University of Copenhagen
Faculty of Science

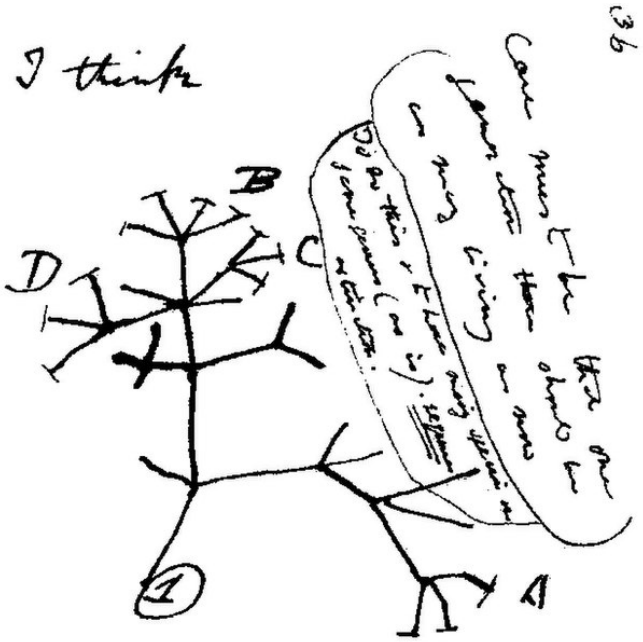
Master's Thesis in Statistics

Stochastic branching trees

Foundations and applications of a time-varying Bellman-Harris process as a macroevolutionary model

Frederik Mølkjær Andersen

December 20th 2023



Supervisors:
Carsten Wiuf & Samir Bhatt

Abstract

Stochastic models for phylogenetic trees are still not particularly well understood and most tractable models in use today do not fit empirical reduced trees very well. Age-dependent speciation models seem better at explaining data but have long been considered intractable. In this thesis, we show that, at least mathematically, this is not so. Inspired by its application in phylogenetics, we define a reduced time-varying Bellman-Harris branching process and the related reduced branching tree rigorously from first principles on branching populations and derive a full distributional characterization of both. In particular, we derive the generating function of the reduced time-varying Bellman-Harris branching process in full generality and we consider the important example of the reduced binary time-varying Bellman-Harris branching process in more detail, from which we obtain a recursive distributional characterization of the corresponding reduced binary branching tree, which we finally show to have a density function. Such density can, if treated as a likelihood function, be used to infer evolutionary dynamics from an observed reduced phylogenetic tree.

Titlepage figure:
The first known tree drawn by Charles Darwin in his notebook [16] with the exhilarating note *"I think"*. Downloaded from Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Darwins_first_tree.jpg].

Preface

> I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense.
>
> Charles Darwin

This thesis is the culmination of many years of study in the math department at the University of Copenhagen. Through these years I believe to have been awarded with *Darwin's extra sense*, I have learned the *great leading principles* of mathematics itself and now know how to exploit these to gain a better understanding of our world. For me, this is the pinnacle of mathematics, an abstract space in which thoughts and arguments are precise and rigorous, but which, when projected onto the real world, gives us insight into the very complex and otherwise incomprehensible systems of nature.

The end product you have at hand started out as a quite different beast than it turned out to be. My original plan was to do as is done in much recent theory of applied branching processes, state the relatively trivial theoretical results, and use them to gain some insight into some field of science, for me, phylogenetics. However, as I started, the deep probabilistic foundations of these processes intrigued me and led me to turn the thesis into a quite formal treatment of branching processes and branching trees, with a smaller emphasis on their use in phylogenetics. I believe that this approach has opened some otherwise obstructed paths for me, that I have only started to explore in this thesis.

Aknowledgements I am deeply grateful to so many people who have been pivotal in getting me where I am today. First, I wouldn't have gotten this far without motivation from a number of amazing teachers, the most influential in chronological order of appearance in my life being, Jens Jensen, Susanne Ditlevsen, Helle Sørensen, Ernst Hansen, Carsten Wiuf, Ben Brown and Samir Bhatt, thank you all, especially Jens for spending many hours after class showing me the beauty of mathematics, and to Carsten and Sam for providing the perfect combination of technical and inspirational supervision, I look forward to working much more together.

Maybe of even higher importance are all the people who have supported me non-academically through my studies. First and foremost, to the most overbearing, supportive, and inspiring person I know, thank you Ida for being in my life, and also, happy anniversary! Thank you to my parents, who have always motivated me to strive for the highest goals and for always supporting me on the way there. To all my family and friends, thank you for giving me a room in which to rest my thoughts, you are all immensely important to me.

Frederik Mølkjær Andersen
December 20th 2023

Contents

Abstract iii

Preface v

Contents vii

1 Introduction 1
1.1 One hundred years of stochastic phylogenetics . . . 1
1.1.1 On the origin of phylogenetics . . . 1
1.1.2 On the origin of stochastic phylogenetics . . . 2
1.1.3 Population dynamics . . . 3
1.1.4 Birth-death processes in phylogenetics . . . 3
1.1.5 Discrete stochastic models in phylogenetics . . . 4
1.2 Problems in modern stochastic phylogenetics . . . 5
1.2.1 Incorporating Age-dependence . . . 6
1.2.2 Estimation in Bellman-Harris phylogenetic trees . . . 7
1.2.3 Mathematical motivation . . . 7
1.3 A preview . . . 8
1.4 de Finetti notation . . . 9

2 Foundations of stochastic branching populations 11
2.1 The population . . . 11
2.1.1 Operations and relations on Ulam-Harris spaces . . . 11
2.1.2 Neveu trees . . . 12
2.1.3 Other subsets of individuals . . . 14
2.2 The basic sample space . . . 14
2.2.1 The realized population . . . 16
2.3 The probability measures . . . 18
2.4 The branching property . . . 22
2.4.1 Information of the non-progeny lives . . . 22
2.4.2 The simple branching property . . . 23
2.4.3 A strong branching property . . . 28

3 A reduced time-varying Bellman-Harris process 29
3.1 Random characteristics . . . 30
3.1.1 Principle of first generation . . . 31
3.2 Defining the reduced process . . . 32
3.3 Distribution of the reduced branching process . . . 33
3.4 The reduced binary branching process . . . 41
3.4.1 Probability of extinction . . . 43
3.4.2 Probability of unity . . . 44
3.4.3 Mean of the reduced process . . . 46
3.5 Death of the least common ancestor . . . 47

4 Reduced binary branching trees 53
4.1 Phylogenetic trees . . . 54
4.1.1 Phylogenetic trees as observed and modeled . . . 54
4.1.2 Graph theory and trees . . . 56
4.1.3 Binary Neveu trees . . . 57

4.2 Marked Neveu trees . . . 59
4.2.1 A σ-algebra of finite marked trees . . . 60
4.2.2 A dominating measure on marked tree space . . . 61
4.3 The reduced branching tree . . . 65
4.3.1 Ancestral lifetime distribution . . . 66
4.3.2 Defining the reduced branching tree . . . 67
4.3.3 The reduced tree distribution . . . 67

5 Conclusion and outlook **73**
5.1 Conclusion . . . 73
5.2 Estimating evolutionary dynamics . . . 74
5.3 Implementation . . . 76
5.4 Further work and limitations . . . 77

Bibliography **79**

Introduction 1

1.1 One hundred years of stochastic phylogenetics . 1
1.2 Problems in modern stochastic phylogenetics . 5
1.3 A preview 8
1.4 de Finetti notation 9

1.1 One hundred years of stochastic phylogenetics

"The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during each former year may represent the long succession of extinct species. [...] The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was small, budding twigs; and this connexion of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. [...] I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever-branching and beautiful ramifications."

Charles Darwin [17]

[17]: Darwin (1859), *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*

1.1.1 On the origin of phylogenetics

Before the mid-eighteenth century, nature was considered hierarchical and linearly ordered, the "most perfect" animals being of a higher order than the "less perfect" animals which in turn had a higher order than the "most perfect" plants, and so on. The evident evolutionary theory being that a species stems from the immediate lower-order species, resulting in conclusions like animals stemming from plants. But acknowledging that even the "most perfect" plants do not resemble even the "least perfect" animals slowly dismantled the linear order of nature. Instead, biologists turned to trees to represent the relationship between species. Now, species are not necessarily directly related, but all have a common ancestor, depicted by the trunk of the tree, which has branched out into the wealth of extant species. Nature is only partially ordered.

Hence, many authors before Darwin thought about the diversity of nature through a tree, but it was not before his and Haeckels works, that trees representing the genealogy of species were widely accepted as the fundamental structure of evolution. Darwin is especially known for two specific trees, his *unrooted* "I Think" tree as depicted on the title page of this thesis[1] and his *rooted* "Great Tree of Life" as depicted in [17]. Both of these are close to the *phylogenetic trees* used today, however, calling them as such was not done until much later.[2]

The trees played an integral role in Darwin's theory of evolution and have since been the backbone of phylogenetic thinking,[3] connecting the processes of evolution with the observed outcome through a sound mathematical framework. Of course, such a mathematical framework

1: As this does not resemble an actual botanical tree, Darwin never called it as such, but in modern-day mathematical terms this is an *unrooted tree*.

2: Haeckel actually coined the term "phylogeny", but used it differently.

3: More recently, and quite surprisingly also before Darwin, the notion of evolution following a tree has been challenged by the more general structure of a *network*, which can take into account phenomenons like *horizontal gene transfer* [67].

wasn't available to Darwin in his time, and the class of models to be considered in this thesis did not appear until almost half a century after Darwin's death. [67]

[67]: Ragan (2009), 'Trees and networks before and after Darwin'

1.1.2 On the origin of stochastic phylogenetics

Exactly 100 years before this thesis was written, a paper by George Udny Yule had been received, but not yet read, by the Royal Society. This paper entitled "A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S." [91] laid the foundation for all subsequent work on stochastic modeling of phylogenetic trees.

Yule was one of the great statisticians of the early 20th century, working on a plethora of different problems in theoretical statistics, probability theory, and applications hereof in genetics and agriculture. He was faced with data on the number of species found in a *genus*[4], which showed the interesting feature that many genera contain only one species, fewer genera had a higher number of species, but there was a few genera containing a very high number of species. A present-day statistician would say that the distribution of the number of species in a genus was *heavy-tailed* and that this particular distribution followed a *power-law distribution*. Yule wanted to find a mechanistic model of evolution that produced such a distribution.

4: A *genus* is the immediate taxonomic rank above *species*, i.e. it is a grouping of closely related species.

The model he came up with, was a continuous-time stochastic process that assumes that all species in a genus stem from a single ancestral species and that,

> *"Within any species, in any interval of time, an "accident" may happen that brings about a (viable) "specific mutation," i.e., the throwing of a new form which is regarded as a new species, but a species within the same genus as the parent. The chance of this occurrence in any assigned interval of time (an hour, or a year or a century) is taken as the same for all species within the group considered and as constant for all time."* [91]

[91]: Yule (1925), 'A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.'

In other words, new species appear at a constant rate, let's say $\beta > 0$, in the genus. This assumption led Yule to a set of differential equations for the distribution of the number of species over time, whose solution is a geometric distribution with success probability $e^{-\beta t}$. This process has since been labeled the *Yule process*. Together with an assumption that each species can also produce a whole new genus at speciation again happening at a constant rate, Yule showed how the distribution of the number of species in a random genus can be modeled by a specific heavy-tailed distribution, now known as the *Yule distribution*. While Yule didn't himself consider the fact that his process produces a genealogical tree of the extant species, his theory has since reigned as the standard model of the evolution of such trees in phylogenetics.

Yule thoroughly examines the simplifications and limitations of his model and notes for example that the model does not directly take into account extinctions of species, and that the assumption of constant rates of speciation across different species, different genus sizes and time is surely not true. Nevertheless, his model fitted the data very closely, in fact as he noted *"more closely than we have any right to expect"*[91].

1.1.3 Population dynamics

A bit later, and apparently completely separate from Yule's work on stochastic models for speciations, a school of stochastic population dynamics was rising with people like Willy Feller and David Kendall. Their main contribution was the birth-death process [19, 45], in which each individual in a population is assumed to give birth at a rate β and die at a rate δ. Kendall [45] even proposed a further generalization, where the birth and death rates are allowed to vary over time. The corresponding population size distributions over time in both the constant rate and time-varying birth-death process are slightly more complicated than those of the Yule process, but still quite manageable. Now, if we by birth mean speciation, and by death mean extinction, this becomes a generalization of the Yule process, where any particular species can either produce new species or go extinct according to the corresponding rates.

[19]: Feller (1939), 'Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung'
[45]: Kendall (1948), 'On the Generalized "Birth-and-Death" Process'

Another branch of population dynamics, which has in recent years also proved a valuable class of models for phylogenetics, is the theory of *branching processes*. The first branching processes, which were evolving in discrete generations, were studied by first by Irénée-Jules Bienaymé [10], and later by Francis Galton and Henry William Watson [88], all trying to answer a question about the extinction of noble families surnames. A little-known curiosity, especially interesting for Danes, is that the first correct proofs of Bienayme, Galton, and Watson's assertions were made by Johan Frederik Steffensen [84]. For more on the Danish contributions to the early theory of branching processes, see [24].

[10]: Bienaymé (1845), 'De la loi de multiplication et de la duree des familles'
[88]: Watson et al. (1875), 'On the Probability of the Extinction of Families'

[84]: Steffensen (1930), 'Om Sandsynligheden for at Afkommet uddør'

The beginning of continuous-time branching processes starts with the birth-death processes, as these, in particular, are a form of branching process, but was extended from here by Richard Bellman and Theodore Harris [9, 28] to be non-Markovian by letting each individual have a non-exponentially distributed lifetime, at the end of which she produces a random number of offspring. If this number of offspring is independent of the lifetime, the process is aptly named the *Bellman-Harris process*. This process will form the foundation of the current work. A situation where the number of individuals born by each individual depends on the lifetime was considered by Boris Aleksandrovich Sevast'yanov [72], and such processes are hence called *Sevat'yanov processes*. This line of research was very much inspired by problems in physics (splitting of atoms) and biology (splitting of cells and bacteria), but not much applied in population dynamics until the generalizations by Kenny Crump, Charles Mode [14, 15] and simultaneously Peter Jagers [35] where births happen during the lifetime of an individual, not necessarily just at the end of its life, according to a point process.

[9]: Bellman et al. (1952), 'On Age-Dependent Binary Branching Processes'
[28]: Harris (1963), *The theory of branching processes*

[72]: Sevast'yanov (1964), 'Age-dependent branching processes'

[14]: Crump et al. (1969), 'A general age-dependent branching process. II'
[15]: Crump et al. (1968), 'A general age-dependent branching process. I'
[35]: Jagers (1969), 'A general stochastic model for population development'

1.1.4 Birth-death processes in phylogenetics

The birth-death process being an apparent generalization of the Yule process as a mechanistic model of evolution, was not evident enough to be considered in phylogenetic studies until [69], partly due to lack of good empirical data, partly because the process has a severe flaw as a model for phylogenetic trees; they usually show no extinctions. If we have no paleontological data, a traditional *reconstructed* or *reduced* phylogenetic tree contains only the extant species and the lineages tracing them back to

[69]: Raup et al. (1973), 'Stochastic Models of Phylogeny and the Evolution of Diversity'

a common ancestor, and we hence do not see any extinctions in this tree, even though they clearly represent an important part of the evolution of the tree, why it is usually not sufficient to use the Yule process.

This complication was solved in [32, 59], deriving the distribution, not only of the number of species through time having at least one extant progeny at observation, but also the distribution of the reduced tree generated by such a process, both for the constant and time-varying situations.[5] Together with these papers, and the influx of higher-quality data, a new interest in using stochastic processes for phylogenetics was spurred, especially with papers deriving estimators for the speciation and extinction rate based on the reconstructed phylogenetic tree [30, 56, 58]. For a good overview of the use of birth-death models in phylogenetics in general, see [57].

[32]: Hey (1992), 'Using Phylogenetic Trees to study Speciation and Extinction'
[59]: Nee et al. (1994), 'The Reconstructed Evolutionary Process'

5: Whenever the process is Markovian, as the birth-death processes and the Yule process, the tree distribution is quite easy to find as a jump is equally likely to happen at any branch due to memorylessness.

[30]: Harvey et al. (1994), 'Phylogenies Without Fossils'
[56]: Nee (2001), 'Inferring Speciation Rates from Phylogenies'
[58]: Nee et al. (1994), 'Extinction rates can be estimated from molecular phylogenies'
[57]: Nee (2006), 'Birth-Death Models in Macroevolution'

A *very* large number of extensions have since been made to the birth-death process as a model of molecular evolution, some interesting directions are the inclusion of different sampling schemes of the extant individuals [33, 76], the inclusion of fossil records (and thus coming closer to modeling the *complete phylogenetic tree*) [55, 77], dependence on the number of co-existing species [65] and species-dependent speciation and extinction rates [4]. For an overview of the different extensions, and how many of them are actually a special case of a more general process called the *coalescent point process*, see [49, 74].

[49]: Lambert et al. (2013), 'Birth–death models and coalescent point processes'
[74]: Stadler (2013), 'Recovering speciation and extinction dynamics based on phylogenies'

Even sticking to the basic constant rate birth-death model however there has been an abundance of papers deriving *different* likelihood equations and other estimation frameworks for the estimation of the speciation and extinction rates, see e.g. [21, 68, 76, 85, 90]. The main difference being what to condition the process on when estimating. Should one condition on the number of extant species? On survival of the process until the present? The time since the origin of the process? The time since the death of the least common ancestor? This certainly leads to a number of different likelihood equations, and a number of different estimators, and extending to the time-varying situation certainly doesn't lower the number of different approaches to estimation, see e.g. [55, 66, 79]. A thorough comparison of these different estimators and recommendations on what to condition on is given in [80].

[80]: Stadler (2013), 'How Can We Improve Accuracy of Macroevolutionary Rate Estimates?'

More general branching processes have only been studied very sparsely as a model for phylogenetics, and most of the work has been on Markovian branching processes [3, 64], which are in many ways just another formulation of birth-death processes. In recent years though, there have been a few publications [22, 26, 41] on the use of age-dependent branching processes in phylogenetics, as they might solve a quite important problem seen in most Markovian processes of tree evolution, namely that the induced distribution over the discrete phylogenetic tree shapes are too *balanced*.

[3]: Aldous et al. (2005), 'A Critical Branching Process Model for Biodiversity'
[64]: Popovic (2004), 'Asymptotic genealogy of a critical branching process'
[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'
[26]: Hagen et al. (2015), 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'
[41]: Jones (2011), 'Calculations for multitype age-dependent binary branching processes'

1.1.5 Discrete stochastic models in phylogenetics

Tree balance is usually a topic that centers around the *tree shape* of a phylogenetic tree, without considering any temporal components. I.e. if we threw away the times at which each branch was initiated under one of our continuous-time processes, we would be left with a tree shape.

Tree balance can be measured in many different ways, no one of which is perfect [20], but all of them essentially measure how well balanced each *clade* of the trees is, that is, if we remove the root of the phylogenetic tree, is there a big difference between the number of branches in the two resulting subtrees or not?

If we consider the Yule process stopped when it has a given number of individuals, the induced distribution over tree shapes is well-known and called the *Yule-Harding distribution*, first derived from a discrete-time version of the Yule process in [27] and connected to the continuous-time version in [2]. The Yule-Harding distribution has a high probability mass at very balanced trees, that is, at trees where each clade is roughly of the same size. Another way to approach this is from a purely combinatorial point of view, where we could consider a discrete distribution over the space of phylogenetic tree shapes with a given number of extant individuals. The simplest distribution we could endow such a space with is the discrete uniform distribution, in phylogenetics often called *proportional to distinguishable* or simply *PDA*. As this is a uniform distribution, it is not especially concentrated at either the very balanced or unbalanced trees, it is simply flat.

[27]: Harding (1971), 'The Probabilities of Rooted Tree-Shapes Generated by Random Bifurcation'
[2]: Aldous (1996), 'Probability Distributions on Cladograms'

When compared to empirical trees as done in [11], it is seen that the PDA distribution has too high a concentration on the very unbalanced trees, while the Yule-Harding distribution has too high a concentration on the very balanced tree, so one would preferably want a model "in between" these two. David Aldous proposed in [2] the β-splitting model for this, which he *"pulled out of thin air"*. Here $\beta \in (-2, \infty)$ is a parameter that can be changed to get as balanced or unbalanced trees as one would like, in particular, it can produce both the PDA distribution ($\beta = -3/2$) and the Yule-Harding distribution ($\beta = 0$), and [11] showed that this distribution fitted the empirical tree well with $\beta = -1$. This model might thus seem favorable to the Yule-Harding distribution or the PDA distribution, but the model has no clear biological interpretation, and in particular, has no natural mechanistic continuous-time stochastic model for evolution been shown to induce this distribution over tree shapes. In contrast, has a very broad class of continuous-time models for phylogenetic trees been shown to induce the, too balanced, Yule-Harding distribution.

[11]: Blum et al. (2006), 'Which Random Processes Describe the Tree of Life?'

[11]: Blum et al. (2006), 'Which Random Processes Describe the Tree of Life?'

1.2 Problems in modern stochastic phylogenetics

The Yule process, the birth-death process, and their many extensions still serve as the backbone of most stochastic phylogenetic thinking, and in particular, these are the models that primarily are implemented in software for estimating speciation and extinction rates [73, 79] as well as a choice for the prior distribution of trees in Bayesian tree reconstruction methods [18, 70], and are thus the most applied models.

But as noted, a large class of models, including most of these processes, have been shown in [48, 49] to all induce the, too balanced, Yule-Harding distribution over tree shapes. This class of models is called *species-speciation-exchangeable* meaning that at any given time, the speciation rate

[48]: Lambert (2010), 'The contour of splitting trees is a Lévy process'
[49]: Lambert et al. (2013), 'Birth–death models and coalescent point processes'

must be the same for all extant species, and they are thus *exchangeable*. This class includes extended birth-death processes which allows for

- Speciation rate changing over time
- Extinction rate changing over time *and with the species' age*
- Number of co-existing species
- Independent sampling of extant individuals

on the other hand, the model does not include processes whose speciation rates are age-dependent.

This result has shown that, if we are only interested in the tree shape distribution, and thus not any temporal component of the phylogenetic tree, then either of these models will give us the same answer, in particular, could we freely choose to use the Yule or birth-death process. Now, as it is fairly well-established that the Yule-Harding distribution does not fit empirical trees very well, these models are surely problematic if we want to represent the true evolution of species, in particular if we will use them for estimating evolutionary dynamics. See [74] for a thorough look at the consequences of these models and proposals for further work.

[74]: Stadler (2013), 'Recovering speciation and extinction dynamics based on phylogenies'

1.2.1 Incorporating Age-dependence

One obvious departure from the class of species-speciation exchangeable models is a model where the speciation rate is age-dependent, this could for example be a binary Bellman-Harris process with a non-exponential lifetime distribution. As the Bellman-Harris process is generally much more intractable than Markovian processes, and the fact they have not been considered as models for phylogenetic trees until quite recently, there are not a lot of theoretical results on e.g. their induced distribution over tree shapes. In [22], which considers the special case with no extinctions, a joint density function for the tree shape and branch lengths is given through a recursive formula. Conditioning this on the number of extant individuals and integrating over the branch lengths gives a formal characterization of the distribution, but it is not in general clear how such a distribution looks, and in particular it is not clear how it is concentrated on more or less balanced trees, as it is highly dependent on the choice of lifetime distribution.

[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'

With the particular choice of a Weibull-distributed lifetime, and again assuming no extinctions, [26] have shown that the induced distribution over tree shapes can actually interpolate between the Yule-Harding distribution and the PDA distribution, much like is possible with Aldous' β-splitting model. The difference from Aldous' β-splitting model, besides being a continuous-time process, is that the Bellman-Harris process has a very clear biological interpretation in that species exist for a random period of time, after which they might speciate and might go extinct.[6] Further simulation studies in [26, 42] with different choices of lifetime distributions, together with altering choices on the speciation and extinction probability at the end of each individual's life, confirm this result for more general classes of lifetime distributions.

6: This is certainly a very crude picture of evolution, but is nonetheless an approximation to what happens in nature.

[26]: Hagen et al. (2015), 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'

[42]: Jones (2011), 'Tree Models for Macroevolution and Phylogenetic Analysis'

In [41], a series of equations regarding the (multi-type) randomly sampled Bellman-Harris process as a phylogenetic model is stated and heuristically argued for, and then a few applications of these equations are proposed.

[41]: Jones (2011), 'Calculations for multitype age-dependent binary branching processes'

In particular, an expression for a given tree-balance statistic is given and computed for a process with Pareto-distributed lifetimes, here showing that altering the sampling rate can change how the tree shape distribution is concentrated. Another proposed application of these equations is to compute the joint density function of the tree shape and branch lengths using a recursive formula like in [22] but with a much more general model. Again, one could in theory marginalize this distribution over the branch lengths to get the distribution over tree shapes, but this is not considered. [41] speculates in this distribution being a possible prior over the tree space (with branch lengths), in a Bayesian reconstruction of a phylogenetic tree, but doesn't consider it as a likelihood function for estimating the speciation and extinction dynamics of the process.

