, E)$. Such a sequence is called a path (resp. directed path) if $x_i x_{i+1}$ is an edge (resp. a directed edge) for $i = 0, \ldots, k-1$. We say that the length of the path is k, the number of edges in it. The path connects x_0 to x_k.

Definition I.18 ([40])**.** We say that a simple undirected (resp. directed) graph is connected (resp. strongly connected) if for every pair of vertices x and y there exists a path (resp. directed path) connecting x to y.
We call a subgraph H of G a component if it is connected and not a proper subgraph of another connected subgraph of G.

Definition I.19 ([40])**.** A vertex which is a component by itself is said to be isolated.

Definition I.20 ([40])**.** Let $x_0, x_1, \ldots, x_k$ be a path (resp. directed path) from x_0 to x_k and assume that $x_k x_0$ is an edge (resp. a directed edge). Then we call $x_0, x_1, \ldots, x_k$ a circuit (resp. a directed circuit).

Definition I.21 ([40])**.** An undirected tree is a connected circuit-free graph. We just write tree for an undirected tree.

Theorem I.22 ([40] Theorem 1.1)**.** *In any tree, there exist exactly one path connecting any pair of vertices.*

Proof. We argue by contradiction. Let $\mathcal{T} = (V, E)$ be a tree and let x and y be two vertices of $\mathcal{T}$. We know that $\mathcal{T}$ is connected, so there must be at least one path from x to y. We now assume, that there is more than one path from x to y. Let P and Q be two such paths. Both of them start at x. At some vertex, the two paths separate and somewhere they come together again, this happens possibly several times. Let z be the first common vertex between P and Q after which the two paths separate, so the vertex after z on P is different from the vertex after z on Q. Let w be the first common vertex after z. Now, we consider only the parts of paths P and Q connecting z and w, we denote these parts by P' and Q' respectively. These parts both start at z and end at w and the only common vertices are z and w. But if we take path P' together with the reverse path of Q', we get a circuit in $\mathcal{T}$. This is a contradiction since $\mathcal{T}$ is circuit-free. □

Theorem I.23 ([40] Theorem 1.2)**.** *In a tree with n vertices, there are exactly $n-1$ edges.*

Proof. Let $\mathcal{T} = (V, E)$ be a tree with n vertices. First, we prove that if $e = xy$ is an edge of $\mathcal{T}$, then $\mathcal{T} \setminus e$ is disconnected. The two ends x and y of the edge e are connected by a unique path according to the theorem before. Because the edge xy connects x to y, there is no path from x to y other than xy. Therefore, there is no path xy in $\mathcal{T} \setminus e$, which implies that $\mathcal{T} \setminus e$ is disconnected. We can classify all the paths from x to all other vertices of $\mathcal{T}$ into two types: paths that use e and paths that do not use e. All the vertices which are not connected to x using e remain connected to x in $\mathcal{T} \setminus e$ and all the vertices which are connected to x with a path using e have to be connected to y, are not connected to x in $\mathcal{T} \setminus e$ and remain connected to y in $\mathcal{T} \setminus e$. Therefore, $\mathcal{T} \setminus e$ is a graph consisting of two trees, having the same number of vertices as $\mathcal{T}$ and having one edge less than $\mathcal{T}$.
We proceed by induction on n. The only tree with one vertex is the empty graph of one vertex and thus the statement holds for $n = 1$. Now, we assume that the statement is true for every tree with less that n vertices, $n \in \mathbb{N}$. Let the two trees of $\mathcal{T} \setminus e$ be $\mathcal{T}_1 = (V_1, E_1)$ and $\mathcal{T}_2 = (V_2, E_2)$. We have $|V_1| < n$ and $|V_2| < n$. By induction, $|E_1| = |V_1| - 1$ and $|E_2| = |V_2| - 1$, which gives us $|E_1| + |E_2| = |V| - 2$. Since $|E| = |E_1| + |E_2| + 1$, we get that $|E| = n - 1$, this completes the induction. □

Theorem I.24 ([40] Theorem 1.3)**.** *Let G be a graph with n vertices. The following statements are equivalent.*

(i) G is a tree.

(ii) G is circuit-free and has $n-1$ edges.

(iii) G is connected and has $n-1$ edges.