A contrasting study to those above is that of [87] which argues that an age-dependent model is not favorable in comparison to a simple Yule process. This study however only considers branch lengths, not tree shapes, and thus has naturally not found the same results as [26, 42]. It should be noted that [26] also considers the distribution of branch lengths in their simulation study, and likewise finds no noticeable change in branch lengths when changing the lifetime or offspring distribution.

[87]: Venditti et al. (2010), 'Phylogenies reveal new interpretation of speciation and the Red Queen'

1.2.2 Estimation in Bellman-Harris phylogenetic trees

Incorporating age-dependence in phylogenetic tree models is thus motivated by the possibility of gaining more flexible models that can better fit the empirically observed tree shapes. In particular, it has been noted that in a Bayesian phylogenetic tree reconstruction, the choice of a Yule or birth-death process prior or the phylogenetic tree space, might bias the posterior distribution in an unwanted way [42]. However, this should also motivate us to use a more flexible model in estimating the temporal components of evolution from empirical trees, in particular, if a model produces the wrong distribution over the tree shapes alone, we should have no hope in it to produce the correct[7] estimates for the speciation and extinction dynamics.

7: There of course is no such thing as a *correct* model, but we might strive to get the model that explains data the best.

Another speculative reason for incorporating age-dependence in a model for estimating evolutionary dynamics is that it has recently been showed that even under a simple time-varying birth-death process, the speciation and extinction rates are not identifiable [53] from the number of lineages through time in a reconstructed phylogeny,[8] which in particular is equivalent to it being unidentifiable from the whole reconstructed tree, as it is well known that the likelihood equation does not depend on the tree shape in a species-speciation exchangeable process [68, 78]. In contrast, is the density, and thus likelihood function, of [41] very much dependent on the tree shape, so one might speculate that if we can exploit more information from an empirical tree, this might make the time-varying rates identifiable. This has to the best knowledge of the author not been studied.

8: Note that [50, 51] has shown that a quite flexible class of rate functions are identifiable.

1.2.3 Mathematical motivation

From a purely mathematical point of view, there has, as mentioned, been little work on incorporating non-Markovian branching processes

into a phylogenetic framework, and even less on time-varying non-Markovian branching processes. Actually, the whole field of continuous-time branching processes contains very little literature on time-varying processes.[9] Specifically, for the time-varying Bellman-Harris process, the only work known to the current author is [46]. A few papers exist for the more general time-varying Crump-Mode-Jagers process [40, 61, 62], but in general researchers are often more interested in limiting behaviors of branching processes, in which general time-variability is hard to incorporate. Hence a more thorough mathematical examination of a time-varying Bellman-Harris process in itself would be interesting, furthermore, if such a process could be applicable in a phylogenetic setting, this would be a generalization of the results of [22, 26, 41].

9: Such processes are often called *branching processes in a varying environment*.

[46]: Kimmel (1983), 'The Point-Process Approach to Age- and Time-Dependent Branching Processes'

[40]: Jagers et al. (1985), 'Branching Processes in Periodically Varying Environment'

[61]: Pakkanen et al. (2023), 'Unifying incidence and prevalence under a time-varying general branching process'

[62]: Penn et al. (2023), 'Intrinsic randomness in epidemic modelling beyond statistical uncertainty'

The natural extension of the reconstructed birth-death model of [59], is called the *reduced branching process*, and is the process counting the number of individuals through time which leaves at least one extant progeny. This process was first proposed for Bellman-Harris processes in [86] and later studied in [71]. Although it is essentially this process that [41] considers, it is not mentioned in the paper, and it generally never seems to have explicitly been used as a model for phylogenetic trees. No known studies have characterized the full distribution of the reduced Bellman-Harris process either or taken time-variability of the lifetime and offspring distribution into account.

[86]: Vatutin (1979), 'Distance to the nearest common ancestor in bellman-harris branching processes'

[71]: Sagitov (1986), 'Reduced Critical Bellman–Harris Branching Process with Several Types of Particles'

1.3 A preview

This thesis tries to shed light on some new directions in models for phylogenetics inspired by the described limitations and problems of the birth-death process and its extensions and to give solid mathematical ground to a field that is often dominated by heuristic arguments.

In Chapter 2 we start by investigating the foundational theory for branching processes and see that a generalization of the Bellman-Harris process, which takes into account time-variability of the lifetime and offspring distribution, fits into a general framework of branching processes proposed in [37]. In particular, will we construct sample spaces and probability measures hereon such that we later can define our branching processes rigorously, and we will prove a Markov-like property of our probability space called the branching property. This branching property will be essential in the remainder of the thesis, as proof after proof will revolve around this property.

[37]: Jagers (1989), 'General branching processes as Markov fields'

Continuing to Chapter 3, we get to define the classical notion of a branching process, together with the reduced branching process, both under the general framework of branching processes counted with a random characteristic. With the definitions at hand, we go on to make a full characterization of the distribution of the reduced branching process by finding an integral equation for its generating function over time. Specializing to a reduced binary branching process, we derive a series of equations for interesting quantities similar to those of [41], but here stated in more generality and derived from the generating function of the process.

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

In Chapter 4, we initially consider how to model reduced phylogenetic trees and come up with a framework for doing so from a branching process. In particular, we define the space of finite trees with edge lengths as a measure space, in which a random reduced branching tree will take values. Finally, we get to a characterization of the distribution of the reduced branching tree through the quantities that we have described in Chapter 3. This distribution is then seen to have a density with respect to the measure we defined on the space of trees. This density is the crown jewel of the thesis.

Finally, in Chapter 5, we look at how all the math we went through might have important consequences for scientific advancements, in particular, of course, in phylogenetics. The advantages and limitations of our model are explored and a look at further directions of study is given. We also consider the feasibility of an implementation of this method, as to make it readily available for practitioners.

This thesis contains, to the best knowledge of the author, several (partly) novel results, in particular, we have throughout our results incorporated time-variability which is not considered anywhere else, and we have formulated and proved all results on a rigorous measure-theoretic foundation, which again is not done anywhere else. In the introduction to each chapter, we specify which results are novel.

1.4 de Finetti notation

We will in general use quite standard notation of probability theory, and aptly label any deviations or special notation throughout the thesis. We will however use a somewhat unconventional convention for writing expectations. All around the thesis, we will be constructing new probability measures, and whenever we want to take an expectation it should be made clear with respect to which probability measure we are integrating. Now, instead of the more traditional way of signifying this, e.g. by using subscripts on the expectation operator such as $\mathbb{E}_{\mathbb{P}}$ and $\mathbb{E}_{\mathbb{Q}}$ for the integrals with respect to $\mathbb{P}$ and $\mathbb{Q}$ respectively, we simply use the symbol for the measure as the expectation operator itself. I.e. if X is a real integrable random variable on some measurable space $(\mathcal{X}, \mathcal{E})$ on which both $\mathbb{P}$ and $\mathbb{Q}$ are probability measures, we write

$$\mathbb{P}X := \int X \, \mathrm{d}\mathbb{P} \quad \text{and} \quad \mathbb{Q}X := \int X \, \mathrm{d}\mathbb{Q}.$$

This is called *de Finetti* notation, and is not only useful in that it can simplify some notation, but when you first get used to it, I personally believe that it makes many arguments clearer, one very simple example is that moving between expectations of indicator functions and the probability of their events become much cleaner. For more motivation on why to use de Finetti notation, see [63, Chapter 1.4].

[63]: Pollard (2001), *A User's Guide to Measure Theoretic Probability*

Foundations of stochastic branching populations

2

2.1 The population 11
2.2 The basic sample space . 14
2.3 The probability measures 18
2.4 The branching property . 22

2.1 The population

The fundamental unit of interest in branching processes is the set of all potential individuals who could possibly live, give birth to a number of new individuals, and die. We will call this set of individuals *the branching population* or simply *the population*.[1] Branching populations are characterized by originating from a single ancestral individual, call her *the ancestor*, this ancestor has a number of potential offspring, each of which can be considered the ancestor of their own sub-population.

To study the branching population, it is necessary to construct a labeling of its members. Any such labeling should needless to say be rich enough to reflect the actual empiric population under study. First, we label the ancestor of any population by 0. If we do not restrict the number of potential offspring of each individual, we can label the potential children of 0 by $\mathbb{N} = \{1, 2, ...\}$, her potential grandchildren by $\mathbb{N}^2$ and so on. Let us for now, not repeat the semi-redundant "potential" in front of all kinship. This labeling will then in general be given as $(x_1, x_2, ..., x_n) \in \mathbb{N}^n$ for some $n \in \mathbb{N}$ being the x_n'th child of the x_{n-1}'th child of $\cdots$ the x_1'st child of the ancestor. If we identify each individual with its labeling, the branching population becomes the *full Ulam-Harris space*

$$I_{\mathbb{N}} := \bigcup_{n=0}^{\infty} \mathbb{N}^n$$

with the convention $\mathbb{N}^0 = \{0\}$.

In applications where we a priori restrict the number of potential offspring of any individual to be $N \in \mathbb{N}$, we can reduce the offspring labeling set to $\{1, ..., N\}$ and obtain the *N-Ulam-Harris space* branching population

$$I_N := \bigcup_{n=0}^{\infty} \{1, ..., N\}^n$$

again letting $\{1, ..., N\}^0 = \{0\}$. A particularly interesting branching population for us is the *binary Ulam-Harris space* where $N = 2$, this arises for example when individuals can only ever die or split into two. Having decided on the choice of branching population, we will often just denote it by I.

2.1.1 Operations and relations on Ulam-Harris spaces

On any of these Ulam-Harris spaces, we can define a series of operations to make life easier in the sequel. For $x = (x_1, ..., x_{n-1}, x_n) \in I \setminus \{0\}$ its *generation* is $\mathfrak{g}(x) = n$ and its *mother* is given as

$$\mathfrak{m}x := (x_1, ..., x_{n-1})$$

1: The population we discuss here is of course just an abstract representation of some collection of actual individuals, groups of individuals, cells, or other kinds of particles. But we will often think of it, and refer to it, as a population of female humans considering that our individuals can "give birth". Thus we use phrases like x is the mother of y or y is the daughter of x to reflect that y is an offspring of x. This is also in the tradition of branching processes and population dynamics more generally [36].

when $\mathfrak{g}(x) > 1$ and $\mathfrak{m}x := 0$ if $\mathfrak{g}(x) = 1$. Naturally, we can iterate the use of $\mathfrak{m}$ and denote by $\mathfrak{m}^{k+1}x$ the k'th *grandmother* of x whenever $\mathfrak{g}(x) > k$. By convention, we let $\mathfrak{g}(0) = 0$, $\mathfrak{m}0 = 0$ and $\mathfrak{m}^0 x = x$ for all $x \in I$. Moreover, the *rank* (in sibship) of x is given as $\mathfrak{r}x = x_n$, but note that the ranking of siblings with a common mother is often completely arbitrary unless we allow for offspring to be born sequentially during the life of their mothers.

For two individuals $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_k)$ the *concatenated individual* xy is given as

$$xy := (x_1, ..., x_n, y_1, ..., y_k)$$

where we let 0 act as an additive identity, $0x = x0 = x$ for all $x \in I$. Note that $\mathfrak{m}x\,\mathfrak{r}x = x$.

The maternal operation $\mathfrak{m}$ induces a partial order $\preceq$ on the Ulam-Harris space where for $x, y \in I$,

$$x \preceq y \iff \exists k \in \mathbb{N}_0 : x = \mathfrak{m}^k y$$

and we say that y *stems from* x. Note that x stems from x, which is not always the intention, so we also define the strict partial order,

$$x \prec y \iff \exists k \in \mathbb{N} : x = \mathfrak{m}^k y.$$

If for $x, y \in I$ we either have $x \preceq y$ or $y \preceq x$, we say that x and y are *comparable*, in particular 0 is comparable with any $x \in I$ while if $\mathfrak{g}(x) = \mathfrak{g}(y)$ then x and y are only comparable if $x = y$.

For two subsets $L, M \subset I$ we write $L \preceq M$ if for each $y \in M$ there is a $x \in L$ such that $x \preceq y$. Seen this way $\preceq$ is also a partial order on 2^I, and whether we denote by $\preceq$ the relation on I or on 2^I should be clear from the context. Under $\preceq$ (as a relation on I), any non-empty finite subset of the population $M \subseteq I$ has a greatest lower bound, that is the greatest element $y \in I$ such that $y \preceq M$.[2] We call this element the *least common ancestor* and denote it $\mathfrak{lca}(M)$. This fact renders I a lower-semilattice.[3]

We will in some cases need a linear order on I. This entails no problems in practice as we can define an arbitrary linear extension of $\prec$, thus preserving the mother-daughter ordering. One possible linear extension of $\prec$ is $<$ which satisfies $0 < 1 < 2 < 11 < 3 < 111 < \cdots$ where $x < y$ if the digit-sum of x is less than the digit-sum of y and the ordering is arbitrarily chosen if they have the same digit-sum. As daughters of x are of the form xk for $k \in \mathbb{N}_0$, they clearly have a higher digit-sum than their mother, and other individuals with the same digit-sum must be non-comparable with the daughter under $\prec$.

2: We can actually talk about y being the *greatest* element as any two lower bounds of S must be comparable.

3: The central motivation for using Ulam-Harris-type spaces to represent branching populations is in fact that they are semilattices. Hence much of the theory just works if we more generally let I be a semilattice, [37] provides examples.

2.1.2 Neveu trees

An Ulam-Harris space will usually represent all potential individuals in a branching population, even those who might never be born. Restricting our attention to only individuals satisfying some criteria, for example, those who are actually born at some point gives rise to certain subsets of Ulam-Harris space called Neveu trees.[4]

4: There is no canonical name for such trees, but when the Ulam-Harris type labeling is used for trees it is often called the Neveu notation as Jacques Neveu seems to be the first to consider trees as generated by branching processes in [60].

In concurrence with graph-theoretic *planted oriented trees*[5] and the intuitive sense of a genealogical tree, a Neveu tree should contain a distinguished root vertex, the ancestor, the tree should be connected, that is the genealogical lines of individuals should be preserved in the subset, and the offspring of each individual should be labeled increasingly. Formally, we say a subset $\gamma \subseteq I$ is a *Neveu tree* or simply a *tree* if the following holds

- $0 \in \gamma$
- $x \in \gamma \implies \mathfrak{m}x \in \gamma$
- $yj \in \gamma \implies yi \in \gamma$ for all $i \leq j$.

5: See Subsection 4.1.2 for more on the connection to graph theory.

moreover, if each generation is finite, we call the Neveu tree *locally finite*, that is,

- $\#(\gamma \cap \mathbb{N}^n) < \infty$ for all $n \in \mathbb{N}_0$

Note that the underlying Ulam-Harris space I is, of course, itself a Neveu tree, but it is only locally finite if we specifically work with an N-Ulam-Harris space, as each generation is otherwise of infinite size. The conditions for a Neveu tree above are sufficient for all of the *retrospective* operations and relations given in Subsection 2.1.1 to be identically defined on Neveu trees, that is, all but the concatenation of individuals, which in general does not work. In particular, every tree is a partially ordered set in which all finite subsets have a greatest lower bound, i.e. a least common ancestor.

Example 2.1.1 Consider the finite Neveu tree $\gamma = \{0, 1, 2, 11, 12, 21, 22\}$. The following statements are obvious, $1 \prec 11$, $\mathfrak{r}(12) = 2$, and $\mathfrak{g}(21) = 2$. Likewise, we can find least common ancestors of any subset of γ, for example

$$\mathfrak{lca}\{12, 21\} = 0, \quad \mathfrak{lca}\{11, 12\} = 1, \quad \text{and } \mathfrak{lca}\{2, 21\} = 2.$$

Contrary to what is possible on I, we cannot concatenate individuals without some care, for example, concatenating 2 and 11 is not possible, as 211 is not contained in the tree, but concatenating 2 and 1 is possible as $21 \in \gamma$. See Figure 2.1 for a graphical depiction of this Neveu tree from which the statements above should be evident.

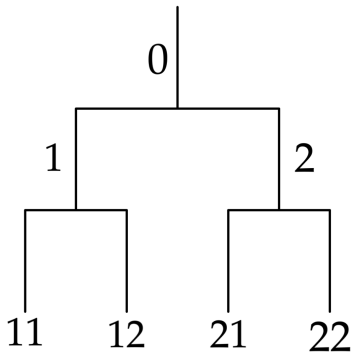


Figure 2.1: The Neveu tree of Example 2.1.1.

We now introduce a few natural collections of Neveu trees,

$$\begin{aligned} \mathcal{N}_\infty &:= \{\gamma \subseteq I : \gamma \text{ Neveu tree}\} \\ \mathcal{N} &:= \{\gamma \subseteq I : \gamma \text{ locally finite Neveu tree}\} \\ \mathcal{N}_0 &:= \{\gamma \subseteq I : \gamma \text{ finite Neveu tree}\} \end{aligned}$$

which are nested with the following inclusions,

$$\mathcal{N}_0 \subseteq \mathcal{N} \subseteq \mathcal{N}_\infty \subseteq 2^I$$

In Chapter 4 there will be much further expansion of the theory of Neveu trees, but for now they are just specific subsets of the population.

2.1.3 Other subsets of individuals

Neveu trees are not the only interesting subsets of individuals. Motivated by the genealogical order $\preceq$ of individuals, we define the *ancestors* and the *progeny* of a given set $M \subset I$,

$$\operatorname{An} M := \{x \in I : x \prec M\}$$
$$\operatorname{Pr} M := \{x \in I : x \succeq M\}.$$

Note that $M \subset \operatorname{Pr} M$ but $M \not\subset \operatorname{An} M$, this is simply a convention, but it is important to remember. Other than this slightly unnatural convention the two sets have exactly the meaning that their names convey.

As a perpendicular concept from genealogically motivated sets, we can define a set V to be a *stopping line* or simply a *line* if no two distinct elements of V are comparable under $\prec$, that is, V is a stopping line if

$$x, y \in V \implies x \not\prec y \text{ and } y \not\prec x.$$

Hence if we think of the genealogical lines of the population as moving vertically wrt. $\prec$, stopping lines are cutting through them horizontally and individuals of a line can hence not be directly related. Trivial examples of stopping lines are $\mathbb{N}^n$ for all $n \in \mathbb{N}$, any singleton individual, and the *head* hM of any subset $M \subset I$ defined as

$$hM = \{x \in M : \forall y \in I : y \prec x \implies y \notin M\}.$$

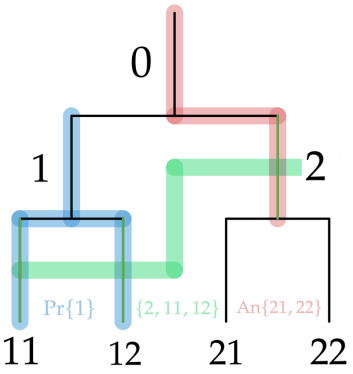


Figure 2.2: Illustration of an ancestor set, progeny set and a stopping line as described in Example 2.1.2.

Example 2.1.2 Consider the Neveu tree of Example 2.1.1. On this tree, we can illustrate the genealogical subsets of individuals and stopping lines. Note that these are of course generally defined on the full Ulam-Harris space, but this is very big and uncomfortable to illustrate. On Figure 2.2 we see superimposed on the Neveu tree, the progeny set of 1, restricted to the tree, $\operatorname{Pr}\{1\} \cap \gamma$ depicted in blue, the stopping line $\{2, 11, 22\}$ depicted in green and the ancestor set of $\{21, 22\}$ depicted in red. Notice that the green stopping line goes in the opposite direction of the blue and red genealogical lines.

Stopping lines will play a pivotal role in our analyses as we will see that when two individuals are non-comparable, and we know the history of those individuals' ancestors, they will behave independently of each other.

2.2 The basic sample space

The following exposition will consider the full Ulam-Harris space $I_{\mathbb{N}}$, however little to no effort lies in restricting the proofs to work on any I_N for $N \in \mathbb{N}$, in particular, all the results still hold under the binary Ulam-Harris space.

To construct the probability spaces on which our processes live, we now endow each individual $x \in I$ with a copy of a measurable space $(\Omega, \mathcal{A})$, which we will call the *life space*.[6] The elements $\omega \in \Omega$ are the possible *life*

6: As of now, $(\Omega, \mathcal{A})$, and in a bit also its extensions, are just measurable spaces, but we will in the prequel define probability measures on them, hence, without trouble, we can already now, at least verbally, treat them as probability spaces

careers of the individual. This entails that all relevant characteristics of the life of an individual should be possible to frame as random elements on $(\Omega, \mathscr{A})$, that is, they must be measurable mappings of the life careers onto some value-space. For our purposes, this means that Ω should be rich enough that we can define a random variable $L : (\Omega, \mathscr{A}) \to (\mathbb{R}_+, \mathscr{B}_+)$ and a stochastic process $N := \{N(t)\}_{t\in\mathbb{R}}$ where $N(t) : (\Omega, \mathscr{A}) \to (\mathbb{N}_0, 2^{\mathbb{N}_0})$ for each $t \in \mathbb{R}$. L will represent the lifetime of an individual while $N(t)$ will represent the number of offspring an individual produces were it to die at time t.

Remark 2.2.1 Contrary to the approach in [61, 62], we let the time dependence of the lifetime distribution be handled by varying the background probability measure instead of constructing a sequence of variables with different distributions. This allows for a rigorous measure-theoretic treatment, especially for formulations of conditional distributions and dependence structures in the population.

[61]: Pakkanen et al. (2023), 'Unifying incidence and prevalence under a time-varying general branching process'
[62]: Penn et al. (2023), 'Intrinsic randomness in epidemic modelling beyond statistical uncertainty'

We can consider all potential individuals' life careers at once through the product space $\left(\Omega^I, \mathscr{A}^I\right) := (\bigtimes_{x\in I} \Omega, \bigotimes_{x\in I} \mathscr{A})$, called the *population space*. The sample space on which the processes under study will be defined, however, should contain all these life careers as well as the origin time of the population, as two populations will evolve differently when initiated at different times. I.e. we define our background sample space, which we will call the *time-population space*, as

$$(\overline{\Omega}, \overline{\mathscr{A}}) := (\mathbb{R} \times \Omega^I, \mathscr{B} \otimes \mathscr{A}^I)$$

where $(\mathbb{R}, \mathscr{B})$ is the real Borel space representing the possible origin times of the process. In this way, an element $\varpi \in \overline{\Omega}$ of the time-population space is of the form $\varpi = (\tau_0, \{\omega_x \in \Omega : x \in I\})$ representing the origin time of the population together with the life careers of all potential individuals.

The population space $\left(\Omega^I, \mathscr{A}^I\right)$ being a product space of the life spaces, it is natural to define the canonical projections onto sub-product spaces, and in particular onto each individual life space. For any $M \subseteq I$ we let $(\Omega^M, \mathscr{A}^M) := (\bigtimes_{x\in M} \Omega, \bigotimes_{x\in M} \mathscr{A})$ and define $U_M \colon \left(\Omega^I, \mathscr{A}^I\right) \to (\Omega^M, \mathscr{A}^M)$ as,

$$U_M(\omega) := \{\omega_x : x \in M\} \tag{2.1}$$

when $\omega = \{\omega_x : x \in I\}$. For any $\varpi = (\tau_0, \omega) \in \overline{\Omega}$ we can likewise define the same projections as being from the time-population space to any $(\Omega^M, \mathscr{A}^M)$ as $U_M(\varpi) := U_M(\omega)$, and we will notationally not distinguish between these two. For further notational ease we will also at times write $\omega^M := U_M(\omega)$. With these projections, we can lift any random variable defined on the life space to the population space or time-population space by function composition of the random variable and the projection onto the life space on which it is defined, i.e. if ξ is a generic random variable defined on $(\Omega, \mathscr{A})$, we let ξ_x be the variable defined on $\left(\Omega^I, \mathscr{A}^I\right)$ or $(\overline{\Omega}, \overline{\mathscr{A}})$ by

$$\xi_x(\omega) := \xi(U_x(\omega))$$

for all $\omega \in \Omega^I$ or likewise for $\varpi \in \overline{\Omega}$.

As the number of offspring of each individual is assumed finite, but each generation in our full Ulam-Harris space is infinite, this space is clearly unnecessarily large for the description of the actually realized population. However as it is still a countable space, most constructions are still fairly simple, and in cases where it is essential only to consider the individuals who are actually born, we can focus on that part of the space.

2.2.1 The realized population

Denote by $\tau_x \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to (\mathbb{R}, \mathscr{B})$ the (possibly not realized) *birth time* of $x \in I$, defined recursively for any $\varpi = (\tau, \{\omega_x : x \in I\}) \in \overline{\Omega}$ as $\tau_0(\varpi) := \tau$ and

$$\tau_x(\varpi) := \tau_{\mathfrak{m}x}(\varpi) + L_{\mathfrak{m}x}(\varpi)$$

for the remaining x in $I \setminus \{0\}$. Notice that τ_x only depends on the origin time τ_0 and the projection $U_{\mathrm{An}\, x}$ onto $(\Omega^{\mathrm{An}\, x}, \mathscr{A}^{\mathrm{An}\, x})$, so we will often think of it as a function of the life of its ancestors, $\tau_x \colon (\mathbb{R} \times \Omega^{\mathrm{An}\, x}, \mathscr{B} \otimes \mathscr{A}^{\mathrm{An}\, x}) \to (\mathbb{R}, \mathscr{B})$ or with the domain being any higher product space than that. In the same way, we can define the *death time* or *splitting time*[7] of each individual $x \in I$ as $\sigma_x \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to (\mathbb{R}, \mathscr{B})$ by

$$\sigma_x(\varpi) := \tau_x(\varpi) + L_x(\varpi).$$

7: We will use these two names interchangeably, as in some situations it is important to note that the individual is dying at this time, while in other situations is it more important that the individual (possibly) produces new individuals at this time.

Notice that $\tau_x(\varpi) = \sigma_{\mathfrak{m}x}(\varpi)$ for all $x \neq 0$, i.e. each individual's birth time is her mother's death time. This, together with the coming assumption that the lifetime variable and the offspring process are independent, is exactly the assumptions making our process a *Bellman-Harris process*.

Projections onto the progeny of each individual will be of special importance as if two individuals are non-comparable under $\preceq$, we will see that their progeny will live on independently of each other. Thus we introduce the shift operator $S_x \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to (\overline{\Omega}, \overline{\mathscr{A}})$ for each $x \in I$ on the time-population space, essentially making x the ancestor of its own sub-population initiated at time τ_x

$$S_x(\varpi) := (\tau_x(\varpi), U_{\mathrm{Pr}\, x}(\varpi)) = \left(\tau_x(\varpi), \{\omega_y : x \preceq y\}\right). \tag{2.2}$$

As noted, the space I contains all potential individuals, even the ones who are never born, and even though we have now given them a "birth time", this is not a natural construct for individuals who are not born. Consequently, we define the set of actually realized individuals and their realized birth times. Initially, the *realized individuals in generation* $n \in \mathbb{N}_0$ can be given recursively as a random subset of I, $\mathscr{R}_n \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to \left(2^{\mathbb{N}^n}, 2^{2^{\mathbb{N}^n}}\right)$

$$\begin{aligned}
&\mathscr{R}_0(\varpi) := \{0\},\\
&\mathscr{R}_n(\varpi) := \left\{x \in \mathbb{N}^n : \mathfrak{m}x \in \mathscr{R}_{n-1}(\varpi),\ rx \leq N_{\mathfrak{m}x}\left(\tau_x(\varpi), \varpi\right)\right\}
\end{aligned}$$

and the random *set of all realized individuals* $\mathscr{R} \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to \left(2^I, 2^{2^I}\right)$ is then given by $\mathscr{R}(\varpi) := \bigcup_{k \in \mathbb{N}} \mathscr{R}_n(\varpi)$. We note that the value space of the full

set of realized individuals can by construction be restricted to the set of locally finite Neveu trees, equipped with the power-set σ-algebra $(\mathcal{N}, 2^{\mathcal{N}})$, as $N(t)$ is a.s. finite for all $t \in \mathbb{R}$.

Remark 2.2.2 In the seminal papers by Jacques Neveu and Brigitte Chauvin [12, 60], another sample space is constructed for the Galton-Watson and Bellman-Harris processes, which is not excessively large for the definition of the processes. The elements of this sample space are on the contrary exactly the realized Neveu trees, where each individual is marked with their life length. In this way, the whole life career, in particular the life length and number of offspring, of each individual is specified from the tree. In the same way, we can *heuristically* think about each element $\omega \in \overline{\Omega}$ as a realized Neveu tree with marks on each individual corresponding to their lifetime.

In contrast to the usual approach of leaving the sample space unspecified, including that of the current thesis, this is a very direct construction of the sample space, but one that is not uncommon in the theory of branching processes, see for example, Harris' construction in [28, Chapter VI Sec. 2].

We conjecture that parts of the theory in this thesis might have been more elegant in the Neveu-Chauvin framework, however, the papers have to the knowledge of the author never been translated from their original French, making the content quite inaccessible.

[12]: Chauvin (1986), 'Arbres et processus de Bellman-Harris'
[60]: Neveu (1986), 'Arbres et processus de Galton-Watson'

[28]: Harris (1963), *The theory of branching processes*

With these definitions at hand, we can define the *realized birth time* of individuals $x \in I$, $\tilde{\tau}_x \colon (\overline{\Omega}, \overline{\mathcal{A}}) \to \left(\mathbb{R} \cup \{\infty\}, \mathcal{B}_\infty\right)$ recursively as

$$\tilde{\tau}_x(\omega) := \begin{cases} \tau_x(\omega), & \text{if } x \in \mathcal{R}(\omega), \\ \infty, & \text{if } x \in I \setminus \mathcal{R}(\omega) \end{cases}$$

and $\tilde{\tau}_0\left((\tau, \{\omega_x : x \in I\})\right) = \tau$. This has the interpretation that individuals with birth time ∞ are never born. The *realized death time* or *realized splitting time* is similarly defined, $\tilde{\sigma}_x \colon (\overline{\Omega}, \overline{\mathcal{A}}) \to (\mathbb{R}, \mathcal{B})$ for each $x \in I$ as

$$\tilde{\sigma}_x(\omega) := \tilde{\tau}_x(\omega) + L_x(\omega)$$

where, again, the realized splitting time of a non-realized individual is ∞.

Finally, like the definition of the realized n'th generation, we might be interested in the *realized individuals, alive at time* $t \in \mathbb{R}$. This can, for all $t \in \mathbb{R}$ be defined as the random set $\mathcal{I}_t \colon (\overline{\Omega}, \overline{\mathcal{A}}) \to \left(2^I, 2^{2^I}\right)$,

$$\mathcal{I}_t(\omega) := \{x \in I : \tilde{\tau}_x(\omega) \leq t < \tilde{\sigma}_x(\omega)\}$$

If we do not care whether or not the individual has died again, we obtain the (finite) Neveu tree of *individuals born before time* t, also called the *complete tree at time* t, $\Upsilon_t \colon (\overline{\Omega}, \overline{\mathcal{A}}) \to (\mathcal{N}_0, 2^{\mathcal{N}_0})$, given by

$$\Upsilon_t(\omega) := \{x \in I : \tilde{\tau}_x(\omega) \leq t\}$$

The assertion that Υ_t takes values in the space of finite Neveu trees is due to the offspring process only taking finite values and the lifetime distribution being strictly positive, so a positive amount of time has to

pass between two births.

2.3 The probability measures

For each life space, we now want to specify the distribution of the two random elements L and $\{N(t)\}_{t\in\mathbb{R}}$ as image measures of some background life law. The distributions of these elements are considered dependent on the individual's birth time, so we must define the probability measure on the life spaces to be dependent on the birth time of the individual too. Then, from these time-dependent measures on the life spaces, we will further establish a probability measure on the population space, again being time-dependent, but now only of the origin time of the full population, i.e. the birth time of the ancestor, as the lives of each individual will then fully determine the birth time of their children. Choosing a prior distribution over the origin times of the population, we also gain a probability measure on the time-population space.

All this is achieved formally by assuming we have a $(\mathbb{R},\mathscr{B})$-Markov kernel $(\Lambda_\tau)_{\tau\in\mathbb{R}}$ on each life space, such that Λ_τ is the *life law* of an individual born at time $\tau\in\mathbb{R}$. We will see that this induces a $(\mathbb{R},\mathscr{B})$-Markov kernel $(\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ on the population space, such that for $\tau\in\mathbb{R}$, $\mathbb{P}_\tau$ gives the life law for the whole population. If we for an individual $x\in I$ in a population originating at time $\tau\in\mathbb{R}$, know her birth time, i.e. τ_x, then the life of x follows the life law Λ_{τ_x} independently of everything else.

The setting here is akin to that of a m-stage experiment where the outcome of the first $m-1$ experiments affects the distribution of outcomes from the m'th experiment. Our case is complicated slightly by being an experiment with a countably infinite number of successive experiments, only some of which affect each other. This can, however, be dealt with by evoking Ionescu-Tulcea's extension theorem and switching between our partial order $\preceq$ on I and the chosen linear extension $\leq$.

Proposition 2.3.1 *Let $(\Lambda_\tau)_{\tau\in\mathbb{R}}$ be an $(\mathbb{R},\mathscr{B})$-Markov kernel on $(\Omega,\mathscr{A})$, then there is a unique $(\mathbb{R},\mathscr{B})$-Markov kernel $(\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ on $\left(\Omega^I,\mathscr{A}^I\right)$ such that for all $\tau\in\mathbb{R}$, finite $M\subset I$ and $A_x\in\mathscr{A}$, for each $x\in M$*

$$\mathbb{P}_\tau\left(U_M^{-1}\left[\bigtimes_{x\in M}A_x\right]\right)=\int_{\tilde{A}_0}\cdots\int_{\tilde{A}_{x_n}}\prod_{y=x_n}^{0}\mathrm{d}\Lambda_{\tau_y(\omega^{\mathrm{An}\,y})}(\omega_y)$$

with $0,x_1,x_2,\ldots,x_n$ being all the elements of $\mathrm{An}\,M\cup M$ ordered according to $\leq$ and with $\tilde{A}_x\in\mathscr{A}$ given as

$$\tilde{A}_x=\begin{cases}A_x, & x\in M\\ \Omega_x, & x\in\mathrm{An}\,M.\end{cases}$$

Proof. Let $(\Omega_x,\mathscr{A}_x)$, $x\in I$ be copies of the life space and for a given $\tau\in\mathbb{R}$ consider Λ_τ a probability measure on $(\Omega_0,\mathscr{A}_0)$.

Under the linear order $\leq$ on I we can identify each $x\in I$ with a unique natural number $i_x\in\mathbb{N}_0$ being its rank under $\leq$. We will in the rest of the

proof interchangeably use x and i_x to mean the appropriate element of I. This allows us to write something like $i_x \in I$ or $\operatorname{An} i_x$ while actually referring to $x \in I$ and $\operatorname{An} x$ respectively.

We can then define the k'th product space, being the product space of the first k individuals under $\leq$,

$$(\Omega^k, \mathscr{A}^k) := \left(\bigtimes_{i=0}^{k} \Omega_i, \bigotimes_{i=0}^{k} \mathscr{A}_i \right)$$

note in particular that in the limit $(\Omega^\infty, \mathscr{A}^\infty) = (\Omega^I, \mathscr{A}^I)$. For each $k \in \mathbb{N}$ we then construct a $(\Omega^{k-1}, \mathscr{A}^{k-1})$-Markov kernel ${\kappa_k}^8$ on $(\Omega_k, \mathscr{A}_k)$ such that for $x \in I \setminus \{0\}$ and $\omega^{i_x - 1} \in \Omega^{i_x - 1}$,

$$\kappa_{i_x}\left(\omega^{i_x-1}, \cdot\right) = \Lambda_{\tau_x(\omega^{i_x-1})}(\,\cdot\,).$$

8: Using the same notation for κ_k as for our life kernel introduces an overload of subscripts, hence we use a slightly different notation where the index enters as the first argument of the kernel seen as a function.

The tensor product of these first k sequential Markov kernels is itself a $(\Omega_0, \mathscr{A}_0)$-Markov kernel $\bigotimes_{i=1}^{k} \kappa_i$ on $(\Omega^k, \mathscr{A}^k)$ and the integration of this Markov kernel with respect to the probability measure Λ_τ then defines a probability measure, as we treat τ as fixed,

$$\mathbb{P}_\tau^k = \Lambda_\tau \otimes \bigotimes_{i=1}^{k} \kappa_i$$

on $(\Omega^k, \mathscr{A}^k)$ [47, Corollary 14.27]. By Ionescu-Tulcea's Theorem [47, Theorem 14.35] we thus have a unique probability measure $\mathbb{P}_\tau$ on $(\Omega^I, \mathscr{A}^I)$ such that for $A \in \mathscr{A}^k$, $k \in \mathbb{N}_0$, we have

$$\mathbb{P}_\tau\left(A \times \bigtimes_{i>k} \Omega_i\right) = \mathbb{P}_\tau^k(A).$$

[47]: Klenke (2020), *Probability Theory*

For finite $M \subset I$, the pre-image of the rectangle $\bigtimes_{x \in M} A_x$ where $A_x \in \mathscr{A}$, $x \in M$, under U_M becomes

$$U_M^{-1}\left[\bigtimes_{x \in M} A_x\right] = \bigtimes_{x \leq M} \tilde{A}_x \times \bigtimes_{x \not\leq M} \Omega_x$$

where $\tilde{A}_x \in \mathscr{A}$, $x \in \operatorname{An} M \cup M$ is given by

$$\tilde{A}_x = \begin{cases} A_x, & x \in M \\ \Omega_x, & x \notin M \end{cases}$$

and so, if the maximal element of M, i.e. the minimal element of I under $\leq$ such that all $x \leq M$ also satisfies $x \leq m$, is of rank $m \in \mathbb{N}_0$ under $\leq$,

the probability of the rectangle under $\mathbb{P}_\tau$ becomes

$$\begin{aligned}
\mathbb{P}_\tau\left(U_M^{-1}\left[\bigtimes_{x\in M} A_x\right]\right) &= \mathbb{P}_\tau^m\left(\bigtimes_{x\leq M}\tilde{A}_x\right) = \int_{\times_{x\in[0,M]}\tilde{A}_x} \mathrm{d}\left(\Lambda_\tau \otimes \bigotimes_{i=1}^m \kappa_i\right) \\
&= \int_{\widetilde{A}_0}\int_{\widetilde{A}_1}\cdots\int_{\widetilde{A}_m} \mathrm{d}\kappa_m(\omega^{m-1},\omega_m)\cdots\mathrm{d}\kappa_1(\omega_0,\omega_1)\,\mathrm{d}\Lambda_\tau(\omega_0) \\
&= \int_{\widetilde{A}_0}\int_{\widetilde{A}_1}\cdots\int_{\widetilde{A}_m} \mathrm{d}\Lambda_{\tau_{x_m}(\omega^{m-1})}(\omega_m)\cdots\mathrm{d}\Lambda_{\tau_1(\omega_0)}(\omega_1)\,\mathrm{d}\Lambda_\tau(\omega_0) \\
&= \int_{\widetilde{A}_0}\int_{\widetilde{A}_1}\cdots\int_{\widetilde{A}_m} \mathrm{d}\Lambda_{\tau_{x_m}(\omega^{\mathrm{An}\,x_m})}(\omega_m)\cdots\mathrm{d}\Lambda_{\tau_1(\omega_0)}(\omega_1)\,\mathrm{d}\Lambda_\tau(\omega_0) \\
&= \int_{\widetilde{A}_0}\int_{\widetilde{A}_1}\cdots\int_{\widetilde{A}_m} \prod_{k=m}^{0}\mathrm{d}\Lambda_{\tau_{x_k}(\omega^{\mathrm{An}\,x_k})}(\omega_k)
\end{aligned}$$

[47]: Klenke (2020), *Probability Theory*

where we have used Tonelli's Theorem for kernels [47, Theorem 14.32], the definition of κ_k and that τ_x can be thought of as a function on $(\Omega^{\mathrm{An}\,x}, \mathcal{A}^{\mathrm{An}\,x})$ when τ is fixed.

Now, many of the integrals above will evaluate to one, as long as they are not ancestral to M. To show this we can find a maximal non-ancestor $y \leq M$, i.e. an element $y \notin \mathrm{An}\,M \cup M$ such that all other non-ancestors have a lower rank under $\leq$ than y. This ensures that $x \in \mathrm{An}\,M \cup M$ for all $x > y$, and in particular we see the remaining nested integrals inside the $\Lambda_{\tau_y(\omega^{\mathrm{An}\,y})}(\omega_y)$ integral cannot depend on ω_y, and can thus be pulled out. Using that Λ_τ is a probability measure for all $\tau \in \mathbb{R}$ and that $\tilde{A}_y = \Omega_y$ as $y \notin M$, the $\Lambda_{\tau_y(\omega^{\mathrm{An}\,y})}(\omega_y)$ integral evaluates as

$$\begin{aligned}
&\int_{\tilde{A}_{i_y}} \underbrace{\left(\int_{\tilde{A}_{i_y+1}}\cdots\int_{\tilde{A}_m}\prod_{k=m}^{i_y+1}\mathrm{d}\Lambda_{\tau_{i_k}(\omega^{\mathrm{An}\,i_k})}(\omega_{i_k})\right)}_{\text{not dependent on }\omega_{i_y}} \mathrm{d}\Lambda_{\tau_{i_y}(\omega^{\mathrm{An}\,i_y})}(\omega_{i_y}) \\
&= \int_{\tilde{A}_{i_y}} \mathrm{d}\Lambda_{\tau_{i_y}(\omega^{\mathrm{An}\,i_y})}(\omega_{i_y}) \int_{\tilde{A}_{i_y+1}}\cdots\int_{\tilde{A}_m}\prod_{k=m}^{i_y+1}\mathrm{d}\Lambda_{\tau_{i_k}(\omega^{\mathrm{An}\,i_k})}(\omega_{i_k}) \\
&= \Lambda_{\tau_y(\omega^{\mathrm{An}\,y})}(\Omega_y)\int_{\tilde{A}_{i_y+1}}\cdots\int_{\tilde{A}_m}\prod_{k=m}^{i_y+1}\mathrm{d}\Lambda_{\tau_{i_k}(\omega^{\mathrm{An}\,i_k})}(\omega_{i_k}) \\
&= \int_{\tilde{A}_{i_y+1}}\cdots\int_{\tilde{A}_m}\prod_{k=m}^{i_y+1}\mathrm{d}\Lambda_{\tau_{i_k}(\omega^{\mathrm{An}\,i_k})}(\omega_{i_k})
\end{aligned}$$

Repeating this argument until there are no more non-ancestor integrals left and denoting the $\leq$-ordered elements of $\mathrm{An}\,M \cup M$ by $0, x_1, \ldots, x_n$, we gain the expression as wanted

$$\mathbb{P}_\tau\left(U_M^{-1}\left[\bigtimes_{x\in M} A_x\right]\right) = \int_{\tilde{A}_0}\cdots\int_{\tilde{A}_{x_n}}\prod_{y=x_n}^{0}\mathrm{d}\Lambda_{\tau_y(\omega^{\mathrm{An}\,y})}(\omega_y)$$

For given $\tau \in \mathbb{R}$, $\mathbb{P}_\tau$ is clearly a probability measure, and for given $A = U_M^{-1}(A_M)$ for finite $M \subset I$ with maximal element of rank m, $A_M \in \mathcal{A}^M$ we have $\tau \mapsto \mathbb{P}_\tau(A) = \mathbb{P}_\tau^m(A_M)$ is $(\mathbb{R}, \mathcal{B})$ measurable as $\mathbb{P}_\tau^m$ is a kernel for all m. Hence $(\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ defines a $(\mathbb{R}, \mathcal{B})$-kernel on $\left(\Omega^I, \mathcal{A}^I\right)$, as it is sufficient to check the $(\mathbb{R}, \mathcal{B})$ measurability of $\tau \mapsto \mathbb{P}_\tau(A)$ for sets in a $\cap$-stable class generating $\mathcal{A}^I$, for example the class of sets of the form

$U_M^{-1}(A_M)$. □

From now on we will hence consider the life and population spaces as probability spaces each endowed with their respective kernel giving them, not only one, but a sequence of probability measures. In most of our treatment, we will solely be considering an a priori known origin time of the whole population, and hence a chosen measure from the kernel. Changing the origin time will hence mostly be relevant when considering shifted populations i.e. when working with S_x for some $x \in I$, in which case the origin time of the shifted population will change to be the birth time of the individual x, now considered the ancestor. If we put a prior distribution over the origin time, however, say $\tau_0 \sim O$ for some probability measure O on $(\mathbb{R}, \mathscr{B})$, then the integration measure $O \otimes (\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ provides us with a probability measure on the time-population space [47, Corollary 14.26]. Back in the situation where we priori know the origin time, it might still come in handy to have a measure on the time-population space, and simply choosing the prior distribution on origin-times as having full measure at the known value give us such a measure,

[47]: Klenke (2020), *Probability Theory*

Remark 2.3.1 For a given $\tau \in \mathbb{R}$, Proposition 2.3.1 defines $\mathbb{P}_\tau$ as a measure on $\left(\Omega^I, \mathscr{A}^I\right)$, but we will often need to treat it a measure on $(\overline{\Omega}, \overline{\mathscr{A}})$ but where we know that $\tau_0 = \tau$. We should thus consider $(\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ a regular conditional distribution of $\bar{\omega} = (\tau_0, \omega)$ given τ_0, and associate with $\mathbb{P}_\tau$ the integration measure $\delta_\tau \otimes (P_\tau)_{\tau\in\mathbb{R}}$ where δ_τ is a Dirac measure on $(\mathbb{R}, \mathscr{B})$ with full measure at τ. I.e. we could for $A \in \mathscr{A}^I$ and $B \in \mathscr{B}$ write

$$\mathbb{P}_\tau(B \times A) = \mathbb{P}_\tau(A)\, \mathbf{1}_B(\tau)$$

or the other way around,

$$\mathbb{P}_\tau(A) = \mathbb{P}_\tau(\{\tau\} \times A)$$

This is of course an abuse of notation, but it will nonetheless make much of our exposition clearer, hence we make the sacrifice.

In the same way, when we by $\mathbb{P}_\tau$ denote an expectation operator, we may apply it to random variables defined on $(\overline{\Omega}, \overline{\mathscr{A}})$. This again poses no real problem as it corresponds to associating $\mathbb{P}_\tau$ with the integration measure $\delta_\tau \otimes (\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$ and we thus have for any integrable or non-negative function $\xi\colon (\overline{\Omega}, \overline{\mathscr{A}}) \to (\mathbb{R}, \mathscr{B})$,

$$\begin{aligned}
\mathbb{P}_\tau \xi &= \int \xi \, \mathrm{d}\delta_\tau \otimes \mathbb{P}_\tau \\
&= \int \int \xi(\tau_0, \omega) \, \mathrm{d}\mathbb{P}_\tau(\omega) \, \mathrm{d}\delta_\tau(\tau_0) \\
&= \int \xi(\tau, \omega) \, \mathrm{d}\mathbb{P}_\tau(\omega) \\
&= \mathbb{P}_\tau[\xi(\tau, \cdot)].
\end{aligned}$$

From a practical point of view, treating $\mathbb{P}_\tau$ as a regular conditional distribution simply lets us change any τ_0 inside of a probability or an

expectation to the deterministic τ and to treat any variable dependent on $\varpi = (\tau_0, \omega) \in \overline{\mathscr{A}}$ as only dependent on $\omega \in \mathscr{A}^I$.

The processes we will consider are intricate combinations of the random variables defined on the probability spaces above, and if we did not know anything about the dependence structure of these variables, analyzing such processes would be inconceivable. The construction of $\mathbb{P}_\tau$, and in particular the dependence structure between each instance of the life law, however, reveals an independence property of the population, which we will return to again and again in characterizing distributions of variables and processes defined from the population. Unmasking this property is the goal of the next section.

2.4 The branching property

The main trouble in the search for this independence property is that it is not grounded in the real-time evolution of the process, as it usually is for continuous-time stochastic processes, e.g. the Markov property that loosely states that the future is independent of the past given the present. Instead, the independence is between the future progeny of an individual and the remaining population, given the birth time of the individual. I.e. the dependence is *genealogical* in the sense that if we know the birth time that the mother gives to her child, the life law of the child and all of its progeny are known and independent of the remaining population. The property is seen by regrouping integrals in the expression in Proposition 2.3.1 such that the distribution of the progeny on the individual appears under its ancestors' integrals determining the birth time of the individual.

We will further see that stopping lines, i.e. sets of individuals that are not directly related, are interesting in generalizing this dependence structure, as they form a set of individuals whose progeny will live on independently of each other and the rest of the process if we know each of their birth times. This again has the consequence that we can regroup the integrals for each individual in the line, as these will not depend on each other, into each conditional distribution given their birth times. This would not be the case for a general set of individuals as the birth time of an individual is certainly not independent of its ancestors.