Proof. (i) implies both (ii) and (iii) by definition and the theorem before.
Suppose that G satisfies (ii) and we assume by contradiction that it is disconnected. Then every component of G is circuit-free. So if there are $k > 1$ such components, then each would be a tree and hence would satisfy the theorem before. If $\mathcal{T}_i = (V_i, E_i)$ is one of these trees, then $|E_i| = |V_i| - 1$. Now, summing over all k components, we get

$$|E| = \sum_{i=1}^{k} |E_i| = \sum_{i=1}^{k} (|V_i| - 1) = n - k < n - 1,$$

which is a contradiction. Therefore, G is connected and thus (ii) implies (i) and (iii).
Suppose that G satisfies (iii). We assume by contradiction that G has a circuit and e is an edge of G contained in the circuit. Then $G \setminus e$ is still connected. If $G \setminus e$ has a circuit, we remove another edge contained in this circuit and the resulting graph is still connected. We continue this process until a connected graph with no circuit is obtained. The final graph is a tree of n vertices and thus $n-1$ edges by the last theorem. By

removing the edges of G which were contained in a circuit, the final graph has less than $n-1$ edges, which can not be possible. Therefore, we have a contradiction and hence G contains no circuits. So G is a tree and (iii) implies both (i) and (ii). □

Definition I.25 ([40])**.** A directed tree is a simple directed graph without circuits, which is connected and all vertices have indegree 1, except for one distinguished vertex which has indegree 0 and which is the root of the tree.

Remark I.26 ([40])**.** When we only consider the graph without the directions, we receive the underlying undirected tree. In a directed tree, every vertex is reachable via a directed path form the root.

Definition I.27 ([40])**.** Let $G = (V, E)$ be a directed tree and $x \in V$ be a vertex. The vertices $\{y \mid xy \in E\}$ are the children of x.
The set of vertices with no children is called the leaves of the tree and one vertex with no children is called a leaf.

Definition I.28 ([40])**.** The number of vertices $|V|$ in a graph $G = (V, E)$ is called the order of the graph.

Definition I.29 ([23], [40])**.** We consider a directed tree $\mathcal{T}$. The largest number of children a vertex of $\mathcal{T}$ may have is called the branching factor.
If the position of the children in $\mathcal{T}$ is important, we call it an ordered tree. So, in an ordered tree with branching factor m, each child of a vertex x is associated with a number from 1 to m. A child is considered to be the first one or the leftmost, the next one is the second or the second left etc. up to the m-th position, which is the rightmost child.

Definition I.30 ([40])**.** An ordered tree with branching factor m is called an m-ary tree. A 2-ary tree is called a binary tree. Each vertex may have no children, one left child, one right child or two children (one left and one right).

The origin of trees lies in computer science. Some things we introduced are called differently in computer jargon, which is the reason for the following definition.

Definition I.31 ([40])**.** We refer to the vertices as nodes, the childless nodes are called the leaves or the terminal nodes. Instead of edges, we talk about links or branches.

Definition I.32 ([23])**.** If we give a unique label to every node in a tree (for instance by different numbers), it is called a labelled tree. If two trees have different labels but the same shape, we still think of them as different trees.

Remark I.33 ([40])**.** Let $\mathcal{T}$ be an ordered tree and consider the set $\mathbb{U}_{\mathbb{N}_0} = \{\mathcal{E}\} \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}_0^n$. We write $\mathcal{E}$ for the root of the tree and $M(\mathcal{T})$ for the number of its children. If $M(\mathcal{T}) \geq 1$, we denote the first child by the letter 0, the second by the letter 1 etc. up until the last child (the $M(\mathcal{T})$-th) is denoted by the letter $(M(\mathcal{T}) - 1)$. Let u be one of these nodes, we denote by $M_u(\mathcal{T})$ the number of children of the

corresponding node u. If $M_u(\mathcal{T}) \geq 1$, we denote by the word $u0$ the first child, by $u1$ the second child etc. up to $u(M_u(\mathcal{T}) - 1)$ for the $M_u(\mathcal{T})$-th child of u. We repeat this process for all the children of the root $\mathcal{E}$. Then we apply this notation recursively for all their children.
Labelling the tree $\mathcal{T}$ gives us some advantages. Every node of the tree $\mathcal{T}$ has a unique representation as a word in the alphabet $\mathbb{N}_0$ and every tree is a language.

I.2 Some properties of the exponential distribution

Remark I.34 ([16])**.** An Exp(λ)-distributed random variable X satisfies the memoryless property, i.e for all $s, t \in \mathbb{R}$, $s, t > 0$ we have $\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s)$.