2.4.1 Information of the non-progeny lives

To state such a result, we need to define an object that contains the known information that we want to condition on, i.e. we want to define a sub-σ-algebra of $\mathscr{A}^I$, containing information about all the population, but the progeny of the individuals in a line. Thus for a stopping line[9] $V \subset I$, the *non-progeny σ-algebra* of V, denoted by $\mathscr{F}_V$ is the σ-algebra generated by the projection mapping onto all life spaces, but the ones for the progeny of the line, i.e.

$$\mathscr{F}_V := \sigma\left(U_x : x \notin \operatorname{Pr} V\right).$$

A property of the non-progeny σ-algebra makes it, at first, slightly unnatural to work with. Consider two lines, one contained in the other,

9: Note that there is nothing stopping us from defining the σ-algebras with respect to sets of individuals which are not lines. We define them for lines as these are the only sets we wish to condition on.

$V' \subset V$. In this situation, V' has more non-progeny than V as the progeny of the individuals in $V \setminus V'$ is also non-progeny of V', hence more projection mappings are measurable with respect to $\mathcal{F}_{V'}$ than with respect to $\mathcal{F}_V$. In other words, the *larger* the line, the *less* information is contained in the corresponding non-progeny σ-algebra, $\mathcal{F}_V \subset \mathcal{F}_{V'}$.

Considering lines that stem from each other, however, defines the natural nesting of these σ-algebras. Notice that if $V \preceq V'$ then $\operatorname{Pr} V' \subseteq \operatorname{Pr} V$ so there are more non-progeny of V' than of V, and hence $\mathcal{F}_V \subset \mathcal{F}_{V'}$. This allows us to consider $\{\mathcal{F}_V : V \subset I \text{ line}\}$ as a *filtration* under $\preceq$. Notice in particular that if V is infinite, then we can look at a sequence $\{V_n\}_{n\in\mathbb{N}}$ where V_n contains the first[10] n individuals of V and thus have that $V_n \subset V_{n+1}$. Thus we have $\mathcal{F}_{V_n} \supset \mathcal{F}_{V_{n+1}}$ and thus $\mathcal{F}_{V_n} \downarrow \mathcal{F}_V$ as needed for *backwards martingale convergence* [47, Theorem 12.14].

10: "First" meaning just some principled way of choosing individuals from V without replacement. One could take the n individuals ranked lowest under the linear order $\preceq$.

[47]: Klenke (2020), *Probability Theory*

A slightly simpler σ-algebra, will be helpful in proofs, namely the *ancestor σ-algebra*, denoted by $\mathcal{H}_V$, which is generated only by the projection mappings onto the ancestors of the line,

$$\mathcal{H}_V := \sigma\left(U_x : x \in \operatorname{An} V\right)$$

note that $\mathcal{H}_V \subset \mathcal{F}_V$ as $\operatorname{An} V \subset (\operatorname{Pr} V)^\complement$. The idea with using $\mathcal{H}_V$ over $\mathcal{F}_V$ is that if V is finite, then the set $\{U_x : x \in \operatorname{An} V\}$ of the generating mappings is finite, and one can show that the σ-algebra is then generated by the $\cap$-stable class

$$\mathcal{G}_{\mathcal{H}_V} := \left\{ U^{-1}_{\operatorname{An} V}\left(\bigtimes_{y\in\operatorname{An} V} A_y\right) : A_y \in \mathcal{A}^I \text{ for each } y \in \operatorname{An} V \right\}$$

which contains Ω^I. The generating sets in $\mathcal{G}_{\mathcal{H}_V}$ are much preferable for example when we want to apply Proposition 2.3.1. The important point is now that sometimes it doesn't matter whether we condition on $\mathcal{F}_V$ or $\mathcal{H}_V$, and in that situation, we surely prefer to work with the easy one.

2.4.2 The simple branching property

The first, and for us most important, form of the branching property is now stated and proved. The proof has some technically detailed steps, but the essence of it is to exploit the partial associativity of Proposition 2.3.1, in that we can collect the integrals for the progeny of each individual, forming the full conditional distribution of those progeny given the individual's birth time, which is itself determined from the integrals for the rest of the population, i.e. the non-progeny.

Theorem 2.4.1 *Let $\tau \in \mathbb{R}$. For any $x \in I$, $\mathbb{P}_{\tau_x}$ is a regular conditional distribution of S_x given $\mathcal{F}_x$ and the Branching Property holds, i.e. for any line $V \subset I$ and functions $\varphi_x : (\overline{\Omega}, \overline{\mathcal{A}}) \to ([0,1], \mathcal{B}_{[0,1]})$ for each $x \in V$,*

$$\mathbb{P}_\tau\left[\prod_{x\in V} \varphi_x(S_x) \,\middle|\, \mathcal{F}_V\right] = \prod_{x\in V} \mathbb{P}_{\tau_x}[\varphi_x].$$

In particular, we have that the shifted populations of a line $\{S_x : x \in V\}$ are conditionally independent given $\mathcal{F}_V$.

Proof. **We start by finding the regular conditional distribution of S_x given $\mathcal{F}_x$,** for a given individual $x \in I$ in the population if it has origin time τ. That is, we want to find a $(\Omega^I, \mathcal{F}_x)$-Markov kernel on $(\overline{\Omega}, \overline{\mathcal{A}})$, let us call it κ_x, such that

$$\kappa_x(\omega, \overline{A}) = \mathbb{P}_\tau\left(S_x \in \overline{A} \,\middle|\, \mathcal{F}_x\right)(\omega)$$

[47]: Klenke (2020), *Probability Theory*

for $\mathbb{P}_\tau$-almost all $\omega \in \Omega^I$ and all $\overline{A} \in \overline{\mathcal{A}}$ [47, Definition 8.28]. Let us initially examine the conditional probability with $\overline{A} = B \times A$ for $B \in \mathcal{B}$ and $A \in \mathcal{A}^I$,

$$\begin{aligned}\mathbb{P}_\tau(S_x \in B \times A \,|\, \mathcal{F}_x) &= \mathbb{P}_\tau((\tau_x, U_{\Pr x}) \in B \times A \,|\, \mathcal{F}_x) \\ &= \mathbf{1}_{(\tau_x \in B)}\, \mathbb{P}_\tau(U_{\Pr x} \in A \,|\, \mathcal{F}_x)\end{aligned} \tag{2.3}$$

where we can pull out the indicator on the τ_x event as it is $\mathcal{F}_x$ measurable. Instead of working on the full conditional probability here we initially just consider $\mathbb{P}_\tau(U_{\Pr x} \in A \,|\, \mathcal{F}_x)$.

As noted above working with the σ-algebra $\mathcal{F}_V$ is not always easy, and the current arguments are only possible as we can choose to work with $\mathcal{H}_V$ instead. Note that the σ-algebra $\sigma\left(\mathcal{F}_x \setminus \mathcal{H}_x\right)$ makes projection mappings onto the life spaces of individuals neither being ancestors or progeny of x measurable. Hence $\sigma\left(\mathcal{F}_x \setminus \mathcal{H}_x\right)$ will be independent of the σ-algebra $\sigma\left(\sigma(U_{\Pr x}), \mathcal{H}_x\right)$ and we can in this case freely choose to either condition on $\sigma(\mathcal{F}_x \setminus \mathcal{H}_x, \mathcal{H}_x) = \mathcal{F}_x$ or on $\mathcal{H}_x$, i.e. $\mathbb{P}_\tau(U_{\Pr x} \in A \,|\, \mathcal{F}_V) = \mathbb{P}_\tau(U_{\Pr x} \in A \,|\, \mathcal{H}_x)$ [34, Exercise 23.16].

[34]: Jacod et al. (2004), *Probability Essentials*

We now propose that $\mathbb{P}_{\tau_x}(A)$ is a version of this conditional probability, i.e for all $H \in \mathcal{G}_{\mathcal{H}_x}$ it should hold that

$$\mathbb{P}_\tau(F, U_{\Pr x} \in A) = \mathbb{P}_\tau[\,\mathbf{1}_F\, \mathbb{P}_{\tau_x}(A)]$$

Hence, let $H \in \mathcal{G}_{\mathcal{H}_x}$. If we consider $A = \bigtimes_{y \in \Pr x} A_y$ and

$$A_n = \bigtimes_{y \in x M_n} A_y \times \bigtimes_{y \notin x M_n} \Omega$$

11: The slightly weird naming of this set is intentional, it should be understood as $M_n \subset I$ that contains the elements y such that xy is one of the first n progeny of x.

where $xM_n \subset \Pr x$ contains the first n of x's progeny under $\leq$,[11] we see that $A_n \downarrow A$ as fewer and fewer dimensions of the set is identified with the full space Ω. By continuity from above we have that

$$\begin{aligned}\mathbb{P}_\tau(H, U_{\Pr x} \in A) &= \lim_{n\to\infty} \mathbb{P}_\tau(H, U_{\Pr x} \in A_n) \\ &= \lim_{n\to\infty} \mathbb{P}_\tau\left(H, U_{xM_n} \in \bigtimes_{y \in x M_n} A_y\right)\end{aligned}$$

As $H \in \mathcal{G}_{\mathcal{H}_x}$, we can write $H = U^{-1}_{\mathrm{An}\, x}[\bigtimes_{y \in \mathrm{An}\, x} A^0_y]$ where $A^0_y \in \mathcal{A}$ for each $y \in \mathrm{An}\, x$, so we have

$$\begin{aligned}\mathbb{P}_\tau\left(H, U_{M_n} \in \bigtimes_{y \in M_n} A_y\right) &= \mathbb{P}_\tau\left(U^{-1}_{\mathrm{An}\, x}\left[\bigtimes_{y \in \mathrm{An}\, x} A^0_y\right], U^{-1}_{xM_n}\left[\bigtimes_{y \in x M_n} A_y\right]\right) \\ &= \mathbb{P}_\tau\left(U^{-1}_{\mathrm{An}\, x \cup xM_n}\left[\bigtimes_{y \in \mathrm{An}\, x} A^0_y \times \bigtimes_{y \in x M_n} A_y\right]\right) = *\end{aligned}$$

We can then apply Proposition 2.3.1 noting that $\operatorname{An} x \cup xM_n$ contains all its ancestors. If we denote the largest element of xM_n by z we have,

$$\begin{aligned}
\ast &= \int_{A_0^0} \cdots \int_{A_{mx}^0} \int_{A_x} \cdots \int_{A_z} \prod_{y \in xM_n \cup \operatorname{An} x} \mathrm{d}\Lambda_{\tau_y(\omega^{\operatorname{An} y})}(\omega_y) \\
&= \int_{A_0^0} \cdots \int_{A_{mx}^0} \left(\int_{A_x} \cdots \int_{A_z} \prod_{y \in xM_n} \mathrm{d}\Lambda_{\tau_y(\omega^{\operatorname{An} y})}(\omega_y) \right) \prod_{y \in \operatorname{An} x} \mathrm{d}\Lambda_{\tau_y(\omega^{\operatorname{An} y})}(\omega_y) \\
&= \int_{A_0^0} \cdots \int_{A_{mx}^0} \left(\int_{A_x} \cdots \int_{A_z} \prod_{y \in M_n} \mathrm{d}\Lambda_{\tau_{xy}(\omega^{\operatorname{An} xy})}(\omega_{xy}) \right) \prod_{y \in \operatorname{An} x} \mathrm{d}\Lambda_{\tau_y(\omega^{\operatorname{An} y})}(\omega_y) \\
&= \int_{A_0^0} \cdots \int_{A_{mx}^0} \mathbb{P}_{\tau_x(\omega^{\operatorname{An} x})} \left(U_{M_n}^{-1} \left[\bigtimes_{y \in M_n} A_{xy} \right] \right) \prod_{y \in \operatorname{An} x} \mathrm{d}\Lambda_{\tau_y(\omega^{\operatorname{An} y})}(\omega_y) = \ast
\end{aligned}$$

where the last equality holds as we can consider τ_x as know under the $\operatorname{An} x$ integrals, and then τ_{xy} can be seen as a function of τ_x and $\{\omega_z : x \preceq z \prec y\}$ for all $y \in M_n$. Now the remaining integrals in the last expression can be seen as an integral with respect to $\mathbb{P}_\tau$ over $\bigtimes_{y \in \operatorname{An} x} A_y^0 \times \bigtimes_{y \notin \operatorname{An} x} \Omega =: U_{\operatorname{An} x}^{-1}[\bigtimes_{y \in \operatorname{An} x} A_y^0]$, i.e.

$$\begin{aligned}
\ast &= \int_{U_{\operatorname{An} x}^{-1}[\bigtimes_{y \in \operatorname{An} x} A_y^0]} \mathbb{P}_{\tau_x(\omega^{\operatorname{An} x})} \left(U_{M_n}^{-1} \left[\bigtimes_{y \in M_n} A_{xy} \right] \right) \mathrm{d}\mathbb{P}_\tau(\omega) \\
&= \mathbb{P}_\tau \left[\mathbf{1}_{U_{\operatorname{An} x}^{-1}[\bigtimes_{y \in \operatorname{An} x} A_y^0]} \mathbb{P}_{\tau_x} \left(U_{M_n} \in \bigtimes_{y \in M_n} A_{xy} \right) \right] \\
&= \mathbb{P}_\tau \left[\mathbf{1}_{U_{\operatorname{An} x}^{-1}[\bigtimes_{y \in \operatorname{An} x} A_y^0]} \mathbb{P}_{\tau_x}(A_n) \right]
\end{aligned}$$

As argued, by continuity from above we then have

$$\begin{aligned}
\mathbb{P}_\tau(H, U_{\operatorname{Pr} x} \in A) &= \lim_{n \to \infty} \mathbb{P}_\tau[\, \mathbf{1}_F \, \mathbb{P}_{\tau_x}(A_n)] \\
&= \mathbb{P}_\tau[\, \mathbf{1}_H \, \mathbb{P}_{\tau_x}(A)]
\end{aligned}$$

where the limit can be moved inside $\mathbb{P}_\tau$ using either monotone or dominated convergence, and $\mathbb{P}_{\tau_x}(A_n) \to \mathbb{P}_{\tau_x}(A)$, $\mathbb{P}_\tau$-almost surely by continuity from above.

Now the image measure $U_{\operatorname{Pr} x}(\mathbb{P}_\tau^H)$ where $\mathbb{P}_\tau^H(\cdot) := \mathbb{P}_\tau(H \cap \cdot)$ is the restricted measure,[12] is uniquely determined by its values on a $\cap$-stable generator containing $\Omega^{\operatorname{Pr} x}$, which certainly the following class is,

$$\left\{ \bigtimes_{y \in \operatorname{Pr} x} A_y : A_y \in \mathscr{A} \right\}.$$

12: $\mathbb{P}_\tau^H$ is no longer a probability measure unless $\mathbb{P}_\tau(H) = 1$, so neither is $U_{\operatorname{Pr} x}(\mathbb{P}_\tau^H)$.

Hence, for all $A \in \mathscr{A}^I$ we have

$$\mathbb{P}_\tau(H, U_{\operatorname{Pr} x} \in A) = U_{\operatorname{Pr} x}(\mathbb{P}_\tau^H) = \mathbb{P}_\tau[\, \mathbf{1}_H \, \mathbb{P}_{\tau_x}(A)]$$

or in other words, $\mathbb{P}_{\tau_x}(A)$ is a version of $\mathbb{P}_\tau(U_{\operatorname{Pr} x} \in A \mid \mathscr{H}_x)$, but then it is also a version of $\mathbb{P}_\tau(U_{\operatorname{Pr} x} \in A \mid \mathscr{F}_x)$. Using this back in Equation 2.3 we get,

$$\mathbb{P}_\tau(S_x \in B \times A \mid \mathscr{F}_x) = \mathbf{1}_{(\tau_x \in B)} \mathbb{P}_{\tau_x}(A) = \mathbb{P}_{\tau_x}(B \times A)$$

Here, the discussion in Remark 2.3.1 is of special importance, as $\mathbb{P}_{\tau_x(\omega)}$ for fixed $\omega \in \Omega^I$ is used as a measure on $\left(\Omega^I, \mathscr{A}^I\right)$ in the second to last expression, while used as a measure on $(\overline{\Omega}, \overline{\mathscr{A}})$ in the last.

Now another analogous uniqueness of measure argument finally shows that

$$\mathbb{P}_\tau\left(S_x \in \overline{A} \,\middle|\, \mathscr{F}_x\right) = \mathbb{P}_{\tau_x}\left(\overline{A}\right) \quad \mathbb{P}_\tau\text{-a.s.}$$

We can thus choose the regular conditional distribution kernel as

$$\kappa_x(\omega, \overline{A}) = \mathbb{P}_{\tau_x(\omega)}\left(\overline{A}\right)$$

for almost all $\omega \in \mathscr{A}^I$. This shows that $\mathbb{P}_{\tau_x}$ is a regular conditional distribution of S_x given $\mathscr{F}_x$.

We will initially argue for the branching property only for finite lines, so let $V = \{x_1, \dots, x_n\}$. Now, notice that $\mathscr{F}_V \subseteq \mathscr{F}_{x_1}$ as $V \preceq x_1$ which is the case as $x_1 \in V$, we can then use the tower property of conditional expectations to see that

$$\begin{aligned}\mathbb{P}_\tau\left[\prod_{i=1}^n \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathscr{F}_V\right] &= \mathbb{P}_\tau\left[\mathbb{P}_\tau\left[\prod_{i=1}^n \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathscr{F}_{x_1}\right] \,\middle|\, \mathscr{F}_V\right] \\ &= \mathbb{P}_\tau\left[\prod_{i=2}^n \varphi_{x_i}(S_{x_i})\, \mathbb{P}_\tau[\varphi_{x_1}(S_{x_1}) \mid \mathscr{F}_{x_1}] \,\middle|\, \mathscr{F}_V\right] = *\end{aligned}$$

as S_x is $\mathscr{F}_{x_1}$-measurable for all $x \in V \setminus \{x_1\}$. Now as we have shown the conditional distribution of S_{x_1} given $\mathscr{F}_{x_1}$ is $\mathbb{P}_{\tau_{x_1}}$, we can find the conditional expectation $\mathbb{P}_\tau[\varphi_{x_1}(S_{x_1}) \mid \mathscr{F}_{x_1}]$ almost surely by integrating φ_{x_1} with respect to this conditional distribution, i.e.

$$\begin{aligned}* &= \mathbb{P}_\tau\left[\prod_{i=2}^n \varphi_{x_i}(S_{x_i})\, \mathbb{P}_{\tau_{x_1}}[\varphi_{x_1}] \,\middle|\, \mathscr{F}_V\right] \\ &= \mathbb{P}_{\tau_{x_1}}[\varphi_{x_1}]\mathbb{P}_\tau\left[\prod_{i=2}^n \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathscr{F}_V\right]\end{aligned}$$

where we have used that $\mathbb{P}_{\tau_{x_1}}[\varphi_{x_1}]$ is $\sigma(\tau_{x_1}) \subset \mathscr{F}_V$ measurable, and can thus be pulled out of the conditional expectation. Repeating this argument n times, each time using the tower property with $\mathscr{F}_{x_i}$ for the next i, gives us the branching property for finite lines,

$$\mathbb{P}_\tau\left[\prod_{i=1}^n \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathscr{F}_V\right] = \prod_{i=1}^n \mathbb{P}_{\tau_{x_i}}[\varphi_{x_i}].$$

Now if V is not finite, we can always choose (in some structured way) a subset $V_n = \{x_1, \dots, x_n\} \subset V$ consisting of $n \in \mathbb{N}$ elements, and this will still be a line. Now notice that as $V_n \subset V_{n+1}$ so as discussed $\mathscr{F}_{V_n} \downarrow \mathscr{F}_V = \bigcap_{i=1}^\infty \mathscr{F}_{V_n}$ as $n \to \infty$. If we now consider the finite product $\varphi^m := \prod_{i=1}^m \varphi_{x_i}(S_{x_i})$ for some $m \in \mathbb{N}$, we see that the sequence of conditional expectations of ϕ^m with respect to the decreasing σ-algebras,

$$(\mathbb{P}_\tau[\varphi^m \mid \mathscr{F}_{V_n}])_{n \in \mathbb{N}}$$

forms a backwards martingale. So by the backwards martingale convergence theorem [47, Theorem 12.14], we know $\mathbb{P}_\tau$-*a.s.* that

[47]: Klenke (2020), *Probability Theory*

$$\mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_V] = \lim_{n\to\infty} \mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_n}] = \mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_m}] = \prod_{i=1}^m \mathbb{P}_{\tau_{x_i}}[\varphi_{x_i}]$$

The second last equality needs an elaboration. Take an $n \geq m$, we thus have $\mathcal{F}_{V_n} \subseteq \mathcal{F}_{V_m}$ and thus by the tower property and the branching property for finite lines,

$$\begin{aligned}
\mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_n}] &= \mathbb{P}_\tau[\mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_m}] \mid \mathcal{F}_{V_n}] && (2.4)\\
&= \mathbb{P}_\tau\left[\prod_{i=1}^m \mathbb{P}_{\tau_{x_i}}[\varphi_{x_i}] \,\middle|\, \mathcal{F}_{V_n}\right] && (2.5)\\
&= \prod_{i=1}^m \mathbb{P}_{\tau_{x_i}}[\varphi_{x_i}] && (2.6)
\end{aligned}$$

using in the last equation that τ_{x_i} is $\mathcal{F}_{V_n}$-measurable as we have chosen $n \geq m$. Hence for large enough n we have that $\mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_n}] = \mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_{V_m}]$ which then in particular must hold in the limit $n \to \infty$.

With the sequence of conditional expectations $(\mathbb{P}_\tau[\varphi^m \mid \mathcal{F}_V])_{m\in\mathbb{N}}$ at hand, all conditioned on the same σ-algebra, we note that $\varphi_x \leq 1$ for all $x \in V$ so in particular, we have that $\varphi^m \leq 1$. Using the dominated convergence theorem for conditional expectations (in the equality with an $*$) [47, Theorem 8.14 (viii)], we finally get the branching property for an infinite line,

$$\begin{aligned}
\prod_{x\in V} \mathbb{P}_{\tau_x}[\varphi_x] &= \lim_{m\to\infty} \prod_{i=1}^m \mathbb{P}_{\tau_{x_i}}[\varphi_{x_i}]\\
&= \lim_{m\to\infty} \mathbb{P}_\tau\left[\prod_{i=1}^m \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathcal{F}_{V_m}\right]\\
&= \lim_{m\to\infty} \mathbb{P}_\tau\left[\prod_{i=1}^m \varphi_{x_i}(S_{x_i}) \,\middle|\, \mathcal{F}_V\right]\\
&\overset{*}{=} \mathbb{P}_\tau\left[\prod_{x\in V} \varphi_x(S_x) \,\middle|\, \mathcal{F}_V\right]
\end{aligned}$$

Conditional independence of $\{S_x : x \in V\}$ follows by looking at a collection of events $\{A_x : x \in V\} \subset \mathcal{A}^I$ and choosing $\varphi_x = \mathbf{1}_{A_x}$ for each $x \in V$,

$$\begin{aligned}
\mathbb{P}_\tau\left(\bigcap_{x\in V}(S_x \in A_x) \,\middle|\, \mathcal{F}_V\right) &= \mathbb{P}_\tau\left[\prod_{x\in V} \mathbf{1}_{A_x}(S_x) \,\middle|\, \mathcal{F}_V\right]\\
&= \prod_{x\in V} \mathbb{P}_{\tau_x}[\,\mathbf{1}_{A_x}]\\
&= \prod_{x\in V} \mathbb{P}_{\tau_x}(A_x)\\
&= \prod_{x\in V} \mathbb{P}_\tau(S_x \in A_x \mid \mathcal{F}_V)
\end{aligned}$$

where the last equality follows from S_x being conditionally distributed as $\mathbb{P}_{\tau_x}$ given F_V for each x. □

Intuitively, the branching property should be relatively self-evident as a kind of Markovian structure on the genealogical tree, and it is certainly also stated as such in several papers on Bellman-Harris processes and generalizations hereof, see for example [12, 13, 37, 39]. On the other hand, the proof is very lengthy and technical, so in practice, these are not found commonly in the literature. Our statement is heavily inspired by the one in [37], the proof is completely rewritten and much elaborated, but essentially contains the same arguments.

[12]: Chauvin (1986), 'Arbres et processus de Bellman-Harris'
[13]: Chauvin (1991), 'Product Martingales and Stopping Lines for Branching Brownian Motion'
[37]: Jagers (1989), 'General branching processes as Markov fields'
[39]: Jagers et al. (1984), 'The Growth and Composition of Branching Populations'

2.4.3 A strong branching property

If the line on whose non-progeny we want to condition on is not known beforehand but is itself a random set, the branching property above is no help. Instead, we want to have an analog to the strong Markov property. The strong Markov property allows us to condition on the past of a random time that is assumed to be *optional*, in the sense that we at any time can determine whether or not the random time has passed, only using the information available at that time. Hence we need an analog to optional stopping times but for lines, i.e. we introduce *optional lines*. We will not put too much emphasis on such optional lines or the strong branching property pertaining thereto as they are merely used to simplify a few proofs, which could also have been done using the simple branching property, in particular, we provide no proof of the strong branching property. For a more thorough treatment see [37].

[37]: Jagers (1989), 'General branching processes as Markov fields'

Consider a random line of individuals in a population started at time $\tau \in \mathbb{R}$, $\mathscr{V} : \left(\Omega^I, \mathscr{A}^I\right) \to (2^I, 2^{2^I})$. We say that $\mathscr{V}$ is an *optional line* if $(\mathscr{V} \preceq V) \in \mathscr{F}_V$, i.e. knowing the lives the non-progeny of V we will know whether V stems from $\mathscr{V}$ or not. In particular, can $\mathscr{V}$ not depend on any future state of the population.

Our primary example of an optional line is the set of realized individuals in some generation $n \in \mathbb{N}_0$ of a population started at time $\tau \in \mathbb{R}$. It has been noted already that $\mathscr{R}_n(\omega)$ is a line for all $\omega \in \Omega^I$. The optionality of it is asserted by writing it as $\mathscr{R}_n = \{x \in \mathbb{N}^n : \tilde{\tau}_x < \infty\}$ and realizing that if we know the lives of the ancestors of $\mathscr{R}_n$ then we know $\tilde{\tau}_x$ for each $x \in \mathbb{N}^n$ and hence $(\mathscr{R}_n \preceq L) \in \mathscr{H}_L \subset \mathscr{F}_L$.

To state the strong branching property we need to define the *optional line non-progeny σ-algebra*, which is done in complete analog to classical optional time σ-algebras. That is, for an optional line $\mathscr{V}$ we let

$$\mathscr{F}_{\mathscr{V}} := \left\{A \in \mathscr{A}^I : A \cap (\mathscr{V} \preceq V) \in \mathscr{F}_V \text{ for all lines } V \subset I\right\}$$

and can then state the strong branching property as,

Theorem 2.4.2 (The strong branching property) *Let $\tau \in \mathbb{R}$. For an optional stopping line $\mathscr{V}$ and functions $\varphi_x \colon (\overline{\Omega}, \overline{\mathscr{A}}) \to ([0,1], \mathscr{B}_{[0,1]})$ for all $x \in I$, it holds that*

$$\mathbb{P}_\tau\left[\prod_{x\in\mathscr{V}} \varphi_x(S_x) \middle| \mathscr{F}_{\mathscr{V}}\right] = \prod_{x\in\mathscr{V}} \mathbb{P}_{\tau_x}[\varphi_x].$$

A reduced time-varying Bellman-Harris process

3

3.1 Random characteristics . 30
3.2 Defining the reduced process 32
3.3 Distribution of the reduced branching process 33
3.4 The reduced binary branching process 41
3.5 Death of the least common ancestor 47

Now that we have a solid foundation for working with branching populations, we are ready to define processes on the time-population space from which we can study the dynamics of the branching population. In this chapter, we will use the framework of branching processes counted by a *random characteristic* to define the *classical branching process* and the *reduced branching process* as continuous time processes in our population. The classical branching process counts the number of individuals alive at any given time whereas the reduced branching process only counts the number of individuals who are alive at any given time who *also* has progeny alive at some specified future time. Neither process actively contains information about the genealogical structure of the population over time, only the number of individuals fulfilling the condition to be counted, but certain properties of the reduced process let us gain partial information about the genealogy of individuals.

The classical branching process is certainly interesting in itself, but will only serve as an intermediate step to gaining knowledge about the reduced process, as this is the process of primary interest when drawing inspiration from phylogenetics. Note in particular that the reduced branching process is a model of the number of ancestors of currently extant species through time, thus if we had an observation of this, our model would in principle allow for estimation of the population's lifetime and offspring distributions, relating to the interesting evolutionary dynamics in phylogenetics.

Besides defining the processes of interest, this chapter's primary aim is to characterize the distribution of the reduced process, with a focus on the binary population case, which simplifies the results a bit and which, in phylogenetics, will often be the model of primary interest. With a characterization of the distribution of the reduced process, we obtain expressions for the mean of the process, for a few interesting probabilities and we finally characterize the distribution of the time since least common ancestor.

This chapter contains several definitions and results which to the best knowledge of the author are novel, of most importance are

- The definition of a reduced *time-varying* Bellman-Harris process using random characteristics
- The integral equation for the generating function of the reduced time-varying Bellman-Harris process
- The characterization of the distribution of the time since least common ancestor conditioned on non-extinction of the process

The section on random characteristics is primarily based on [36, 39].

[36]: Jagers (1975), *Branching processes with biological applications*
[39]: Jagers et al. (1984), 'The Growth and Composition of Branching Populations'

3.1 Random characteristics

[28]: Harris (1963), *The theory of branching processes*

The classical definition of a branching process (as for example that in [28]) is the process counting the number of individuals in a branching population being alive, or being alive and under a given age, over time. A natural extension is to count individuals in the population having some more general characteristics, i.e. we could consider a random element on $(\overline{\Omega}, \overline{\mathcal{A}})$ which in a certain sense scores each individual according to some characteristic. Adding up these scores for each individual in the population over time defines a stochastic process, which we will call the *branching process counted with random characteristic*.

Consider a stochastic process $\chi := \{\chi(a)\}_{a\in\mathbb{R}}$ on $(\overline{\Omega}, \overline{\mathcal{A}})$ with values in $(\{0,1\}, 2^{\{0,1\}})$, such that $\chi(a) = 0$ a.s. for $a < 0$. We will call such a process a *random characteristic*[1] and let it represent a score of an individual of age a, thus we associate each individual x in I with its own random characteristic $\chi_x := \chi(S_x)$, called the *individual random characteristic*. Remark that the random characteristic is defined on the full time-population space, not on the life space, thus the characteristic might not only depend on the life of an individual but also on its progeny. Here we have restricted the score to be either 0 or 1, we will see that this corresponds to the score simply being whether or not to count an individual when it is of age a.

1: Our definition of a random characteristic is very limited in comparison to their general definition. In particular, is it possible to let them take values in much more intricate spaces, see e.g. [36].

The branching process counted by χ is then the stochastic process $Z^\chi := \{Z^\chi(t)\}_{t\in\mathbb{R}}$ on $(\overline{\Omega}, \overline{\mathcal{A}})$ given for $t \in \mathbb{R}$, $\omega \in \overline{\Omega}$,

$$Z^\chi(t,\omega) := \sum_{x\in I} \chi(t - \tilde{\tau}_x(\omega), S_x(\omega)) = \sum_{x\in I} \chi_x(t - \tilde{\tau}_x(\omega), \omega) \tag{3.1}$$

as $t - \tilde{\tau}_x(\omega)$ is certainly the age of an individual $x \in I$ at time t, if we convene on non-realized individuals having age $-\infty$ for all t. Note in particular that this formulation never counts non-realized individuals or individuals that are not yet born as their characteristics are 0 by assumption, we could hence equivalently write

$$Z^\chi(t,\omega) = \sum_{x\in\mathcal{R}_1(\omega)} \chi(t - \tau_x(\omega), S_x(\omega)) = \sum_{x\in\mathcal{R}_1(\omega)} \chi_x(t - \tau_x(\omega), \omega) \tag{3.2}$$

where it is no longer necessary to distinguish between $\tilde{\tau}_x$ and τ_x.

Example 3.1.1 The classical branching process, counting living individuals can be recovered by counting individuals whose age is less than their lifetime, i.e. individuals who are alive,[2]

$$\chi(a) = \mathbf{1}_{[0,L)}(a)$$

As this is such an important example we will simply denote the process counted with this characteristic by $Z := \{Z(t)\}_{t\in\mathbb{R}} := \{Z^{\mathbf{1}_{[0,L)}}(t)\}_{t\in\mathbb{R}}$. By construction this process counts the number of individuals in $\mathcal{Z}_t$ over time, i.e.

$$Z(t,\omega) = \#(\mathcal{Z}_t(\omega))$$

Ignoring whether or not an individual has died, i.e. counting with the

2: Remark that an individual "dies instantaneously", that is, we do not consider an individual to be alive when it has age L.

characteristic

$$\chi(a) = \mathbf{1}_{[0,\infty)}(a)$$

gives us the process counting the cumulative number of realized individuals in the process. Note that this last characteristic does not itself depend on ω, i.e. it is non-random, but when we compose it with the random age, it of course becomes a random variable.

3.1.1 Principle of first generation

A foundational lemma for most work on branching processes is the *fundamental equation* which for Bellman-Harris processes is usually called the *principle of first generation*. This equation, together with the branching property, lays the way for the renewal-type governing equations, which we will see characterizes the full distributions of the process.

The lemma, which is simply a regrouping of the terms in Equations 3.1 and 3.2, states that if we for each individual $x \in I$ define the process $Z_x^\chi = Z^\chi(S_x)$ as the progeny process of x, we can consider the original process Z^χ as the contribution from the ancestor 0 and the contribution from each first generation progeny process,[3]

3: Rearranging Equations 3.1 and 3.2 according to the first generation is just a simple, and for our purposes useful, choice, but we could do just the same with any other generation, or more generally as in [37] with any stopping line M for which we either have $x > M$ or $\Pr x \cap M \neq \emptyset$ for all $x \in I$, i.e. a line which *covers* the whole population.

Lemma 3.1.1 (The principle of first generation) *Let Z^χ be a branching process counted with the random characteristic χ, then*

$$\begin{aligned} Z^\chi(t) &= \chi_0(t - \tau_0) + \sum_{x \in \mathbb{N}} Z_x^\chi(t) \\ &= \chi_0(t - \tau_0) + \sum_{x \in \mathscr{R}_1} Z_x^\chi(t) \end{aligned}$$

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

Proof. Let $t \in \mathbb{R}$, $\omega \in \overline{\Omega}$ and consider Equation 3.1, the lemma follows as

$$\begin{aligned} Z^\chi(t, \omega) &= \sum_{x \in I} \chi_x(t - \tilde{\tau}_x(\omega), \omega) \\ &= \chi_0(t - \tilde{\tau}_0(\omega), \omega) + \sum_{x \in I \setminus \{0\}} \chi_x(t - \tilde{\tau}_x(\omega), \omega) \\ &= \chi_0(t - \tilde{\tau}_0(\omega), \omega) + \sum_{x \in \mathbb{N}} \sum_{y \in I} \chi_{xy}(t - \tilde{\tau}_{xy}(\omega), \omega) \\ &= \chi_0(t - \tilde{\tau}_0(\omega), \omega) + \sum_{x \in \mathbb{N}} Z_x^\chi(t, \omega) \end{aligned}$$

where we have used that any individual in $I \setminus \{0\}$ is a progeny of some first-generation individual and the definition of Z_x^χ. The proof of the second part is completely analogous, just working with Equation 3.2 instead. □

With this *self-similarity* property at hand, it is, at least intuitively, evident that we can formulate the distribution of the full process by some combination of the distributions of the sub-processes. However, the distribution of a sum of random variables is not a trivial object to work with, especially not when the number of terms in the sum is also random,

but this can be remedied by considering the generating function of the distribution, and hence for the process. We will describe in detail such a formulation for the reduced branching process, but first, we need to define the process rigorously.

3.2 Defining the reduced process

A particular choice of random characteristic, which defines the reduced branching process, will form the backbone of the rest of this thesis. Consider a branching population whose extant members are *observed* at fixed time T after the origin-time of the population, these individuals make up the set $\mathfrak{X}_T$. We might be interested in the genealogical and temporal structure of the ancestors of these individuals, i.e. the random Neveu tree $\mathrm{An}\,\mathfrak{X}_T$[4] and the branching process counting the number of living individuals of that tree through time, or as we shall later see, a coupling of these. This Neveu tree is closely related to a *reduced phylogenetic tree* from which we draw much of our inspiration, and the mean of the branching process will correspond to the *lineages-through-time* function, but for now, let's stay in the abstract population world only taking inspiration from this application. Definitions, and further analyses, of the reduced non-time-varying Bellman-Harris process have previously been made [71, 86], but these have not used the unifying framework of random characteristics, and has not considered any time-variability.

4: Not to be confused with Υ_T which is a superset of $\mathrm{An}\,\mathfrak{X}_T$.

[71]: Sagitov (1986), 'Reduced Critical Bellman–Harris Branching Process with Several Types of Particles'
[86]: Vatutin (1979), 'Distance to the nearest common ancestor in bellman-harris branching processes'

The random characteristic for the reduced branching process is more complex than those considered in Example 3.1.1 in that it is not determined by the life of each individual by itself, but rather by the lives of all the individual's progeny. Recall that the classical branching process, counting the number of individuals alive over time, is denoted by $Z := \{Z(t)\}_{t\in\mathbb{R}}$. Then the *reduced characteristic*[5] at time T can be defined generally as

$$\chi^T(a,\omega) := \mathbf{1}_{[0,L(\omega))}(a)\ \mathbf{1}_{(Z(T)>0)}(\omega)$$

for $a \in \mathbb{R}$ and $\omega \in \overline{\Omega}$. Thus for each x in I, the individual characteristic is

$$\chi^T_x(a,\omega) := \chi^T(a, S_x(\omega)) = \mathbf{1}_{[0,L_x(\omega))}(a)\ \mathbf{1}_{(Z_x(T)>0)}(\omega)$$

An individual is then counted at time t *if it is alive* at time t *and it has extant progeny* at the observation, this will surely eradicate any individuals whose lineage has gone extinct before the population is observed.

5: There is nothing *reduced* about the characteristic itself, but we name it as such as it defines the reduced process.

The *reduced branching process*, which we will denote $Z^T := \{Z^T(t)\}_{t\in\mathbb{R}}$, is now simply defined as the branching process counted by χ^T,

$$Z^T(t,\omega) := \sum_{x\in I} \chi^T_x(t-\tilde{\tau}_x(\omega),\omega) = \sum_{x\in\mathcal{R}(\omega)} \chi^T_x(t-\tau_x(\omega),\omega). \tag{3.3}$$

Remark 3.2.1 The reduced process is intrinsically connected with the classical branching process, and if we know the classical branching, we can always gain the reduced process. The other way around is clearly not the case, but a central aim in studying the reduced process

is to gain some insight into the dynamics of the population and hence the classical process. Note in particular that we define the reduced process from the exact same background measures, so if we fit the reduced process to some appropriate empirical observation, we can study the classical process based on these estimated distributions, and gain insight into the full population dynamics.

The reduced branching process satisfies, on top of being a branching process, some almost defining features, which will be appropriate to collect in a proposition.

Proposition 3.2.1 *Let* $\varpi \in \overline{\Omega}$

1. *The function* $t \mapsto Z^T(t, \varpi)$ *is weakly increasing for* $t \in [\tau_0(\varpi), \infty)$
2. *If* $Z^T(s, \varpi) = 0$ *for some* $s \in [\tau_0(\varpi), \infty)$ *then* $Z^T(t, \varpi) = 0$ *for all* $t \in [\tau_0(\varpi), \infty)$.
3. $Z^t(t, \varpi) = Z(t, \varpi)$ *for all* $t \in [\tau_0(\varpi), \infty)$
4. *If* $N(t, \varpi) > 0$ *for all* $t \in \mathbb{R}$ *then* $Z^T(t, \varpi) = Z(t, \varpi)$ *for all* $t \in \mathbb{R}$ *(this also holds for* $\mathbb{P}_\tau$*-almost all* ϖ *with given* $\tau \in \mathbb{R}$*).*

The proofs of these statements follows easily from Equation 3.3.

3.3 Distribution of the reduced branching process

To find the distribution of the reduced branching process, we have to be more explicit about the probability kernels $(\Lambda_\tau)_{\tau\in\mathbb{R}}$ and thus $(\mathbb{P}_\tau)_{\tau\in\mathbb{R}}$, and in particular the image measures under the lifetime and offspring process. For fixed $\tau \in \mathbb{R}$, we consider Λ_τ such that L and N are independent, and such that the distribution of L has m_+-density[6] g_τ, that we assume to be continuous, and that the distribution of $N(t)$ has $c_{\mathbb{N}_0}$-density[7] $\nu \mapsto q_\nu(t)$ for all $t \in \mathbb{R}$, that is $(L, N)(\Lambda_\tau) = L(\Lambda_\tau) \otimes N(\Lambda_\tau)$ and

$$L(\Lambda_\tau)(B) = \Lambda_\tau(L \in B) = \int_B g_\tau \, \mathrm{d}m_+, \qquad \text{for all } B \in \mathcal{B}_+$$
$$N(t)(\Lambda_\tau)(\{\nu\}) = \Lambda_\tau(N(t) = \nu) = q_\nu(t), \qquad \text{for all } \nu \in \mathbb{N}_0.$$

6: By m we denote the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$. With a subscript such as here is it the Lebesgue measure on the appropriately restricted real Borel space.

7: By $c_{\mathbb{N}_0}$ we denote the counting measure on $(\mathbb{N}_0, 2^{\mathbb{N}_0}$. Likewise c_M would denote the counting measure on some discrete measurable space $(M, 2^M)$

Remark 3.3.1 To characterize the distribution of some of our variables taking values in $(\mathbb{N}_0, 2^{\mathbb{N}_0})$, we will often turn to *generating functions*. For completeness, we here include a definition of such a function and a few important properties.

Let us in general consider a random variable X on some probability space with probability measure $\mathbb{P}$ which takes values in $(\mathbb{N}_0, 2^{\mathbb{N}_0})$, the *(probability) generating function* of $X(\mathbb{P})$ (or simply of X) is the map $\psi_X \colon [0, 1] \to [0, 1]$ given by

$$\psi_X(s) := \mathbb{P}\left[s^X\right] = \sum_{k\in\mathbb{N}_0} s^x \mathbb{P}(X = k)$$

where we convene on $0^0 = 1$. Notice that ψ_X is a power series that converges at least on $[0,1]$ as the point probabilities of $X(\mathbb{P})$ must sum to one, thus ψ_X actually exists for all $s \in [0,1]$. Another way to realize this is that $s^X \leq 1$ for all $s \in [0,1]$ and is thus integrable.

Several results about generating functions stem from basic results about power series. A few, which will be essential for our work, are given here,

1. ψ_X is continuous on $[0,1]$ and infinitely often differentiable on $[0,1)$
2. The distribution $X(\mathbb{P})$ is completely characterized by ψ_X through its power series coefficients

$$X(\mathbb{P})(\{k\}) = \mathbb{P}(X = k) = \frac{\frac{\partial^k}{\partial s^k}\psi_X(0)}{k!}$$

3. ψ_X has a (possibly infinite) left-derivative at 1, being the expectation of X,

$$\frac{\partial}{\partial s}\psi_X(1^-) = \sum_{k\in\mathbb{N}} k\mathbb{P}(X = k) = \mathbb{P}[X]$$

Proofs of these statements and further properties of generating functions can be found e.g. in [47, Theorem 3.2].

[47]: Klenke (2020), *Probability Theory*

Now as $N(t)$ takes values in $\mathbb{N}_0$ for each $t \in \mathbb{R}$, its distribution is completely determined by the point probabilities here, but we could also characterize its distribution through its generating function at each time $t \in \mathbb{R}$, so let $\zeta(\cdot;t)\colon [0,1] \to [0,1]$ be its generating function[8], i.e.

$$\zeta(s;t) := \Lambda_\tau[s^{N(t)}] = \sum_{\nu\in\mathbb{N}_0} s^\nu \Lambda_\tau(N(t) = \nu)$$

for $s \in [0,1]$ and all $t \in \mathbb{R}$. We will keep these assumptions throughout the remainder of the thesis, restricting them further when necessary in special cases.

8: Note that even though notationally the distribution $N(t)(\Lambda_\tau)$ looks like it is dependent on τ it is not, as $q_\nu(t)$ is not dependent on τ. Consequentially, ζ is not indexed by τ.

Example 3.3.1 (The Yule and birth-death processes) Two particularly interesting examples of lifetime and offspring distributions are the ones that simplify our process into the classical frameworks of the Yule process and the birth-death process. Both cases are considered time-homogeneous, but their time-varying counterparts are also special cases of our processes.

For the Yule process with birth rate $\beta > 0$, we consider exponentially distributed lifetimes, i.e. $g_\tau(\ell) = \beta e^{-\beta\ell}$, and let all individuals have exactly two offspring at the end of their lives, i.e. $q_2(t) = 1$. Note that in this case, no lineages will ever go extinct, so the reduced process Z^T reduces to the classical branching process Z, this is the result of Proposition 3.2.1 statement 4.

For the birth-death model with birth rate $\beta > 0$ and death rate $\delta > 0$[9], we now let an individual either have zero or two offspring at the end of their life, corresponding to a death or a birth, i.e. $q_2(t) = 1 - q_0(t) = q$ for some $q \in [0,1]$. The lifetime will still be exponentially distributed,

9: β and δ are not traditional names for the birth and death rates, but the more common λ and μ are used for other purposes in this thesis.

but with rate parameter $\beta\delta$ as the life of an individual ends when it either dies or gives birth, i.e. it is the minimum of two exponentially distributed variables. In contrast to the Yule process, extinctions of lineages can happen in the birth-death model, so the reduced process is not identical to the classical process. The reduced process of a birth-death model[10] has previously been studied as a model for reconstructed phylogenetic trees and for the number of lineages through time.

10: This is the same as e.g. [21, 59] calls the *reconstructed process*.

With these assumptions at hand, the principle of first generation and the branching property, we will now obtain an expression for the generating function of Z^T, and thus, in particular, have its distribution characterized. Hence let $\psi_\tau^T(\cdot;t)\colon [0,1] \to [0,1]$ for each set of times $\tau \leq t \leq T$ be the generating function of the reduced branching process originating at time τ and observed at time T, i.e.

$$\psi_\tau^T(s;t) := \mathbb{P}_\tau\left[s^{Z^T(t)}\right] = \sum_{z\in\mathbb{N}_0} s^z \mathbb{P}_\tau\Big(Z^T(t) = z\Big)$$

for $\tau \leq t \leq T$ and $s \in [0,1]$.

The main idea here is to split up the expectation based on whether or not the ancestral individual has died at time t. If she has not died yet, the process must either take value one or zero, depending on whether or not it eventually goes extinct before observation, this reduces the expectation into a combination of probabilities of survival until time t and the population going extinct. The probability of extinction can be found recursively using the principle of first generation and the branching property as the probability of extinction of all the ancestor's potential progeny processes, which all originate at the time of death of the ancestor, which happens sometime after t according to the distribution $L(\Lambda_\tau)$. The first two terms of the generating functions integral equation below are given by this argument.

If the ancestor has died before time t, she is not counted by the reduced characteristic, but we can consider each of her potential progeny processes, all originating at her time of death, sometime between time τ and time t, according to the distribution $L(\Lambda_\tau)$, and count them and their progeny just as the original process. Again this uses the principle of first generation and the branching property. The last term of the integral equation for the generating function is given by this argument.

Theorem 3.3.1 *Let $\tau \in \mathbb{R}$. The generating function $\psi_\tau^T(s;t) := \mathbb{P}_\tau\left[s^{Z^T(t)}\right]$ satisfies*

$$\begin{aligned}\psi_\tau^T(s;t) = {} & s\,(1 - G_\tau(t-\tau)) + \\ & (1-s)\int_{(t-\tau,T-\tau]} \zeta\left(\mathbb{P}_{\tau+\ell}(Z(T)=0);\tau+\ell\right)\, g_\tau(\ell)\,\mathrm{d}m(\ell) + \\ & \int_{[0,t-\tau]} \zeta(\psi_{\tau+\ell}^T(s;t);\tau+\ell)\, g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

for $s \in [0,1]$ and $\tau \leq t \leq T$.

Proof. We start by splitting up the expectation based on the value of whether the ancestral individual has died or not, i.e. we split up the

expectation based on the value of $\tau_0 + L_0$, which under $\mathbb{P}_\tau$ is almost surely equal to $\tau + L_0$,

$$\begin{aligned}\psi_\tau^T(s;t) &= \mathbb{P}_\tau\left[s^{Z^T(t)}\right] \\ &= \mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0\le t)}\right] \\ &= \underbrace{\mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0>t)}\right]}_{=*} + \underbrace{\mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0\le t)}\right]}_{=**} \end{aligned} \tag{3.4}$$

We will treat each term here separately, starting with the former which is working on the event that the ancestral individual has not yet died by time $t \ge \tau$. We start by using the definition of a branching process counted with random characteristic and singling out the ancestor of the sum

$$\mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0>t)}\right] = \mathbb{P}_\tau\left[s^{\chi_0(t-\tau)+\sum_{x\in I\setminus\{0\}}\chi_x(t-\tilde{\tau}_x)}\,\mathbf{1}_{(\tau+L_0>t)}\right] = *$$

Working under $(\tau + L_0 > t)$ we have that $\tilde{\tau}_x \ge \tau + L_0 > t$ a.s. for all $x \in I \setminus \{0\}$ so $t - \tilde{\tau}_x < 0$ and we have that $\chi_x(t - \tilde{\tau}_x) = 0$,

$$\begin{aligned} * &= \mathbb{P}_\tau\left[s^{\chi_0(t-\tau)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= \mathbb{P}_\tau\left[s^{\mathbf{1}_{[0,L_0)}(t-\tau)\,\mathbf{1}_{(Z(T)>0)}}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= \mathbb{P}_\tau\left[s^{\mathbf{1}_{(Z(T)>0)}}\,\mathbf{1}_{(\tau+L_0>t)}\right] = * \end{aligned}$$

where we in the last line have used that under $(\tau + L_0 > t)$ we clearly have $\mathbf{1}_{[0,L_0)}(t - \tau) = 1$ a.s.