Remark I.35 ([16])**.** Let X and Y be two independent exponentially distributed random variables with parameters λ and μ respectively. Then the minimum $\min(X, Y)$ of these random variables is an exponentially distributed random variable with parameter $\lambda + \mu$. Moreover,

$$\mathbb{P}(\min(X, Y) = X) = \frac{\lambda}{\lambda + \mu} \text{ and } \mathbb{P}(\min(X, Y) = Y) = \frac{\mu}{\lambda + \mu}.$$

An Exp(λ)-distributed clock rings after a certain amount of time, where the ringing time is Exp(λ)-distributed. A particular application using the minimum of exponentially distributed random variables is the following. If we consider n Exp(1)-distributed independent clocks, the time when the first clock rings is an Exp(n)-distributed random variable and the ringing clock is uniformly among all the n clocks.

Lemma I.36 ([21])**.** *Let X and Y be independent* Exp(1)*-distributed random variables. Let*

$$U = \frac{X}{X + Y},$$

then U is Uni($[0, 1]$)*-distributed.*

Proof. We need to show that $\mathbb{P}(U \leq u) = u$ for $u \in [0, 1]$. Let $0 < u < 1$, then

$$\begin{aligned}
\mathbb{P}(U \leq u) = \mathbb{P}\left(\frac{X}{X + Y} \leq u\right) &= \mathbb{P}\left(X \leq \frac{uY}{1 - u}\right) \\
&= \int_0^\infty \left(\int_0^{\frac{uy}{1-u}} e^{-y} \cdot e^{-x} \, dx \right) dy \\
&= \int_0^\infty e^{-y} \cdot \left(1 - e^{-\frac{uy}{1-u}x}\right) dy \\
&= 1 - (1 - u) = u.
\end{aligned}$$

For $u = 0$, $\mathbb{P}(U \leq u) = \mathbb{P}(X \leq 0) = 0$ and for $u = 1$, $\mathbb{P}(U \leq u) = \mathbb{P}(0 \leq Y) = 1$. This proves the statement. □

I.3 A result about conditional expectation and equality in law

Lemma I.37 ([13] Exercise 1.12)**.** *Let $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A}$ be a sub-σ-algebra of $\mathcal{F}$. We assume that X and*

$$Y = \mathbb{E}(X \mid \mathcal{A})$$

have the same distribution, then X is $\mathcal{A}$-measurable and hence $X = Y$ almost surely.

Proof. We part this proof in two parts. First of all, we show that this result holds if X belongs to L^2.
Since X and Y have the same distribution, we have $\mathbb{E}(X^2) = \mathbb{E}(Y^2)$. Moreover, due to $\mathcal{A}$-measurability, we have

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}(\mathbb{E}(XY \mid \mathcal{A})) = \mathbb{E}(Y \cdot \mathbb{E}(X \mid \mathcal{A})) \\ &= \mathbb{E}(Y^2).\end{aligned}$$

Therefore,

$$\mathbb{E}((X-Y)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(XY) + \mathbb{E}(Y^2) = \mathbb{E}(X^2) - \mathbb{E}(Y^2) = 0,$$

which implies $X = Y$ almost surely.
For the second step, we let $a, b \in \mathbb{R}_{>0}$ and we want to show that

$$\mathbb{E}(\max(\min(X, a), (-b)) \mid \mathcal{A}) = \max(\min(Y, a), (-b)).$$

To prove this, we set $X' = \max(\min(X, a), (-b))$ and $Y' = \max(\min(Y, a), (-b))$. We notice that Y' is $\mathcal{A}$-measurable.
We apply Jensens inequality in the following conditional expectation.

$$\mathbb{E}(\min(X, a) \mid \mathcal{A}) \leq \min(\mathbb{E}(X \mid \mathcal{A}), a) = \min(Y, a).$$

Since $\min(X, a)$ and $\min(Y, a)$ have the same law, this inequality is actually an equality. Since $\mathbb{E}(\min(X, a)) = \mathbb{E}(\min(Y, a))$, the above inequality can not be strict on a set of positive probability. To prove this, we assume that $\min(Y, a) > \mathbb{E}(\min(X, a) \mid \mathcal{A})$. Then

$$\mathbb{E}(\min(Y, a)) > \mathbb{E}(\mathbb{E}(\min(X, a) \mid \mathcal{A})) = \mathbb{E}(\min(X, a)),$$

which gives us a contradiction.
Hence we get

$$\mathbb{E}(\min(X, a) \mid \mathcal{A}) = \min(Y, a)$$

and similarly

$$\mathbb{E}(\max(\min(X, a), (-b)) \mid \mathcal{A}) = \max(\min(Y, a), (-b)).$$

Now, with X' and Y', the truncated versions of X and Y, we have $X' = Y'$ almost surely by the first step. By letting $a \to \infty$ and $b \to \infty$, we get $X = Y$ almost surely. □