This expectation can again be split up into two, now based on the event $(Z(T) > 0)$ and its complement $(Z(T) = 0)$, under which the indicator in the exponent is a.s. determined. The following long sequence of basic algebra and working with indicator functions give us two nice and workable terms,

$$\begin{aligned} * &= \mathbb{P}_\tau\left[s^{\mathbf{1}_{(Z(T)>0)}}\,\mathbf{1}_{(Z(T)>0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[s^{\mathbf{1}_{(Z(T)>0)}}\,\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= \mathbb{P}_\tau\left[s^{1}\,\mathbf{1}_{(Z(T)>0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[s^{0}\,\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= s\mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)>0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= s\mathbb{P}_\tau\left[\left(1-\mathbf{1}_{(Z(T)=0)}\right)\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= s\mathbb{P}_\tau\left[\mathbf{1}_{(\tau+L_0>t)}\right] - s\mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] + \mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= s\mathbb{P}_\tau\left[\mathbf{1}_{(\tau+L_0>t)}\right] + (1-s)\mathbb{P}_\tau\left[\mathbf{1}_{(Z(T)=0)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= \underbrace{s\,\mathbb{P}_\tau\left(\tau+L_0>t\right)}_{=*_1} + \underbrace{(1-s)\,\mathbb{P}_\tau\left(Z(T)=0,\tau+L_0>t\right)}_{=*_2} = * \end{aligned}$$

note that these probabilities don't depend on the reduced process anymore. We start by looking at the first expectation,

$$\begin{aligned} *_1 &= \mathbb{P}_\tau(L_0 > t-\tau) \\ &= L_0(\mathbb{P}_\tau)\left((t-\tau,\infty)\right) \\ &= L(\Lambda_\tau)\left((t-\tau,\infty)\right) \\ &= 1 - G_\tau(t-\tau) \end{aligned}$$

where G_τ is the distribution function of $L(\Lambda_\tau)$, i.e. $1 - G_\tau(t-\tau)$ is the probability that an individual born at time τ is still alive at time t.

More work is required for the second expectation. We start by using the principle of first generation on $\{Z(t)\}_{t\in\mathbb{R}}$, which is the process counted with the random characteristic $\chi(a) = \mathbf{1}_{[0,L)}(a)$,

$$\begin{aligned}
\ast_2 &= \mathbb{P}_\tau\left(\mathbf{1}_{[0,L_0)}(T-\tau) + \sum_{x\in\mathscr{R}_1} Z_x(T) = 0,\ \tau + L_0 > t\right) \\
&= \mathbb{P}_\tau\left((T-\tau \geq L_0) \cap \bigcap_{x\in\mathscr{R}_1} (Z_x(T)=0) \cap (\tau + L_0 > t)\right) \\
&= \mathbb{P}_\tau\left((L_0 \in (t-\tau, T-\tau]) \cap \bigcap_{x\in\mathscr{R}_1} (Z_x(T)=0)\right) = \ast_2
\end{aligned}$$

Reformulating as expectations and using the law of total expectation, conditioning on $\mathscr{F}_{\mathscr{R}_1}$, we get

$$\begin{aligned}
\ast_2 &= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0) \prod_{x\in\mathscr{R}_1} \mathbf{1}_{(Z_x(T)=0)}\right] \\
&= \mathbb{P}_\tau\left[\mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0) \prod_{x\in\mathscr{R}_1} \mathbf{1}_{(Z(T,S_x)=0)} \,\middle|\, \mathscr{F}_{\mathscr{R}_1}\right]\right] = \ast_2
\end{aligned}$$

where we see that L_0 is $\mathscr{F}_{\mathscr{R}_1}$-measurable, and can be pulled out of the conditional expectation, leaving us with an expression on which we can use the strong branching property[11] as $\omega \mapsto \mathbf{1}_{(Z_x(T,\omega)=0)}$, is clearly bounded and measurable,

11: Here we utilize the strong branching property, but this is only as a shortcut. One could instead argue by conditioning on the whole first generation $\mathbb{N}$ and then arguing that we only need to consider the finite number of children of the ancestor.

$$\begin{aligned}
\ast_2 &= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0)\, \mathbb{P}_\tau\left[\prod_{x\in\mathscr{R}_1} \mathbf{1}_{(Z(T,S_x)=0)} \,\middle|\, \mathscr{F}_{\mathscr{R}_1}\right]\right] \\
&= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0) \prod_{x\in\mathscr{R}_1} \mathbb{P}_{\tau_x}\left[\mathbf{1}_{(Z(T)=0)}\right]\right] \\
&= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0) \prod_{x\in\mathscr{R}_1} \mathbb{P}_{\tau+L_0}[Z(T)=0]\right] \\
&= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0)\, \mathbb{P}_{\tau+L_0}[Z(T)=0]^{\#\mathscr{R}_1}\right] \\
&= \mathbb{P}_\tau\left[\mathbf{1}_{(t-\tau,T-\tau]}(L_0)\, \mathbb{P}_{\tau+L_0}[Z(T)=0]^{N_0(\tau+L_0)}\right] = \ast_2
\end{aligned}$$

in the last three lines we use that $\mathscr{R}_1$ is finite $\mathbb{P}_\tau$-a.s. as $q_\infty(t) = 0$ for all $t \geq \tau$ and that $\tau_x = \tau + L_0$ for all $x \in \mathscr{R}_1$.

The integrand under $\mathbb{P}_\tau$ is now clearly a (L_0, N_0)-measurable non-negative real variable, so we can use the abstract change-of-variable [47, Theorem 4.10] to see

[47]: Klenke (2020), *Probability Theory*

$$\begin{aligned}
\ast_2 &= \int \mathbf{1}_{(t-\tau,T-\tau]}(L_0)\, \mathbb{P}_{\tau+L_0}[Z(T)=0]^{N_0(\tau+L_0)}\, \mathrm{d}\mathbb{P}_\tau \\
&= \int \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathbb{P}_{\tau+\ell}[Z(T)=0]^{\eta(\tau+\ell)}\, \mathrm{d}(L_0, N_0)(\mathbb{P}_\tau)(\ell,\eta) = \ast_2
\end{aligned}$$

so let us now examine the image measure $(L_0, N_0)(\mathbb{P}_\tau)$.

We start by noticing that $(L_0, N_0) = (L, N) \circ U_0$ where L and N are the random elements defined directly on the life space. This allows us to express the image measure $(L_0, N_0)(\mathbb{P}_\tau)$ through the life law Λ_τ. Let $B \in \mathcal{B}_+ \otimes \sigma(N)$, we then have

$$\begin{aligned}(L_0, N_0)(\mathbb{P}_\tau) &= \mathbb{P}_\tau\Big(((L, N) \circ U_0)^{-1}(B)\Big) \\ &= \mathbb{P}_\tau\left(U_0^{-1}\left((L, N)^{-1}(B)\right)\right) \\ &= \int_{(L,N)^{-1}(B)} \mathrm{d}\Lambda_\tau(\omega) \\ &= \Lambda_\tau\left((L, N)^{-1}(B)\right) \\ &= (L, N)(\Lambda_\tau)(B)\end{aligned}$$

where we have used Proposition 2.3.1, having in mind that 0 has no ancestors, and have thus showed that $(L_0, N_0)(\mathbb{P}_\tau) = (L, N)(\Lambda_\tau)$. As we assume that L and N are independent, their joint image measure is a product measure $(L, N)(\Lambda_\tau) = L(\Lambda_\tau) \otimes N(\Lambda_\tau)$. These facts and Tonelli's theorem then let us express the integral as,

$$\begin{aligned}*_2 &= \int \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\eta(\tau+\ell)}\, \mathrm{d}\left(L(\Lambda_\tau) \otimes N(\Lambda_\tau)\right)(\ell, \eta) \\ &= \int\int \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\eta(\tau+\ell)}\, \mathrm{d}N(\Lambda_\tau)(\eta)\, \mathrm{d}L(\Lambda_\tau)(\ell) \\ &= \int\int \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\eta(\tau+\ell)}\, \mathrm{d}N(\Lambda_\tau)(\eta)\ \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathrm{d}L(\Lambda_\tau)(\ell) = *_2\end{aligned}$$

We have assumed that $L(\Lambda_\tau) = g_\tau \cdot m_+$, so the outer integral can easily be transformed into an integral with respect to the Lebesgue measure. For the inner integral, we have not specified a density of $N(\Lambda_\tau)$ directly, but if we consider the exponent inside the integral $\eta(\tau+\ell)$ as a function of η for fixed ℓ, that is we consider the evaluation map $\theta_{\tau+\ell}\colon \left(\mathbb{N}_0^{\mathbb{R}}, \sigma(N)\right) \to \left(\mathbb{N}_0, 2^{\mathbb{N}_0}\right)$ such that $\theta_{\tau+\ell}(\eta) = \eta(\tau + \ell)$ we can again use the abstract change-of-variable and that we know the distribution of $\theta_{\tau+\ell}(N) = N(\tau + \ell)$ to be $\theta_{\tau+\ell}\left(N(\Lambda_\tau)\right) = q_\bullet(\tau + \ell) \cdot c_{\mathbb{N}_0}$, where $c_{\mathbb{N}_0}$ is the counting measure on $\mathbb{N}_0$. Using these assumptions, we get,

$$\begin{aligned}*_2 &= \int\int \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\nu}\, \mathrm{d}\theta_{\tau+\ell}(N(\Lambda_\tau))(\nu)\ \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathrm{d}(g_\tau \cdot m_+)(\ell) \\ &= \int\int \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\nu}\, \mathrm{d}(q_\bullet(\tau + \ell) \cdot c_{\mathbb{N}_0})(\nu)\ \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\, \mathrm{d}(g_\tau \cdot m_+)(\ell) \\ &= \int\int \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\nu} q_\nu(\tau + \ell)\, \mathrm{d}c_{\mathbb{N}_0}(\nu)\ \mathbf{1}_{(t-\tau,T-\tau]}(\ell)\ g_\tau(\ell)\, \mathrm{d}m_+(\ell) \\ &= \int_{(t-\tau,T-\tau]} \sum_{\nu\in\mathbb{N}_0} \mathbb{P}_{\tau+\ell}[Z(T) = 0]^{\nu} q_\nu(\tau + \ell)\ g_\tau(\ell)\, \mathrm{d}m_+(\ell) = *_2\end{aligned}$$

where we have used the laws for integrating wrt. a measure with density and that an integral wrt. a counting measure is just a sum.

We then realize that the inner sum defines the probability generating function of $N(\tau + \ell)$, $\zeta(s; t) := \Lambda_\tau[s^{N(t)}]$ evaluated at $s = \mathbb{P}_{\tau+\ell}(Z(T) = 0) \leq 1$

and $t = \tau + \ell$, so we finally have,

$$\ast_2 = \int_{(t-\tau, T-\tau]} \zeta\left(\mathbb{P}_{\tau+\ell}(Z(T) = 0); \tau + \ell\right)\ g_\tau(\ell)\,\mathrm{d}m_+(\ell)$$

giving us the final expression of $\ast$,

$$\begin{aligned}\ast &= \mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0>t)}\right] \\ &= s(1 - G_\tau(t-\tau)) + (1-s)\int_{(t-\tau, T-\tau]} \zeta\left(\mathbb{P}_{\tau+\ell}(Z(T) = 0); \tau + \ell\right)\ g_\tau(\ell)\,\mathrm{d}m_+(\ell).\end{aligned}$$

For the second expectation of Equation 3.4 which is working on the event that the ancestral individual has died at time $t \geq \tau$ we do a completely analogous proof to that of $\ast_2$, the main steps are highlighted below, but the details are skipped. The principle of first generation and the fact that we work on the event $(\tau + L_0 \leq t)$ give us

$$\begin{aligned}\mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0\leq t)}\right] &= \mathbb{P}_\tau\left[s^{\chi_0(t-\tau)+\sum_{x\in\mathscr{R}_1} Z_x^T(t)}\ \mathbf{1}_{(\tau+L_0\leq t)}\right] \\ &= \mathbb{P}_\tau\left[\prod_{x\in\mathscr{R}_1} s^{Z_x^T(t)}\ \mathbf{1}_{(\tau+L_0\leq t)}\right] = \ast\ast\end{aligned}$$

where $\chi_0(t-\tau) = \mathbf{1}_{[0,L_0)}(t-\tau)\,\mathbf{1}_{(Z(T)>0)} = 0 \cdot \mathbf{1}_{(Z(T)>0)} = 0$ under $(\tau + L_0 \leq t)$.

By the law of total expectation, conditioning on $\mathscr{F}_{\mathscr{R}_1}$, and that L_0 is $\mathscr{F}_{\mathscr{R}_1}$ measurable, we again get an expression on which we can use the strong branching property, now with $\varphi_x = (\omega \mapsto s^{Z^T(t,\omega)})$ which is clearly positive, bounded by 1 and measurable as $s \leq 1$. From this, we obtain that

$$\begin{aligned}\ast\ast &= \mathbb{P}_\tau\left[\mathbb{P}_\tau\left[\prod_{x\in\mathscr{R}_1} s^{Z^T(t,S_x)}\,\middle|\,\mathscr{F}_{\mathscr{R}_1}\right]\ \mathbf{1}_{(\tau+L_0\leq t)}\right] \\ &= \mathbb{P}_\tau\left[\mathbb{P}_{\tau+L_0}\left[s^{Z^T(t)}\right]^{N_0(\tau+L_0)}\ \mathbf{1}_{(\tau+L_0\leq t)}\right] = \ast\ast\end{aligned}$$

with the same argument as before to justify the product being over a finite number of identical terms giving us the last line. Identifying the integrand under $\mathbb{P}_\tau$ here as a (L_0, N_0) measurable non-negative variable we can again use our result about the image measure $(L_0, N_0)(\mathbb{P}_\tau) = (L, N)(\Lambda_\tau) = L(\Lambda_\tau) \otimes N(\Lambda_\tau)$ together with Tonelli's theorem to see that

$$\begin{aligned}\ast\ast &= \int\int \mathbb{P}_{\tau+\ell}\left[s^{Z^T(t)}\right]^{\eta(\tau+\ell)}\mathrm{d}N(\Lambda_\tau)(\eta)\ \mathbf{1}_{(\tau+\ell\leq t)}\,\mathrm{d}L(\Lambda_\tau)(\ell) \\ &= \int\int \mathbb{P}_{\tau+\ell}\left[s^{Z^T(t)}\right]^{\theta_{\tau+\ell}(\eta)}\mathrm{d}N(\Lambda_\tau)(\eta)\ \mathbf{1}_{(\tau+\ell\leq t)}\,\mathrm{d}L(\Lambda_\tau)(\ell) \\ &= \int\int \mathbb{P}_{\tau+\ell}\left[s^{Z^T(t)}\right]^{\nu}\mathrm{d}\theta_{\tau+\ell}(N(\Lambda_\tau))(\nu)\ \mathbf{1}_{(\tau+\ell\leq t)}\,\mathrm{d}L(\Lambda_\tau)(\ell) = \ast\ast\end{aligned}$$

where we have also reformulated the exponent as the evaluation map $\theta_{\tau+\ell}(\eta) = \eta(\tau+\ell)$ and used the abstract change-of-variables [47, Theorem 4.10].

[47]: Klenke (2020), *Probability Theory*

The density assumptions $\theta_{\tau+\ell}(N(\Lambda_\tau)) = q_\bullet(\tau+\ell)\cdot c_{\mathbb{N}_0}$ and $L(\Lambda_\tau) = g_\tau \cdot m_+$

give us an expression in which we recognize the probability generating function of $N(\tau+\ell)$, $\zeta(s;t) := \Lambda_\tau[s^{N(t)}]$ evaluated at $s = \mathbb{P}_{\tau+\ell}\left[s^{Z^T(t)}\right] = \psi_{\tau+\ell}(s;t) \leq 1$ and $t = \tau + \ell$, so we finally have,

$$\begin{aligned}
** &= \mathbb{P}_\tau\left[s^{Z^T(t)}\,\mathbf{1}_{(\tau+L_0\leq t)}\right] \\
&= \int_{[0,t-\tau]} \sum_{\nu\in\mathbb{N}_0} \mathbb{P}_{\tau+\ell}\left[s^{Z^\chi(t)}\right]^\nu q_\nu(\tau+\ell)\; g_\tau(\ell)\,\mathrm{d}m_+(\ell) \\
&= \int_{[0,t-\tau]} \zeta\left(\psi_{\tau+\ell}(s;t);\tau+\ell\right)\; g_\tau(\ell)\,\mathrm{d}m_+(\ell)
\end{aligned}$$

Collecting the resulting expressions for $*$ and $**$ we obtain the wished for integral equation for ψ_τ,

$$\begin{aligned}
\psi_\tau^T(s;t) &= s\,(1-G_\tau(t-\tau)) \\
&\quad + (1-s)\int_{(t-\tau,T-\tau]} \zeta\left(\mathbb{P}_{\tau+\ell}(Z(T)=0);\tau+\ell\right)\; g_\tau(\ell)\,\mathrm{d}m_+(\ell) \\
&\quad + \int_{[0,t-\tau]} \zeta(\psi_{\tau+\ell}(s;t);\tau+\ell)\; g_\tau(\ell)\,\mathrm{d}m_+(\ell)
\end{aligned}$$

Using that $\mathbb{P}_{\tau+\ell}(Z(T)=0) = \psi_{\tau+\ell}^T(0;T)$, we also gain the alternative expression. □

Remark 3.3.2 One might speculate as to whether a solution to the integral equation exists and if it is unique, as is done in [8] for traditional Bellman-Harris processes, and for similar integral equations, but for different functions in [41]. But as we have defined the generating function $\psi_\tau^T(\cdot,t)$ for all $\tau \leq t \leq T$ directly as the expectation of a bounded random variable it must exist, and we have in the theorem above simply shoved that this function satisfies the integral equation. However, whether this is the only solution to the integral equation is not certain. We conjecture that a proof similar to that of [8, Chapter IV, Theorem 1] would show uniqueness in the set of bounded functions.

[8]: Athreya et al. (1972), *Branching Processes*

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

[8]: Athreya et al. (1972), *Branching Processes*

As noted, the reduced and the classical branching processes are tightly linked, and from this generating function for the reduced process we can obtain the simpler generating function for the classical time-varying Bellman-Harris process,

Example 3.3.2 Using the property that $Z^t(t) = Z(t)$ for all $t \in \mathbb{R}$ we can obtain the generating function of the non-reduced time-varying Bellman-Harris process, which we denote $\psi_\tau(\cdot;t)\colon [0,1] \to [0,1]$ following the notational difference between Z^T and Z,

$$\begin{aligned}
\psi_\tau(s;t) &= \psi_\tau^t(s;t) \\
&= s\,(1-G_\tau(t-\tau)) + \int_{[0,t-\tau]} \zeta(\psi_{\tau+\ell}^T(s;t);\tau+\ell)\; g_\tau(\ell)\,\mathrm{d}m(\ell)
\end{aligned}$$

which is exactly the generating function for the time-varying Bellman-Harris process of [46, 61], when taking into account our slightly different framework.

[46]: Kimmel (1983), 'The Point-Process Approach to Age- and Time-Dependent Branching Processes'

[61]: Pakkanen et al. (2023), 'Unifying incidence and prevalence under a time-varying general branching process'

By the properties of generating functions in Remark 3.3.1 we can completely characterize the distribution $Z^T(\mathbb{P}_\tau)$ for each $\tau \in \mathbb{R}$ with our equation for the generating function. But evidently, it is not a simple closed-form characterization, so if we wish to compute the distribution in practice we either need extra strict distributional assumptions, for example, to reduce our process to a birth-death process, or computational tools to solve the integral equation numerically. One simple distributional assumption, which seems reasonable in many applications, is to only work on a binary population, i.e. each individual either has two children or none. This does not result in a closed-form solution of the integral equation but simplifies it, making computational approaches more feasible.

3.4 The reduced binary branching process

As noted in Section 2.2, all constructions we have done with the full Ulam-Harris space are very simple, however tedious, to redo with I being the binary Ulam-Harris space, several proofs even become much simpler as each generation is by definition finite. If we define a branching process on the binary time-population space, in which individuals with probability one produce either none or two offspring, we will call it a *binary branching process*.[12] However, there are no substantive differences between working with a binary branching population and a full branching population on which the offspring process is almost surely either 0 or 2 other than technical details,[13] so we will not distinguish the binary branching process from the branching process on the full space with this assumption.[14]

From this point on, we only consider binary branching processes, regardless of whether they are defined from the full Ulam-Harris space or the binary Ulam-Harris space. For notational convenience, we will not distinguish between the binary process and the "full" process, so we will simply be working on the full Ulam-Harris space, but always assuming that $q_0(t) + q_2(t) = 1$ for all $t \in \mathbb{R}$. We will also often just talk about the reduced process instead of the reduced binary process.

Restricting the offspring process to almost surely only take values in $\{0, 2\}$ corresponds to assuming $q_0(t) = 1 - q_2(t)$ for all $t \in \mathbb{R}$ and consequently also having a generating function $\zeta(s;t) := \mathbb{P}_\tau\left(s^{N(t)}\right)$ given by

$$\begin{aligned}\zeta(s;t) &= s^0 q_0(t) + s^2 q_2(t) \\ &= q_0(t) + s^2 q_2(t) \qquad (3.5) \\ &= 1 + (s^2 - 1) q_2(t) \qquad (3.6)\end{aligned}$$

for $s \in [0, 1]$ and $t \in \mathbb{R}$. This makes it clear that we in principle only need to specify the time-varying probability of having two children, and some equations will reduce a bit by using this fact, with the expense that they have a less clear interpretation than if we were to write them out with both q_0 and q_2, hence we will generally show both versions of the equations.

12: Note that it is not allowed for an individual to produce only one offspring.

13: In particular can we see that the set of realized individuals from the binary process defined from the full Ulam-Harris space must be a subset of the binary Ulam-Harris space treated as a Neveu tree in the full Ulam-Harris space. As all our processes are defined through the set of realized individuals, we see that both constructions are essentially equivalent.

14: Introducing the binary branching process through the binary branching population might seem redundant, however when we restrict our attention to binary branching population in Chapter 4 this will be helpful.

With a simple expression for the generating function of the offspring process, the generating function of the reduced binary branching process simplifies a bit from the very general form of Theorem 3.3.1,

Corollary 3.4.1 *Let $\tau \in \mathbb{R}$. The generating function for the reduced binary branching process started in time τ, $\psi_\tau^T(s;t) := \mathbb{P}_\tau\left[s^{Z^T(t)}\right]$ satisfies*

$$\begin{aligned}\psi_\tau^T(s;t) &= s\,(1 - G_\tau(t-\tau)) \\ &+ (1-s)\int_{(t-\tau,T-\tau]} \left(q_0(\tau+\ell) + q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T)=0)^2\right) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &+ \int_{[0,t-\tau]} \left(q_0(\tau+\ell) + q_2(\tau+\ell)\psi_{\tau+\ell}(s;t)^2\right) g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

or using that $q_0(t) = 1 - q_2(t)$,

$$\begin{aligned}\psi_\tau^T(s;t) &= s + (1-s)G_\tau(T-\tau) \\ &+ (1-s)\int_{(t-\tau,T-\tau]} \left(\mathbb{P}_{\tau+\ell}(Z(T)=0)^2 - 1\right) q_2(\tau+\ell)\, g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &+ \int_{[0,t-\tau]} \left(\psi_{\tau+\ell}(s;t)^2 - 1\right) q_2(\tau+\ell)\, g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

for $s \in [0,1]$ and $\tau \le t \le T$.

Proof. The first form follows directly from plugging Equation 3.5 into Theorem 3.3.1. For the second form, we do the same but with Equation 3.6,

$$\begin{aligned}\psi_\tau^T(s;t) &= s\,(1 - G_\tau(t-\tau)) \\ &\quad + (1-s)\int_{(t-\tau,T-\tau]} \left(1 + \left(\mathbb{P}_{\tau+\ell}(Z(T)=0)^2 - 1\right) q_2(\tau+\ell)\right) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + \int_{[0,t-\tau]} \left(1 + \left(\psi_{\tau+\ell}^T(s;t)^2 - 1\right) q_2(\tau+\ell)\right) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &= s\,(1 - G_\tau(t-\tau)) \\ &\quad + (1-s)\int_{(t-\tau,T-\tau]} g_\tau\,\mathrm{d}m \\ &\quad + (1-s)\int_{(t-\tau,T-\tau]} \left(\mathbb{P}_{\tau+\ell}(Z(T)=0)^2 - 1\right) q_2(\tau+\ell) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + \int_{[0,t-\tau]} g_\tau\,\mathrm{d}m \\ &\quad + \int_{[0,t-\tau]} \left(\psi_{\tau+\ell}^T(s;t)^2 - 1\right) q_2(\tau+\ell) g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

Using $G_\tau(0) = 0$ for all τ, that $\int_{(a,b]} g_\tau\,\mathrm{d}m = G_\tau(b) - G_\tau(a)$ for any $a < b$, and that we can arbitrarily choose to integrate over open, closed, or half-open intervals when doing so with respect to m, as the singletons

have Lebesgue measure 0, we get

$$\begin{aligned}\psi_\tau^T(s;t) &= s\,(1 - G_\tau(t-\tau)) \\ &\quad + (1-s)\,(G_\tau(T-\tau) - G_\tau(t-\tau)) \\ &\quad + (1-s)\int_{(t-\tau,T-\tau]} \left(\mathbb{P}_{\tau+\ell}(Z(T)=0)^2 - 1\right) q_2(\tau+\ell) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + G(t-\tau) \\ &\quad + \int_{[0,t-\tau]} \left(\psi_{\tau+\ell}^T(s;t)^2 - 1\right) q_2(\tau+\ell) g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

If we now consider all the distribution functions in isolation, we realize that many of them will cancel each other,

$$\begin{aligned}&\cancel{-\,sG_\tau(t-\tau)} + G_\tau(T-\tau) - \cancel{G_\tau(t-\tau)} \\ &\quad - sG_\tau(T-\tau) + \cancel{sG_\tau(t-\tau)} + \cancel{G_\tau(t-\tau)} \\ &= (1-s)G_\tau(T-\tau)\end{aligned}$$

So we are left with the generating function for the reduced binary branching process,

$$\begin{aligned}\psi_\tau^T(s;t) &= s + (1-s)G_\tau(T-\tau) \\ &\quad + (1-s)\int_{(t-\tau,T-\tau]} \left(\mathbb{P}_{\tau+\ell}(Z(T)=0)^2 - 1\right) q_2(\tau+\ell)\, g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + \int_{[0,t-\tau]} \left(\psi_{\tau+\ell}(s;t)^2 - 1\right) q_2(\tau+\ell)\, g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

and using that $\mathbb{P}_{\tau+\ell}(Z(T)=0) = \psi_{\tau+\ell}^T(0;T)$, we gain the alternative expression. □

From this equation for the generating function of the reduced binary process, we can obtain similar equations for several interesting quantities of the process using the properties of generating functions from Remark 3.3.1.[15] Some of these quantities are interesting in their own right, and some of them are merely going to be helpful in the sequel. The proofs are straightforward calculations and are thus left out.

15: No real extra difficulty comes in computing these quantities for the general, non-binary, branching process, however, we will not need those quantities, and their expressions are much less elegant as we don't know the s-derivative of ζ in general.

3.4.1 Probability of extinction

In the study of stochastic processes, especially Markov chains, we are often interested in *absorbing states*. That is, a state (i.e. a value) of the process in which the process will be indefinitely with probability one once it enters. For the classical branching process, 0 is an absorbing state, since once no individuals are alive, no new individuals can be born, i.e. if there is a $t \in \mathbb{R}$ such that $\mathbb{P}_\tau(Z(t)=0) = 1$ then $\mathbb{P}_\tau(Z(s)=0) = 1$ for all $s \geq t$. This probability of extinction is thus an interesting quantity in itself, but we also see that it appears in the generating equations for the reduced processes, and likewise, it will be present in the other quantities that we will study.

For the reduced process, 0 is also an absorbing state, however, it has an even more special role here as if the non-reduced process has gone extinct at any time before observation, then the reduced process will be

identically zero at any time as argued in Proposition 3.2.1. Thus we must have that $(Z^T(t) = 0) = (Z^T(T) = 0)$ for all $t \geq \tau_0$, which we can find the probability of under $\mathbb{P}_\tau$ by evaluating the generating function in $s = 0$ and $t = T$,

Corollary 3.4.2 *For $\tau \in \mathbb{R}$, let Z^T be the reduced binary branching process. The extinction probability of the process $\mathbb{P}_\tau\left(Z^T(t) = 0\right)$ satisfies the integral equation*

$$\begin{aligned}\mathbb{P}_\tau\Big(Z^T(t) = 0\Big) &= \mathbb{P}_\tau(Z(T) = 0) = \psi_\tau^T(0;T) \\ &= G_\tau(T-\tau) + \int_{[0,T-\tau]} \Big(\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2 - 1\Big)\, q_2(\tau+\ell)\, g_\tau(\ell)\, \mathrm{d}m(\ell) \\ &= \int_{[0,T-\tau]} \Big(q_0(\tau+\ell) + \mathbb{P}_{\tau+\ell}(Z(T) = 0)^2 q_2(\tau+\ell)\Big)\, g_\tau(\ell)\, \mathrm{d}m(\ell)\end{aligned}$$

for $\tau \leq t \leq T$.

The last form of the integral equation has a clear interpretation. Consider the process at some time $t \in [\tau, T]$. If the ancestor leaves no extant progeny by time T, she must, in particular, have died at some age $\ell \in [0, T-\tau]$, at which she either produced no offspring (with probability $q_0(\tau+\ell)$) and hence has no extant progeny, or she produced two offspring ($q_2(\tau+\ell)$) and both of these produce no extant progeny ($\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2$). Integrating over the ancestor's age at death, with respect to the lifetime distribution (with m-density g_τ), gives us the integral equation.

Example 3.4.1 (The Yule process (continued)) Now as we have stated several times, the Yule process cannot go extinct. Hence it should be quite evident that $\mathbb{P}_\tau\left(Z^T(t) = 0\right) = \mathbb{P}_\tau(Z(t) = 0) = 0$ satisfies the equation in Corollary 3.4.2 as we also have $q_0(t) = 0$ for all $t \geq \tau$.

For the non-reduced birth-death process, the probability of extinction before time T of a population initiated at time 0 is found in [41], where it is denoted by $X(T)$. If we take into account the slight differences in our framework, in particular, that we have arbitrary origin and observation times, the equations of Corollary 3.4.2 and [41, Eq. (5)] are equivalent. Hence we must by [41, Theorem 3] have that

$$\mathbb{P}_\tau\Big(Z^T(t) = 0\Big) = \frac{(1-q)(1-e^{(1-2q)\beta\delta(T-\tau)})}{q-(1-q)e^{(1-2q)\beta\delta(T-\tau)}}$$

is a solution to Corollary 3.4.2.

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

3.4.2 Probability of unity

The probability $\mathbb{P}_\tau\left(Z^T(t) = 1\right)$ might seem like an arbitrary quantity to work with at first, but if we consider it a function of $t \geq \tau$ it gives us some insight into the jump times of the reduced process, given that it has not gone extinct before observation, namely the probability that the first jump has yet to happen at time t for a population born at time τ. The distribution of jump times is, just like the study of absorbing states, a central part of the theory of stochastic processes, and for us in particular,

jump times give us death times of least common ancestors of (subsets of) the extant population. The probability is found by taking the s-derivative of the generating function and evaluating in $s = 0$,

Corollary 3.4.3 *Let $\tau \in \mathbb{R}$. The probability that the reduced binary branching process takes the value one at time t satisfies the integral equation*

$$\begin{aligned}\mathbb{P}_\tau\left(Z^T(t) = 1\right) &= 1 - G_\tau(t-\tau) \\ &\quad - \int_{(t-\tau, T-\tau]} \left(q_0(\tau+\ell) + q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2\right) g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + 2\int_{[0,t-\tau]} q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T) = 0)\mathbb{P}_{\tau+\ell}\left(Z^T(t) = 1\right) g_\tau(\ell)\,\mathrm{d}m(\ell).\end{aligned}$$

for $\tau \le t \le T$.

Again we can give an interpretation of the equation. Let us look at the process for some $t \in [\tau, t]$. First, if the ancestor is not yet dead (with probability $1 - G_\tau(t-\tau)$, the process will only equal one if she produces some extant progeny at observation, thus from the probability of survival of the ancestor we subtract the probability of her not producing any extant progeny, i.e. if she dies at age $\ell \in (t-\tau, T-\tau]$ she either have to produce no offspring at her death ($q_0(\tau+\ell)$) or she can produce two offspring ($q_2(\tau+\ell)$), but neither of them leaves any extant progeny ($\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2$). Integrating over her possible age at death, with respect to the lifetime distribution (with m-density g_τ), accounts for the first two terms. If the ancestor, on the other hand, has died, i.e. she died at some age $\ell \in [0, t-\tau]$, she would not contribute to the process anymore, and then the process can only equal one if she actually produced two offspring ($q_2(\tau+\ell)$), but one of them produces no extant progeny ($\mathbb{P}_{\tau+\ell}(Z(T) = 0)$), while the progeny reduced process started by the other equals one at time t ($\mathbb{P}_{\tau+\ell}\left(Z^T(t) = 1\right)$), which child takes which path is immaterial, hence the factor of 2. Integrating over the ancestor's age at death, with respect to the lifetime distribution (with m-density g_τ), accounts for the last term.

Example 3.4.2 (Yule and birth-death processes) Again for the Yule process, extinctions cannot happen, i.e. $\mathbb{P}_\tau\left(Z^T(t) = 0\right) = \mathbb{P}_\tau(Z(T) = 0) = 0$, so the integral equation of Corollary 3.4.3 simplifies to

$$\mathbb{P}_\tau\left(Z^T(t) = 1\right) = 1 - G_\tau(t-\tau) = e^{-\beta t}$$

i.e. the survival function of an exponential distribution. This is also what we would expect, as the process can only ever be one if the ancestor has not died, otherwise it must take higher values.

For the birth-death processes, we initially remark that this is not the same equation as in [41, Eq. (2)] as this equation treats the probability of unity in a non-reduced branching process. We conjecture that a version of the proof of [41, Theorem 3] would yield an explicit solution to the integral equation of Corollary 3.4.3, but have not succeeded in finding it.

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

3.4.3 Mean of the reduced process

To have a rough sketch of the dynamics of the reduced process we might be interested in characterizing the mean function over time of the reduced process. Moreover, for the reconstructed birth-death model, this is exactly the function known in phylogenetics as the *deterministic lineages-through-time function*[59, 75], which we might want to fit to an observed lineages-through-time function to estimate the underlying distributions of our model. Fortunately, the mean can be found from the generating function as the first s-derivative evaluated in $s = 1$.

[59]: Nee et al. (1994), 'The Reconstructed Evolutionary Process'
[75]: Stadler (2008), 'Lineages-through-time plots of neutral models for speciation'

Corollary 3.4.4 *Let $\tau \in \mathbb{R}$. The reduced binary branching process has mean function $t \mapsto \mathbb{P}_\tau\left[Z^T(t)\right]$ satisfying the integral equation*

$$\begin{aligned}\mathbb{P}_\tau\left[Z^T(t)\right] &= 1 - G_\tau(t-\tau) \\ &\quad - \int_{(t-\tau, T-\tau]} \left(q_0(\tau+\ell) + q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T)=0)^2\right)\, g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &\quad + 2\int_{[0,t-\tau]} q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}\left[Z^T(t)\right] g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

for $\tau \le t \le T$.

Again we lead with an interpretation of the integral equation. The expectation of the process at any given time $t \in [\tau, T]$ consists again of a contribution from the situation where the ancestor is still alive and a contribution from when she have died. The first two terms are exactly the expected value of the process if the ancestor were still alive, which happens with probability $1 - G_\tau(t-\tau)$, from which we subtract the probability of the ancestor producing no extant progeny at observation, as argued earlier can be written as in the second term here. The last term is the expected value in the situation where the ancestor has died, only if she produces two offspring will it add to the expectation, and it will do so corresponding to the expectation of each of the two offspring produced at her time of death.

Example 3.4.3 (Mean of a Yule process) Let us choose $q_2(t) = 1$ for all $t \in \mathbb{R}$, $g_\tau(\ell) = \beta e^{-\beta\ell}$ and thus $G_\tau(\ell) = 1 - e^{-\beta\ell}$ for $\ell \in \mathbb{R}_+$. In Example 3.3.1 this was associated with the Yule process. We will show that the mean of the reduced branching process coincides with the mean of the traditional Yule process, by seeing that it satisfies the integral equation in Corollary 3.4.4. One can find expressions for the mean of a Yule process started in time 0 in most textbooks on stochastic processes [44] and of course in Yule's original paper [91]. A slight modification for the process started in time τ give us the mean function $M_\tau(t) = e^{\beta(t-\tau)}$.

As noted, when choosing $q_2(t) = 1$ and thus $q_0(t) = 0$ for all $t \in \mathbb{R}$, the process will almost surely never go extinct, i.e. $\mathbb{P}_\tau(Z(T) = 0) = 0$ for all $\tau \in \mathbb{R}$, this removes the second term of the integral equation for the mean. Plugging the distributional assumptions and $\mathbb{P}_{\tau+\ell}\left[Z^T(t)\right] = M_{\tau+\ell}(t) = e^{\beta(t-(\tau+\ell))}$ into the right hand side of the integral equation

[44]: Karlin et al. (1975), *A first course in stochastic processes*

[91]: Yule (1925), 'A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.'

give us

$$
\begin{aligned}
&1 - G_\tau(t-\tau) + 2\int_{[0,t-\tau]} q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}\left[Z^T(t)\right] g_\tau(\ell)\,\mathrm{d}m(\ell) \\
&\quad = e^{-\beta(t-\tau)} + 2\int_{[0,t-\tau]} e^{\beta(t-(\tau+\ell))}\beta e^{-\beta\ell}\,\mathrm{d}m(\ell) \\
&\quad = e^{-\beta(t-\tau)} + e^{\beta(t-\tau)}\int_{[0,t-\tau]} 2\beta e^{-2\beta\ell}\,\mathrm{d}m(\ell) \\
&\quad = e^{-\beta(t-\tau)} + e^{\beta(t-\tau)}\left(1 - e^{-2\beta(t-\tau)}\right) \\
&\quad = e^{\beta(t-\tau)} = M_\tau(t) = \mathbb{P}_\tau\left[Z^T(t)\right]
\end{aligned}
$$

where we have recognized the final integral as the distribution function of an exponential distribution with rate parameter 2β.

3.5 Death of the least common ancestor

The time of death of the least common ancestor of the extant individuals will play a special role in the construction and analysis of the *reduced branching trees* of the extant individuals. A heuristic sneak peek at the main constructions in Chapter 4 will show why. The reduced branching process counts individuals who have extant progeny at observation, or in other words, it counts the number of ancestors of the extant individuals who are alive at a given time before observation. In the same way, when we want to study the trees formed by such a branching process, we only want to consider the ancestors' branches and their branch lengths as given by the ancestors' lifetimes. Such a tree is called the *reduced* or *reconstructed branching tree*.

Consider now the root branch of the reduced tree, i.e. the branch representing the most recent common ancestor of the extant individuals. The length of this branch should hence be determined by the origin-time of the population and the death time of the least common ancestor, that is, the first time at which two individuals are born who both have extant progeny. Considering the subtrees generated by each of these individuals, we can of course do the same recursively down each line of descent. Hence, if we can characterize the death time of the least common ancestor, we can recursively characterize the branch lengths of the reduced branching trees. Furthermore, the time of death of the least common ancestor also reveals the time since the least common ancestor in the process, i.e. the death time subtracted from the observation time, a variable that in itself is interesting and has been studied immensely.

Defining the death time of the least common ancestor has one obstruction, what if the population has gone extinct before observation? In this case, the set of extant individuals is the empty set, whose least common ancestor is not defined, and no clear extension of the definition seems to fit.[16] In other words, $\mathrm{lca}(\mathfrak{X}_T)$ is only defined on the event $(Z(T) > 0)$, and we can thus only define the death time of the least common ancestor $\sigma_{\mathrm{lca}(\mathfrak{X}_T)}$ on the same event. When working with this variable, we will thus always condition on the event of non-extinction by observation, so we can consider it a variable on $(\overline{\Omega}, \overline{\mathcal{A}})$, but ignore the non-defined part of the

16: One might suggest that we define $\mathrm{lca}\,\emptyset := \emptyset$ but then we would just run into trouble when defining the death time of the empty set.

variable. All of this is certainly only sensible when $\mathbb{P}_\tau(Z(T) > 0) \neq 0$, but populations for which this does not hold are not of primary interest.

Consider the reduced branching process of a population originating at time $\tau \in \mathbb{R}$ which we condition on non-extinction by observation time T. From the discussion above, the reduced branching process is counting the number of ancestors of the extant individuals alive at a given time, so the death time of the least common ancestor must be the time the process jumps from 1 to 2. Thus, using that the reduced process is weakly increasing (Proposition 3.2.1) and starts in 1 when conditioning on non-extinction by observation, we treat the probability that the reduced process equals one as a function of t, and get a partial characterization of the survival function of $\sigma_{\mathrm{lca}\,\mathscr{X}_T}$, seeing that

$$\mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} > t \,\middle|\, Z(T) > 0\right) = \mathbb{P}_\tau\left(Z^T(t) = 1 \,\middle|\, Z(T) > 0\right)$$

for $t \in [\tau, T]$. For $t < \tau$, $\mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} > t \,\middle|\, Z(T) > 0\right) = 0$ as the least common ancestor cannot die before the population originated, but for $t > T$ the reduced process is not defined and thus cannot give us any characterization of the survival function. However, for the variable $\sigma_T := \sigma_{\mathrm{lca}(\mathscr{X}_T)} \wedge T$, which is again only defined on the event $(Z(T) > 0)$, the reduced process can give us a full characterization of the distribution through its survival, and thus distribution function.

Lemma 3.5.1 *Let $\tau \in \mathbb{R}$ and assume that $\mathbb{P}_\tau(Z(T) > 0) \neq 0$. The distribution $\sigma^T\left(\mathbb{P}_\tau^{(Z(T)>0)}\right)$ on $(\mathbb{R}, \mathscr{B})$ have distribution function*

$$\begin{aligned} F_\tau^T(t) &= 1 - \mathbf{1}_{(-\infty,T)}(t)\mathbb{P}_\tau\left(Z^T(t) = 1 \,\middle|\, Z(T) > 0\right) \\ &= 1 - \mathbf{1}_{(-\infty,T)}(t)\frac{\mathbb{P}_\tau\left(Z^T(t) = 1\right)}{\mathbb{P}_\tau(Z(T) > 0)} \end{aligned}$$

for all $t \in \mathbb{R}$. F_τ^T is absolutely continuous on $[\tau, T)$ and jumps to one at $t = T$.

Proof. For $t < T$ we have from the discussion above

$$\begin{aligned} F_\tau^T(t) &= \mathbb{P}_\tau\left(\sigma^T \leq t \,\middle|\, Z(T) > 0\right) \\ &= \mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} \leq t \,\middle|\, Z(T) > 0\right) \\ &= 1 - \mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} > t \,\middle|\, Z(T) > 0\right) \\ &= 1 - \mathbb{P}_\tau\left(Z^T(t) = 1 \,\middle|\, Z(T) > 0\right) \end{aligned}$$

[47]: Klenke (2020), *Probability Theory*

Note that as $\sigma_{\mathrm{lca}(\mathscr{X}_T)}\left(\mathbb{P}_\tau^{(Z(T)>0)}\right)$ is an absolutely continuous distribution with respect to m, it must have an m-density by the Radon-Nikodym theorem [47, Corollary 7.34] and in particular, must its distribution function be absolutely continuous. As this distribution function coincides with F_τ^T on $[\tau, T)$ the same is true on this interval.

For $t \geq T$ we see that

$$\begin{aligned} F_\tau^T(t) &= \mathbb{P}_\tau\left(\sigma^T \leq t \,\middle|\, Z(T) > 0\right) \\ &= \mathbb{P}_\tau\left(\sigma^T \wedge T \leq t \wedge T \,\middle|\, Z(T) > 0\right) \\ &= \mathbb{P}_\tau\left(\sigma^T \leq T \,\middle|\, Z(T) > 0\right) = 1 \end{aligned}$$

so the distribution has a point mass at T, and thus a jump in the distribution function of size

$$\begin{aligned} \mathbb{P}_\tau\left(\sigma^T = T \,\middle|\, Z(T) > 0\right) &= \mathbb{P}_\tau\left(Z^T(T) = 1 \,\middle|\, Z(T) > 0\right) \\ &= \mathbb{P}_\tau(Z(T) = 1 \mid Z(T) > 0) \end{aligned}$$

□

This lemma proves that $\sigma^T\left(\mathbb{P}_\tau^{(Z(T)>0)}\right)$ has a decomposition into an absolutely continuous part with respect to m, namely its restriction to $\left([\tau, T), \mathcal{B}([\tau, T))\right)$ and a discrete (also singular with respect to m) part, being a point mass at T[17]. Now by the Radon-Nikodym theorem, the absolutely continuous part must have an m-density, which we can find by taking the t-derivative of $F_\tau^T(t)$ over $[\tau, T)$.

17: This can also be seen as a consequence of the Lebesgue decomposition theorem [47, Theorem 7.33], where the singular part of the distribution is measure with the point mass at T.

The conditional probability $\mathbb{P}_\tau\left(Z^T(t) = 1 \,\middle|\, Z(T) > 0\right)$, from which we have described F_τ^T, has not been directly characterized, but as we condition on an event, which we assume to have positive probability, and furthermore that $(Z^T(t) = 1) \subset (Z(T) > 0)$, the conditional probability is simple to find from the quantities characterized in the last sections.

Theorem 3.5.2 *Let $\tau \in \mathbb{R}$ and assume that $\mathbb{P}_\tau(Z(T) > 0) \neq 0$. The restriction of the distribution*[18] *$\sigma^T\left(\mathbb{P}_\tau^{(Z(T)>0)}\right)$ to $\left([\tau, T), \mathcal{B}([\tau, T))\right)$ has m-density $f_\tau^T(t)$ given as,*

$$\begin{aligned} f_\tau^T(t) = \frac{g_\tau(t-\tau)}{\mathbb{P}_\tau(Z(T) > 0)}\Bigg(1 - q_0(t) - q_2(t)\,\mathbb{P}_t(Z(T) = 0)^2 \\ - 2\,q_2(t)\,\mathbb{P}_t(Z(T) = 0)\,\mathbb{P}_t\left(Z^T(t) = 1\right)\Bigg) \\ + 2\int_{[0,t-\tau]} q_2(\tau + \ell)\mathbb{P}_{\tau+\ell}(Z(T) = 0) f_{\tau+\ell}^T(t) g_\tau(\ell)\,\mathrm{d}m(\ell) \end{aligned}$$

for $t \in [\tau, T)$ and which vanishes outside this interval.

18: Note that when restricted, it is no longer a probability distribution as $[\tau, T)$ does not have measure one under the (unrestricted) distribution.

Proof. We have argued that this density exists in the discussion above, differentiability (m-almost everywhere) of the distribution function over $[\tau, T)$ follows from the lemma which proves that F_τ^T is absolutely continuous on the interval.

Notice now that $(Z^T(t) = 1) \subset (Z(T) > 0)$ as if the population is extinct at observation, the reduced process must equal 0 at all times, so the

distribution function becomes,

$$\begin{aligned} F_\tau^T(t) &= 1 - \mathbb{P}_\tau\Big(Z^T(t) = 1 \,\Big|\, Z(T) > 0\Big) \\ &= 1 - \frac{\mathbb{P}_\tau\big(Z^T(t) = 1, Z(T) > 0\big)}{\mathbb{P}_\tau(Z(T) > 0)} \\ &= 1 - \frac{\mathbb{P}_\tau\big(Z^T(t) = 1\big)}{\mathbb{P}_\tau(Z(T) > 0)} \end{aligned}$$

for $t \in [\tau, T)$, where each of these quantities has been characterized in the preceding sections.

To find the density of the distribution we can take the t-derivative of this distribution function. I.e. we have that

$$f_\tau^T(t) = -\frac{\frac{\partial}{\partial t}\mathbb{P}_\tau\big(Z^T(t) = 1\big)}{\mathbb{P}_\tau(Z(T) > 0)} \tag{3.7}$$

Let us consider the numerator here,

$$\begin{aligned} \frac{\partial}{\partial t}\mathbb{P}_\tau\Big(Z^T(t) = 1\Big) &= \frac{\partial}{\partial t}(1 - G_\tau(t-\tau)) \\ &- \frac{\partial}{\partial t}\int_{(t-\tau, T-\tau]} \Big(q_0(\tau+\ell) + q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2\Big)\, g_\tau(\ell)\,\mathrm{d}m(\ell) \\ &+ 2\frac{\partial}{\partial t}\int_{[0, t-\tau]} q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T) = 0)\mathbb{P}_{\tau+\ell}\Big(Z^T(t) = 1\Big)\, g_\tau(\ell)\,\mathrm{d}m(\ell) \end{aligned}$$

[47]: Klenke (2020), *Probability Theory*

19: Any singleton is a m-null set, so we can without trouble integrate over $[t-\tau, T-\tau]$ instead of over $(t-\tau, T-\tau]$.