I.4 Some convergence results in probability theory

Theorem I.38 (Fatou's lemma, [37] Theorem 2.8)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a sequence of random variables with values in $\mathbb{R}_{\geq 0}$. Then*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

Theorem I.39 ([37] Corollary 12.16)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a non-negative martingale. Then X_n converges almost surely as $n \to \infty$ and its limit X_∞ is in L^1.*

Theorem I.40 (Doob's martingale convergence theorem, [37] Corollary 12.22)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a martingale and let $p \in (1,\infty)$. Suppose that*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}(|X_n|^p) < \infty.$$

Then X_n converges almost surely and in L^p as $n \to \infty$ to a random variable X_∞ such that

$$\mathbb{E}(|X_\infty|^p) = \sup_{n\in\mathbb{N}_0} \mathbb{E}(|X_n|^p).$$

Theorem I.41 (Scheffé's lemma, [37] Proposition 10.5)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a sequence of non-negative random variables in L^1. We assume that X_n converges in probability to X_∞ as $n \to \infty$ and $X_\infty \in L^1$. Then the condition $\mathbb{E}(X_n) \to \mathbb{E}(X_\infty)$ as $n \to \infty$ implies that X_n converges to X_∞ in L^1.*
Moreover, since $|\mathbb{E}(X_n) - \mathbb{E}(X_\infty)| \leq \mathbb{E}(|X_n - X_\infty|)$, it also holds that $\mathbb{E}(X_n) \to \mathbb{E}(X_\infty)$ if X_n converges to X_∞ in L^1 as $n \to \infty$.

Definition I.42 ([37])**.** A collection $(X_k)_{k\in K}$ of random variables in L^1 is said to be uniformly integrable if

$$\lim_{a\to\infty} \left(\sup_{k\in K} \mathbb{E}(|X_k| \mathbb{1}_{\{|X_k|>a\}}) \right) = 0.$$

Theorem I.43 ([37] Theorem 12.27)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a sequence of random variables in L^1 that converges to X_∞ almost surely. The following are equivalent:*

(i) $X_n \to X_\infty$ almost surely and in L^1.

(ii) The sequence $(X_n)_{n\in\mathbb{N}_0}$ is uniformly integrable.

Proposition I.44 ([37])**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a sequence of random variables such that $|X_n| \leq Z$ for all $n \in \mathbb{N}_0$ and Z a random variable in L^1. Then, $(X_n)_{n\in\mathbb{N}_0}$ is uniformly integrable.*
In particular, if $\mathbb{E}(\sup_{n\in\mathbb{N}_0} |X_n|) < \infty$, then $(X_n)_{n\in\mathbb{N}_0}$ is uniformly integrable with $Z = \sup_{n\in\mathbb{N}_0} |X_n|$.

I.5 A property of the continuous time Markov chain

Lemma I.45 ([3] Proposition 2.8)**.** *Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a continuous time Markov chain with state space S. We assume that it is homogeneous. Let $i \in S$ and we assume that $X_0 = i$. Then the holding time T_i in state i is exponentially distributed if i is not an absorbing state.*

Proof. We write $\mathbb{P}_i(\cdot)$ for $\mathbb{P}(\cdot \mid X_0 = i)$. Then

$$\begin{aligned}
\mathbb{P}_i(T_i > t+s) &= \mathbb{P}_i(X_u = i,\ 0 \leq u \leq t+s) \\
&= \mathbb{P}_i(X_u = i,\ t \leq u \leq t+s \mid X_u = i,\ 0 \leq u \leq t) \cdot \mathbb{P}_i(X_u = i,\ 0 \leq u \leq t) \\
&= \mathbb{P}(X_u = i,\ t \leq u \leq t+s \mid X_t = i) \cdot \mathbb{P}_i(T_i > t) \\
&= \mathbb{P}(X_u = i,\ 0 \leq u \leq s \mid X_0 = i) \cdot \mathbb{P}_i(T_i > t) \\
&= \mathbb{P}_i(T_i > s) \cdot \mathbb{P}_i(T_i > t)
\end{aligned}$$

where we used the Markov property and the homogeneity of the process. The only acceptable solution of this equation is $\mathbb{P}_i(T_i > t) = e^{-\alpha t}$ for some $\alpha \in \mathbb{R}_{\geq 0}$. Hence the holding time is exponentially distributed. □

I.6 Kolmogorov's existence theorem

Remark I.46 ([12])**.** Let $(X_t)_{t\in T}$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where T is arbitrary and the X_t takes values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra of $\mathbb{R}$. For each k-tuple $(t_1, \ldots, t_k)$ of distinct elements of T, the random vector $(X_{t_1}, \ldots, X_{t_k})$ has over $\mathbb{R}^k$ some distribution $\mu_{t_1,\ldots,t_k}$:

$$\mu_{t_1,\ldots,t_k}(A) = \mathbb{P}((X_{t_1}, \ldots, X_{t_k}) \in A),\ A \in \mathcal{B}(\mathbb{R}^k), \tag{I.1}$$

where $\mathcal{B}(\mathbb{R}^k)$ is the Borel σ-algebra of $\mathbb{R}^k$. These probability measures $\mu_{t_1,\ldots,t_k}$ are the finite-dimensional distributions of the process $(X_t)_{t\in T}$. Suppose that A in Equation I.1 is of the form $A = A_1 \times \cdots \times A_k$, with $A_i \in \mathcal{B}(\mathbb{R})$, and consider a permutation $\pi \in S_k$. Since $\{(X_{t_1}, \ldots, X_{t_k}) \in (A_1 \times \cdots \times A_k)\}$ and $\{(X_{t_{\pi(1)}}, \ldots, X_{t_{\pi(k)}}) \in (A_{\pi(1)} \times \cdots \times A_{\pi(k)})\}$ are the same event, it follows that

$$\mu_{t_1,\ldots,t_k}(A_1 \times \cdots \times A_k) = \mu_{t_{\pi(1)},\ldots,t_{\pi(k)}}(A_{\pi(1)} \times \cdots \times A_{\pi(k)}). \tag{I.2}$$

Moreover, $(X_{t_1}, \ldots, X_{t_{k-1}})$ lies in $(A_1 \times \cdots \times A_{k-1})$ if and only if $(X_{t_1}, \ldots, X_{t_k})$ lies in $(A_1 \times \cdots \times A_{k-1} \times \mathbb{R})$, so this gives us

$$\mu_{t_1,\ldots,t_{k-1}}(A_1 \times \cdots \times A_{k-1}) = \mu_{t_1,\ldots,t_k}(A_1 \times \cdots \times A_{k-1} \times \mathbb{R}). \tag{I.3}$$

Measures $\mu_{t_1,\ldots,t_{k-1}}$ coming from a process $(X_t)_{t\in T}$ always satisfy Equations I.2 and I.3.

Remark I.47 ([12])**.** Let T be an arbitrary set. Let $x \in \mathbb{R}^T$ be an element. We denote by $x(t)$ or x_t the value of x at $t \in T$. The map $Z_t \colon \mathbb{R}^T \to \mathbb{R}$ is defined by $Z_t(x) = x(t) = x_t$ and called the projection or the coordinate function. If we have defined a probability measure on $\mathbb{R}^T$, then Z_t is a random variable, called the coordinate variable.

Theorem I.48 (Kolmogorov's existence theorem, [12] Theorem 36.1)**.** *If $\mu_{t_1,\ldots,t_k}$ are a system of distributions satisfying the consistency conditions I.2 and I.3, then there is a probability measure $\mathbb{P}^T$ on $\mathbb{R}^T$ such that the process $(Z_t)_{t\in T}$ of the coordinate variables on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), \mathbb{P}^T)$ has the $\mu_{t_1,\ldots,t_k}$ as its finite-dimensional distributions.*

I.7 Some analytical results

Proposition I.49 ([28] Satz 20.2)**.** *The Gamma function satisfies*

$$\Gamma(n+1) = n!$$

for all $n \in \mathbb{N}$ *and*

$$x\Gamma(x) = \Gamma(x+1)$$

for all $x \in \mathbb{R}_{>0}$*, which gives inductively*

$$\Gamma(x+n) = \Gamma(x) \cdot x \cdot (x+1) \cdots (x+n-1)$$

for $n \in \mathbb{N}$ *and* $x \in \mathbb{R}_{>0}$*.*

Theorem I.50 (Stirling's approximation, [28] Satz 20.6)**.** *The factorial of* $n \in \mathbb{N}$ *has the asymptotic behaviour*

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

Theorem I.51 ([28] Satz 20.5)**.** *For all* $x \in \mathbb{R}_{>0}$*, the Gamma function satisfies*

$$\Gamma(x) = \lim_{n\to\infty} \frac{n! n^x}{x(x+1)\cdots(x+n)}.$$

Remark I.52 ([28])**.** Let $x \in \mathbb{R}_{>0}$. Using the above formula, we find the following asymptotic behaviour

$$\Gamma(x) \sim \frac{\Gamma(x)\Gamma(n+1)n^x}{\Gamma(x+n+1)},$$

which gives us

$$\frac{\Gamma(n)n^x}{\Gamma(x+n)} \sim 1.$$

Theorem I.53 (Banach fixed-point theorem, [26] Satz 8.1)**.** *Let* (V, d) *be a complete metric space. Let* $K\colon V \to V$ *be a contraction, i.e there exists a constant* $0 < c < 1$*, such that for all* $f, g \in V$*,*

$$d(K(f), K(g)) \leq cd(f, g).$$

Then K *admits a unique fixed point, meaning there exists a unique* $f_0 \in V$ *such that* $K(f_0) = f_0$*.*

Definition I.54 ([34])**.** Let μ be a finite measure on $\mathbb{R}$. Its Fourier transform is defined for $u \in \mathbb{R}$ by

$$F(\mu)(u) = \int e^{iux}\, \mu(dx).$$

Let $f \in L^1(\mathbb{R})$ be a integrable function with respect to the Lebesgue measure. Then its Fourier transform is defined for $u \in \mathbb{R}$ by

$$F(f)(u) = \int_{\mathbb{R}} e^{iux} f(x)\, dx.$$

Theorem I.55 ([34] Theorem 7.15)**.** *If the Fourier transform $F(\mu)$ of a finite measure μ on $\mathbb{R}$ is an integrable function with respect to the Lebesgue measure on $\mathbb{R}$, then the measure μ has a bounded, uniformly continuous density g given by*

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} F(\mu)(u) e^{-iux}\, du.$$

List of Figures

1 A complete binary tree of 15 nodes, where all leaves have the same level. 13
2 A binary search tree constructed from the keys $x_1 = 0.4$, $x_2 = 0.5$, $x_3 = 0.3$, $x_4 = 0.6$, $x_5 = 0.9$, $x_6 = 0.2$, $x_7 = 0.35$, $x_8 = 0.55$, $x_9 = 0.45$, $x_{10} = 0.33$, $x_{11} = 0.38$ and below its corresponding ranked tree. 15
3 In a binary tree of 11 nodes, we can add the 12 leaves that correspond to the possibilities where the next key can be inserted. 16
4 A 4-ary search tree constructed from the following sequence of keys: $x_1 = 0.3$, $x_2 = 0.4$, $x_3 = 0.7$, $x_4 = 0.5$, $x_5 = 0.6$, $x_6 = 0.9$, $x_7 = 0.1$, $x_8 = 0.45$, $x_9 = 0.8$, $x_{10} = 0.55$, $x_{11} = 0.42$, $x_{12} = 0.56$. 64

References

[1] G. Alsmeyer. *Kapitel I: Der einfache Galton-Watson Prozess*. Universität Münster, 2011. URL: https://www.uni-muenster.de/Stochastik/lehre/WS1011/SpezielleStochastischeProzesse/ (visited on 12/26/2024).

[2] G. Alsmeyer. *Part I: The simple Galton-Watson process: Classical approach*. Universität Münster, 2011. URL: https://www.uni-muenster.de/Stochastik/lehre/WS1011/SpezielleStochastischeProzesse/ (visited on 12/26/2024).

[3] W. J. Anderson. *Continuous-Time Markov Chains: An Applications-Oriented Approach*. 1st ed. Springer Series in Statistics. Springer New York, NY, 1991. ISBN: 978-1-4612-3038-0. DOI: 10.1007/978-1-4612-3038-0.

[4] S. Asmussen and H. Hering. *Branching Processes*. 1st ed. Birkhäuser Boston, MA, 1983. ISBN: 978-1-4615-8155-0. DOI: 10.1007/978-1-4615-8155-0.

[5] K. B. Athreya. "Some Results on Multitype Continuous Time Markov Branching Processes". In: *The Annals of Mathematical Statistics* 39.2 (1968), pp. 347–357. DOI: 10.1214/aoms/1177698395.

[6] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer Berlin, Heidelberg, 1972. ISBN: 978-3-642-65371-1. DOI: 10.1007/978-3-642-65371-1.

[7] J. Barral, X. Jin, and B. Mandelbrot. "Convergence of complex multiplicative cascades". In: *The Annals of Applied Probability* 20.4 (2010). ISSN: 1050-5164. DOI: 10.1214/09-AAP665.

[8] J. Bertoin and A. Rouault. "Discretization methods for homogeneous fragmentations". In: *Journal of the London Mathematical Society* 72.1 (2005), pp. 91–109. ISSN: 1469-7750. DOI: 10.1112/s0024610705006423.

[9] J. D. Biggins. "Martingale Convergence in the Branching Random Walk". In: *Journal of Applied Probability* 14.1 (1977), pp. 25–37. ISSN: 00219002. DOI: 10.2307/3213258.

[10] J. D. Biggins. "Uniform Convergence of Martingales in the Branching Random Walk". In: *The Annals of Probability* 20.1 (1992), pp. 137–151. DOI: 10.1214/aop/1176989921.

[11] J. D. Biggins. "Uniform Convergence of Martingales in the One-Dimensional Branching Random Walk". In: *Lecture Notes - Monograph Series* 18 (1991), pp. 159–173. ISSN: 07492170. DOI: 10.1214/lnms/1215459294.

[12] P. Billingsley. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley, 1995. ISBN: 978-0-471-00710-4.

[13] L. Chaumont and M. Yor. *Exercises in Probability: A Guided Tour from Measure Theory to Random Processes, via Conditioning*. 2nd ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2012. ISBN: 978-1-107-60655-5.

[14] B. Chauvin. *Martingales discrètes et applications à l'analyse d'algorithmes*. 2002. URL: https://chauvin.perso.math.cnrs.fr/ (visited on 07/25/2024).

[15] B. Chauvin. *Random trees for analysis of algorithms*. 2011. URL: https://chauvin.perso.math.cnrs.fr/ (visited on 02/19/2024).

[16] B. Chauvin, J. Clément, and D. Gardy. *Arbres pour l'Algorithmique*. Vol. 83. Mathématiques et Applications. Springer International Publishing, 2018. ISBN: 978-3-319-93725-0. DOI: 10.1007/978-3-319-93725-0.

[17] B. Chauvin, Q. Liu, and N. Pouyanne. "Limit distributions for multitype branching processes of m-ary search trees". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 50.2 (2014), pp. 628–654. ISSN: 0246-0203. DOI: 10.1214/12-AIHP518.

[18] B. Chauvin and N. Pouyanne. "m-ary search trees when $m \geq 27$: A strong asymptotics for the space requirements". In: *Random Structures & Algorithms* 24.2 (2004), pp. 133–154. ISSN: 1042-9832. DOI: 10.1002/rsa.10108.

[19] B. Chauvin, N. Pouyanne, and R. Sahnoun. "Limit distributions for large Pólya urns". In: *The Annals of Applied Probability* 21.1 (2011), pp. 1–32. ISSN: 1050-5164. DOI: 10.1214/10-AAP696.

[20] B. Chauvin and A. Rouault. "Connecting Yule Process, Bisection and Binary Search Tree via Martingales". In: *Journal of The Iranian Statistical Society* 3.2 (2004), pp. 89–116. DOI: 10.48550/arXiv.math/0410318.

[21] B. Chauvin et al. "Martingales and Profile of Binary Search Trees". In: *Electronic Journal of Probability* 10.12 (2005), pp. 420–435. ISSN: 1083-6489. DOI: 10.1214/EJP.v10-257.

[22] L. Devroye. "Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms". In: *Probabilistic Methods for Algorithmic Discrete Mathematics*. Ed. by M. Habib et al. Springer Berlin Heidelberg, 1998, pp. 249–314. ISBN: 978-3-662-12788-9. DOI: 10.1007/978-3-662-12788-9_7.

[23] M. Drmota. "Stochastic Analysis of Tree–like Data Structures". In: *Proceedings: Mathematical, Physical and Engineering Sciences* 460.2041 (2004), pp. 271–307. ISSN: 1364-5021. DOI: 10.1098/rspa.2003.1243.

[24] R. M. Dudley. *Real analysis and probability*. 2nd ed. Cambridge Studies in Advanced Mathematics 74. Cambridge University Press, 2002. ISBN: 978-0-511-75534-7. DOI: 10.1017/CBO9780511755347.

[25] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. 4th ed. Cambridge University Press, 2009. ISBN: 978-0-521-89806-5.

[26] O. Forster. *Analysis 2: Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgleichungen*. 11th ed. Springer Spektrum Wiesbaden, 2017. ISBN: 978-3-658-19411-6. DOI: 10.1007/978-3-658-19411-6.

[27] O. Forster. *Analysis 3: Mass- und Integrationstheorie, Integralsätze im $\mathbb{R}^n$ und Anwendungen*. 8th ed. Springer Spektrum Wiesbaden, 2017. ISBN: 978-3-658-16746-2. DOI: 10.1007/978-3-658-16746-2.

[28] O. Forster and F. Lindemann. *Analysis 1: Differential- und Integralrechnung einer Veränderlichen*. 13th ed. Springer Spekrum Wiesbaden, 2023. ISBN: 978-3-658-40130-6. DOI: 10.1007/978-3-658-40130-6.

[29] H.-K. Hwang. "Asymptotic expansions for the stirling numbers of the first kind". In: *Journal of Combinatorial Theory, Series A* 71.2 (1995), pp. 343–351. ISSN: 0097-3165. DOI: 10.1016/0097-3165(95)90010-1.

[30] J. Jabbour-Hattab. "Martingales and large deviations for binary search trees". In: *Random Structures & Algorithms* 19.2 (2001), pp. 112–127. ISSN: 1042-9832. DOI: 10.1002/rsa.1023.

[31] S. Janson. "Functional limit theorems for multitype branching processes and generalized Pólya urns". In: *Stochastic Processes and their Applications* 110.2 (2004), pp. 177–245. ISSN: 0304-4149. DOI: 10.1016/j.spa.2003.12.002.

[32] R. Kalpathy and H. Mahmoud. "Degree profile of m-ary search trees: A vehicle for data structure compression". In: *Probability in the Engineering and Informational Sciences* 30.1 (2015), pp. 113–123. ISSN: 1469-8951. DOI: 10.1017/S0269964815000303.

[33] J. F. C. Kingman. "The First Birth Problem for an Age-dependent Branching Process". In: *The Annals of Probability* 3.5 (1975), pp. 790–801. ISSN: 00911798. DOI: 10.1214/aop/1176996266.

[34] S. Lalley. *Fourier Transforms*. The University of Chicago, 2019. URL: https://galton.uchicago.edu/~lalley/Courses/381/index.html (visited on 12/02/2024).

[35] S. Lang. *Algebra*. Rev. 3rd ed. Graduate Texts in Mathematics 211. Springer New York, 2002. ISBN: 978-0-387-95385-4.

[36] G. Last and M. Penrose. *Lectures on the Poisson Process*. 1st ed. Cambridge University Press, 2017. ISBN: 978-1-316-10447-7. DOI: 10.1017/9781316104477.

[37] J.-F. Le Gall. *Measure Theory, Probability, and Stochastic Processes*. Vol. 295. Graduate Texts in Mathematics. Springer International Publishing, 2022. ISBN: 978-3-031-14205-5. DOI: 10.1007/978-3-031-14205-5.

[38] W. C. Lynch. "More Combinatorial Properties of Certain Trees". In: *The Computer Journal* 7.4 (1965), pp. 299–302. ISSN: 0010-4620. DOI: 10.1093/comjnl/7.4.299.

[39] R. Lyons and Y. Peres. *Probability on Trees and Networks*. 1st ed. Cambridge University Press, 2017. ISBN: 978-1-316-67281-5. DOI: 10.1017/9781316672815.

[40] H. M. Mahmoud. *Evolution of random search trees*. Wiley, 1992. ISBN: 978-0-4715-3228-6.

[41] H. M. Mahmoud. "The size of random bucket trees via urn models". In: *Acta Informatica* 38.11 (2002), pp. 813–838. ISSN: 1432-0525. DOI: 10.1007/s00236-002-0096-1.

[42] J. Neveu. "Multiplicative Martingales for Spatial Branching Processes". In: *Seminar on Stochastic Processes, 1987*. Ed. by E. Çinlar et al. Birkhäuser Boston, 1988, pp. 223–242. ISBN: 978-1-4684-0550-7. DOI: 10.1007/978-1-4684-0550-7_10.

[43] V. V. Petrov. *Sums of Independent Random Variables*. 1st ed. Berlin, Boston: Springer Berlin, Heidelberg, 1975. ISBN: 978-3-642-65809-9. DOI: 10.1007/978-3-642-65809-9.

[44] N. Pouyanne. "An algebraic approach to Pólya processes". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 44.2 (2008). ISSN: 0246-0203. DOI: 10.1214/07-AIHP130.

[45] J. L. Schiff. *The Laplace Transform: Theory and Applications*. 1st ed. Undergraduate Texts in Mathematics. Springer New York, 1999. ISBN: 978-0-387-22757-3. DOI: 10.1007/978-0-387-22757-3.

[46] R. B. Schinazi. *Classical and Spatial Stochastic Processes: With Applications to Biology*. 2nd ed. Birkhäuser New York, NY, 2014. ISBN: 978-1-4939-1869-0. DOI: 10.1007/978-1-4939-1869-0.