Wanting to interchange the derivatives and integrals, we transform the integrals from Lebesgue integrals to Riemann integrals. By [47, Theorem 4.23] there arise no complications in this transformation as we can consider both integrals to be over the closed intervals[19], on which the integrands, considered as functions of t are continuous, and thus Riemann integrable,

$$\begin{aligned} \frac{\partial}{\partial t}\mathbb{P}_\tau\Big(Z^T(t) = 1\Big) &= \frac{\partial}{\partial t}(1 - G_\tau(t-\tau)) \\ &- \frac{\partial}{\partial t}\int_{t-\tau}^{T-\tau} \Big(q_0(\tau+\ell) + q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T) = 0)^2\Big)\, g_\tau(\ell)\,\mathrm{d}\ell \\ &+ 2\frac{\partial}{\partial t}\int_0^{t-\tau} q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T) = 0)\mathbb{P}_{\tau+\ell}\Big(Z^T(t) = 1\Big)\, g_\tau(\ell)\,\mathrm{d}\ell \end{aligned}$$

20: One could argue that a more coherent treatment could have been given with a theorem on interchanging Lebesgue integrals and derivatives e.g. [47, Theorem 6.28], but problems arise as the indicator function on the intervals we integrate over, are not differentiable in the traditional sense, hence one would have to work with *generalized functions* and *weak derivatives*. In that light, it does not seem too incoherent to transform the problem from Lebesgue to Riemann integrals.

Now using Leibniz' rule for interchanging derivatives and Riemann integrals[20], we get

$$\begin{aligned} \frac{\partial}{\partial t}\mathbb{P}_\tau\Big(Z^T(t) = 1\Big) &= -g_\tau(t-\tau) \\ &+ \Big(q_0(t) + q_2(t)\,\mathbb{P}_t(Z(T) = 0)^2\Big)\, g_\tau(t) \\ &+ 2\,q_2(t)\,\mathbb{P}_t(Z(T) = 0)\,\mathbb{P}_t\Big(Z^T(t) = 1\Big)\, g_\tau(t) \\ &+ 2\int_0^{t-\tau} q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T) = 0)\,\frac{\partial}{\partial t}\mathbb{P}_{\tau+\ell}\Big(Z^T(t) = 1\Big) g_\tau(\ell)\,\mathrm{d}\ell \end{aligned}$$

where the integral can, for the sake of consistency, be transformed back

into a Lebesgue integral

$$\begin{aligned}\frac{\partial}{\partial t}\mathbb{P}_\tau\Big(Z^T(t)=1\Big) &= -g_\tau(t-\tau)\\ &+\Big(q_0(t)+q_2(t)\,\mathbb{P}_t(Z(T)=0)^2\Big)\,g_\tau(t-\tau)\\ &+2\,q_2(t)\,\mathbb{P}_t(Z(T)=0)\,\mathbb{P}_t\Big(Z^T(t)=1\Big)\,g_\tau(t-\tau)\\ &+2\int_{[0,t-\tau]} q_2(\tau+\ell)\,\mathbb{P}_{\tau+\ell}(Z(T)=0)\,\frac{\partial}{\partial t}\mathbb{P}_{\tau+\ell}\Big(Z^T(t)=1\Big)g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

Plugging this into Equation 3.7 we get

$$\begin{aligned}f_\tau^T(t) &= \frac{g_\tau(t-\tau)}{\mathbb{P}_\tau(Z(T)>0)}\Big(1-q_0(t)-q_2(t)\,\mathbb{P}_t(Z(T)=0)^2\\ &\qquad -2\,q_2(t)\,\mathbb{P}_t(Z(T)=0)\,\mathbb{P}_t\Big(Z^T(t)=1\Big)\Big)\\ &\quad -2\int_{[0,t-\tau]} q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T)=0)\frac{\frac{\partial}{\partial t}\mathbb{P}_{\tau+\ell}\big(Z^T(t)=1\big)}{\mathbb{P}_{\tau+\ell}(Z(T)>0)}g_\tau(\ell)\,\mathrm{d}m(\ell)\\ &= \frac{g_\tau(t-\tau)}{\mathbb{P}_\tau(Z(T)>0)}\Big(1-q_0(t)-q_2(t)\,\mathbb{P}_t(Z(T)=0)^2\\ &\qquad -2\,q_2(t)\,\mathbb{P}_t(Z(T)=0)\,\mathbb{P}_t\Big(Z^T(t)=1\Big)\Big)\\ &\quad +2\int_{[0,t-\tau]} q_2(\tau+\ell)\mathbb{P}_{\tau+\ell}(Z(T)=0)f^T_{\tau+\ell}(t)g_\tau(\ell)\,\mathrm{d}m(\ell)\end{aligned}$$

□

We finish with an interpretation of this integral equation. The first split in the reduced process could actually happen at the time of the ancestor's death, this happens only if both the offspring, which the ancestor must have, produce extant progeny, this happens with probability

$$1-\Big(q_0(t)+q_2(t)\,\mathbb{P}_t(Z(T)=0)^2+2\,q_2(t)\,\mathbb{P}_t(Z(T)=0)\,\mathbb{P}_t\Big(Z^T(t)=1\Big)\Big)$$

as we here subtract from one the probability of the complementary event that the ancestor either has no offspring, has two offspring but both have no extant progeny or has two offspring and one of them has no extant progeny. In this case, where the death time of the ancestor is actually the splitting time in the reduced process, the splitting time has density $g_\tau(t-\tau)$ as this is the density of the death time of the ancestor.

If the splitting time is not at the ancestor's death time the ancestor must have died at some time before t, i.e. with an age ℓ less than $t-\tau$, at this point it must produce two offspring one of which cannot have any extant progeny, while the reduced process started with the other individual has its first split at time t with density $f^T_{\tau+\ell}(t)$. Which of the ancestor's offspring takes which path doesn't matter, hence the factor of 2.

Reduced binary branching trees 4

4.1 Phylogenetic trees 54
4.2 Marked Neveu trees . . . 59
4.3 The reduced branching tree 65

As has been hinted at several times up until now, there is a deep connection between branching processes and trees. This connection was already seen partly in Chapter 2 where we defined certain subsets of individuals in a branching population to form a Neveu tree, and noted that each element in our population sample space can be thought of as a tree of the realized individuals marked with their lifetimes. These notions were however left when characterizing the classical and reduced branching processes in Chapter 3, where only the number of individuals satisfying some rule over time was of interest. In this chapter, we will return to investigate trees as formed by branching populations, and in particular, we will study the analog to the reduced branching process for branching trees, namely, the *reduced branching tree*, which is the ancestral tree of the extant individuals, not containing any individuals without extant progeny.

This path of endeavor is clearly inspired by the evolution of phylogenetic trees. Where we in the last chapter extended a classical model for the number of ancestral lineages through time, we will here extend the classical models of phylogenetic trees to be based on our reduced time-varying Bellman-Harris process instead of e.g. the Yule or birth-death process. Another aim of this chapter, as with the last, is to formalize and partly generalize the results of [22, 41], in particular the tree density proposed in Section 4.2 of [41] and in Section 3.1 of [22].

[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'
[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

We start with an introduction to trees in general with a particular focus on the type of trees we want to model. Then we turn to construct a measure space of the considered trees, such that we can define them as random variables with values in this space, and such that we, in the last section, can consider the distribution of the reduced branching trees, and in particular find a density function for them.

As in the last part of Chapter 3, we will, silently, only work with binary branching processes, but in contrast to that chapter we will actually work on the binary Ulam-Harris space, and not just restrict the offspring process to be binary.

This chapter contains several constructions and results, which to the best knowledge of the author, are novel, of most importance are

- The dominating measure constructed on marked tree space.
- The construction of the reduced branching tree as a tree of least common ancestors.
- The density function of the reduced branching trees, with respect to the aforementioned dominating measure.

The introductory section on phylogenetic trees is primarily based on [1, 6] and [82].

[1]: Abraham et al. (2020), *An introduction to Galton-Watson trees and their local limits*
[6]: Andersen (2023), 'Models for the evolution of a phylogenetic tree'
[82]: Steel (2016), *Phylogeny*

4.1 Phylogenetic trees

In Chapter 2 we defined a certain subset of the Ulam-Harris space to be a Neveu tree if its individuals satisfied a set of characteristics, and we have noted that such a tree can be marked with lifetimes of the individuals to represent both the genealogy and temporal component of the population's evolution. Instead of thinking of trees as being subsets of the Ulam-Harris space, they are usually studied using *graph theory*, but we will see that our Neveu trees are just a specific type of graph theoretic tree called a *planted oriented tree*, which in turn is a type of *rooted tree*, which we can endow with edge lengths to represent the lifetimes of each individual. We start with a description of the trees that we actually want to model, and how to define them as random variables on the population space.

4.1.1 Phylogenetic trees as observed and modeled

As discussed loosely in Chapter 1, there exists a plethora of different trees analyzed in the field of phylogenetics, each with the possibility of answering its own set of questions. We will primarily be interested in the *reduced phylogenetic tree* sometimes also called the *reconstructed* or *ultrametric phylogenetic tree*, which we initially will characterize from the *complete phylogenetic tree*[1].

1: We do not necessarily think of a phylogenetic tree as the strict mathematical object defined e.g. in [82], but rather the general concept of a tree consisting of species and their ancestral relations.

[31]: Hedges et al. (2009), *The timetree of life*

Remark 4.1.1 Phylogenetic trees are not per se defined as having edge lengths in all cases, we will however mostly treat so-called phylogenetic time-trees [31] where the time each species in the tree lived is marked on the tree by its length, as these are the trees intrinsically produced by our branching population. To minimize the number of adjectives we have to say every time we refer to a tree, we will often leave out the time component from the naming, but certainly still have it in the definition. If we strictly talk about a phylogenetic tree without edge lengths we say a *tree shape*.

Consider a tree[2] consisting of extant and extinct species, up to some observation point, all stemming from a common ancestor. This *complete phylogenetic tree* describes the full evolutionary trajectory starting at the common ancestor, even with species that are not extant at the observation time. As it is often not possible to gain information about the species that have gone extinct before observation, it is a quite uncommon tree to observe, and biologists would in most situations only have access to the tree in which all extant individuals have been deleted, this is the *reduced phylogenetic tree* (see Figure 4.1). The reduced phylogenetic tree is in other words the tree containing the least common ancestors of each clade and their relations to each other, and it is important to note that the reduced phylogenetic tree in itself does not contain any information about what happened between each of these least common ancestors, e.g. we do not know how many species that have since gone extinct each of these common ancestors have produced.

2: For now, just the loosely defined concept of a (phylogenetic) tree, we return to a rigorous definition below.

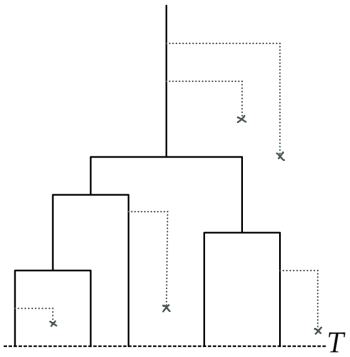


Figure 4.1: A reduced phylogenetic tree observed at time T with extinct individuals, that have no extant progeny, illustrated with dashed lines.

Connecting this to our foundational approach of branching populations, consider a population (of species) originating at time $\tau \in \mathbb{R}$ and look at the random set of individuals born before time $T \geq \tau$, in Section 2.2 denoted

by Υ_T, this forms the complete phylogenetic tree of the population up until time T. The reduced phylogenetic tree should in contrast to the complete tree not contain any extinct species that do not leave any extant progeny, and in particular, if we for any extant individual $x \in \mathfrak{X}_T$ traced back all of its ancestors, and combined their lifetimes, this should sum to $T - \tau$, which is also why some authors call it an ultrametric tree.

This last point could be read as assuming that the death time of all the extant individuals is exactly at time T, but that they haven't yet produced any offspring, essentially making them non-extant, which with our assumptions about reproduction is certainly nonsense. Instead what we mean is that the lifetime of the extant individuals are *censored*, and can from our coming definition of a phylogenetic tree never be observed.

To get the shape of the reduced tree, one might naturally suggest to take this tree and deleting all individuals not having any extant progeny, i.e. the ancestors of the extant individuals $\operatorname{An}\mathfrak{X}_t$, but this suggestion is troublesome in three regards. First, the set of ancestors does not necessarily constitute a Neveu tree, as we could have two realized individuals $x1$ and $x2$ where $x1$ produce no extant progeny, but $x2$ does, so the labeling the ancestor set would not actually be a proper sibship ranking (see Figure 4.2). This could maybe be fixed by some appropriate relabeling of each individual, as the labeling in itself is not of any importance and as the other conditions for being a Neveu tree are satisfied, but this is seemingly not a trivial task to formalize.

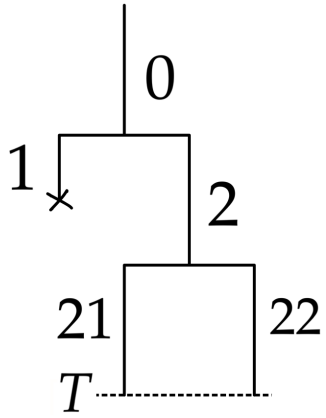


Figure 4.2: A complete tree, e.g. an observation of Υ_T, in which the Neveu labeling would not work if we deleted the extinct individuals.

Besides the labeling issue, the ancestral set will possibly contain individuals that we would not observe in a reduced phylogenetic tree, in particular, those individuals that produced offspring, but only one of which themselves produced any extant progeny. That is, the set, seen as a tree, might contain two or more sequential edges that are only connected to each other and no other individuals, and we would in particular have several splits before the observed split as the least common ancestor of all the extant individuals. One might suggest identifying all individuals between each split at the least common ancestors, and this is a good idea, but will not solve the last, and for our probabilistic branching framework, most damning problem with this "definition" of the reduced tree.

Working from the ancestor 0,[3] we could identify all the individuals between this and the least common ancestor of all the extant individuals. Now to do any distributional analysis on a random object defined from our branching population, we will have to get some recursive characterization using either the ordinary or the strong branching property of Theorems 2.4.1 and 2.4.2. The natural line on which to condition is then the immediate children of the least common ancestor of all extant individuals, and then have a distribution of the full reduced tree given as some combination of the distribution of what happened between 0 and the least common ancestor (which would be measurable with respect to what we condition on), and the distribution of the resulting trees of the progeny processes given by the branching process. But to use the branching process we need either a fixed line of individuals or an optional line, neither of which the children of the least common ancestor is, as the least common ancestor is very dependent on the future of the process. It might be possible to prove the branching property for such specific random lines which are not optional, but the author have not succeeded in doing so.

3: This is not necessarily a least common ancestor of any clade of the extant individuals.

Instead, we return to the second heuristic construction of the reduced phylogenetic tree from above, where we look at each branch[4] as representing the least common ancestor of one or another subset of the extant individuals,[5] and treating these as being the individuals of a branching population themselves. The very nature of being a least common ancestor means that it must have at least two offspring, and exactly two in the binary population case, so if we can characterize the lifetime distribution of this least common ancestor, we have a new well-defined branching population, for which we absolutely can use the ordinary branching property on the immediate potential children of the least common ancestor as this the fixed line $\mathbb{N}$, or in the binary case $\{1, 2\}$. Fortunately, we characterized the distribution of the time of splitting of the least common ancestor in Section 3.5, which if we know the origin time of the population characterizes the lifetime distribution.

4: Or equivalently each node, see Subsection 4.1.2.

5: This in itself is just the identification of individuals from before

Before a formal definition of this ancestor population and the reduced branching tree on it is apt, we have to expand our theory of Neveu trees a bit and set them into the context of general graph-theoretical trees.

4.1.2 Graph theory and trees

A *tree* or as we will sometimes call it a *tree shape* is a *graph* (V, E) with a finite or countable set of *vertices* V and *edges* $E \subseteq V \times V$, in which any two vertices are connected by a unique *path*, i.e. if for any two $v_1, v_n \in V$ there exists a unique sequence of distinct vertices $v_2, ..., v_{n-1} \in V$ such that $\{v_1, v_2\}, \{v_2, v_3\}, ..., \{v_{n-1}, v_n\} \in E$. If the tree has a singled-out vertex $\rho \in V$ called the *root*, then we have to do with a *rooted tree*. We will say that an edge $\{u, v\} \in E$ in a rooted tree is directed from u towards v if the unique path from ρ to v contains u. Each vertex in a rooted tree then has an *in-degree* and an *out-degree* counting the number of edges directed towards and from the vertex, respectively.

We can then divide V into the set of *interior vertices* denoted by $\mathring{V}$, those who have non-zero out-degree, and *leaves* denoted by $\check{V}$, which have out-degree 0. If all non-leaves have an out-degree of exactly 2, the tree is *binary*.[6] A subgraph of a tree, in which any two vertices are connected by a unique path, is called a *subtree*, note that any subtree of a rooted tree is rooted at the unique vertex contained in the path from the original root to any vertex in the subtree.

6: Many authors allow for a binary tree to have interior vertices with out-degree 1 also, but we stick to this stricter definition.

Note that while we have and will often draw trees (in the plane $\mathbb{R}^2$) to represent them, this is not usually a unique representation. If such a drawing contains no crossing edges, it is formally called a *planar embedding*, and for any tree such an embedding exists,[7] but may not be unique. A planar embedding of a rooted tree is equivalent to augmenting the vertices with an *orientation*, which gives a rule for where to draw each vertex, and thus each edge.

7: For general graphs we have no guarantee of the existence of a planar embedding.

A classical result of combinatorics (see e.g. [81, Example 5.3.12]) is that there exist

$$C_{n-1} := \frac{(2n-2)!}{n!(n-1)!}$$

oriented rooted trees with n leaves.[8] We now define a *symmetry vertex*

[81]: Stanley (1999), *Enumerative Combinatorics*

8: $C_n := \frac{(2n)!}{(n+1)!n!}$ is the n'th Catalan number

of a binary rooted tree (or any extension of such) to be a vertex with out-degree of two, such that both subtrees rooted each of these are indistinguishable. By indistinguishable, we mean that there should be a natural bijection between the two graphs, which does not alter any additional characteristic we might have endowed each vertex or branch with. For example, an oriented tree has no symmetry vertices, as we can always distinguish the two subtrees of any vertex by their orientation. This notion of symmetry vertices allows us for a given tree shape, to count the number of oriented trees that have that exact tree shape. If we by s denote the number of symmetry vertices in our tree shape with n leaves, there are 2^{n-1-s} oriented trees with that tree shape. For more details on this see [6, 82].

[6]: Andersen (2023), 'Models for the evolution of a phylogenetic tree'
[82]: Steel (2016), *Phylogeny*

A possible choice of orientation for each $v \in V$ with out-degree k is to label each vertex that has an incoming edge from v by $v1, v2, ..., vk$, starting from the root ρ, which we might label 0, and continuing this labeling for all vertices, we get that any vertex has the labeling $0v_1v_2...v_k$ for some $k \in \mathbb{N}_0$, i.e. we can identify each vertex with an element of the Ulam-Harris space. If the tree is binary, we can make do with the binary Ulam-Harris space. Note that the conditions of a subset of Ulam-Harris space for it to be a Neveu tree $\gamma \in \mathcal{N}_\infty$ are created exactly such that the graph $(\gamma, \{\{x, y\} \in \gamma \times \gamma : \exists n \in \mathbb{N} : y = xn\})$ is a rooted tree.

Our treatment of Neveu trees up until now has not specified the edges of the tree, due to them being given intrinsically by the vertices' labeling, and since we can identify each edge with its incoming edge. This doesn't fit fully into our definition of a rooted tree as the root has in-degree 0, and thus cannot have any incoming edge. This is mitigated by introducing *planted trees*, in which we add a "pseudo-edge" going into the root, but not coming from any other vertex. What we finally have is that a Neveu tree is a *planted oriented tree* (V, E) where $V = \gamma$ and $E \cong \gamma$.

Remark 4.1.2 We have seen that in a Neveu tree we can identify the vertices and edges, furthermore, these also represent individuals of the population (which we often interpret as being species or ancestral species) and thus branches (which of course is just another name for an edge) of a phylogenetic tree. We will use all of these terms interchangeably.

4.1.3 Binary Neveu trees

With some graph theory at hand, we return to Neveu trees as defined in Subsection 2.1.2, in the special case of a binary Ulam-Harris population, and in particular to define some operations on the space of binary Neveu trees. Much of the theory below will hold for trees on the full Ulam-Harris space, but binary trees will simplify some results and are all we need. Adjusting the former definition just slightly, we have, $\gamma \subseteq I$ is a *binary Neveu tree* if the following conditions are met

- $0 \in \gamma$
- $x \in \gamma \implies mx \in \gamma$
- $y2 \in \gamma \implies y1 \in \gamma$

and if each generation is finite, we call the binary Neveu tree *locally finite*, that is,

- $\#(\gamma \cap \{1,2\}^n) < \infty$ for all $n \in \mathbb{N}_0$.

Any binary Neveu tree is locally finite as the possible offspring is always limited to at most two. Thus we will now only work with the two spaces of, respectively, all Neveu trees $\mathcal{N} = \mathcal{N}_\infty$, and of finite Neveu trees $\mathcal{N}_0$.

For the treatment of the reduced branching trees, a few notational instruments and operations will be helpful. Let $\gamma \in \mathcal{N}$, the n'th generation of a tree is denoted γ_n, that is

$$\gamma_n := \gamma \cap \{1,2\}^n$$

and note that this might be an empty set, in particular, it is so if the branching population has gone extinct before generation n. A tree cut at height n is given by

$$\gamma_{|n} := \{x \in \gamma : \mathfrak{g}(x) \leq n\}$$

which on the contrary cannot be empty as $0 \in \gamma_{|n}$ for all $n \in \mathbb{N}_0$. We note that the $\gamma_{|0} = \gamma_0 = \{0\}$ for all $\gamma \in \mathcal{N}$.

Analogously to the shift operator, S_x we have defined on $(\overline{\Omega}, \overline{\mathcal{A}})$ for each $x \in I$, an even more fundamental shift operator works directly on the Ulam-Harris space or a Neveu subset of it. The shift operator on a Neveu tree[9] $\gamma \in \mathcal{N}$ called $\mathcal{S}_x$ for each $x \in \gamma$ treats the rooted subtree under x as its own Neveu tree, i.e.

9: Recall that the Ulam-Harris space is itself a Neveu tree.

$$\mathcal{S}_x(\gamma) := \{y \in I : xy \in \gamma\}$$

We remark that there naturally is a connection between the two shift operators, for any $x \in I$ we have

$$S_x(\varpi) = \left(\tau_x(\varpi), \{\omega_{xy} : y \in \mathcal{S}_x(I)\}\right).$$

A special operator for assembling trees on $\mathcal{N}$ is handy when working with trees generated by branching populations. Consider a finite tree $\gamma \in \mathcal{N}_0$ with non-empty leaf set $\check{\gamma}$, then for each $x \in M \subseteq \check{\gamma}$ let $\gamma_x \in \mathcal{N}$, we defined the *grafted tree* at M as

$$\gamma \circledast_{x \in M} \gamma_x := \gamma \cup \bigcup_{x \in M} \{xy : y \in \gamma_x\}$$

note that if $\gamma_x = \{x\}$ for each $x \in M$ the grafting does nothing, i.e. $\gamma \circledast_{x \in M} \gamma_x = \gamma$.

Together, the notion of shifted trees and grafting gives us a useful decomposition of a given tree $\gamma \in \mathcal{N}$, namely as the tree up to some generation n grafted with the shifted trees for each individual of generation $n+1$,

$$\gamma = \gamma_{|n} \circledast_{x \in \gamma_{n+1}} \mathcal{S}_x(\gamma) \tag{4.1}$$

for all $n \in \mathbb{N}$.

4.2 Marked Neveu trees

All of these constructions are working on discrete genealogical (Neveu) trees, but as noted in Remark 4.1.1, we are fundamentally more interested in trees that carry some temporal component of the evolution of the population. In particular, will it be interesting to consider random Neveu trees where we assign each individual with a branch length corresponding to their (censored) lifetime, and we will do so by introducing *marked Neveu trees* and constructing a measure space over such trees, such that we can eventually define random objects with values in this space and study their distribution.

A *marked tree* is a Neveu tree γ augmented with a sequence of positive *marks*[10] $\{\ell_x\}_{x\in\gamma} \in \mathbb{R}_+^\gamma$, one for each individual in the tree

$$\gamma^* := \left(\gamma, \{\ell_x\}_{x\in\gamma}\right)$$

10: In theory we could mark each individual with a more general object than a real number, but a real number is exactly what we need for the further development.

We should think of the marks as some characteristic of the individual, and for our purposes, it should of course reflect some temporal component of the population. With the marks all being positive real numbers, we can graphically represent a marked tree as the planted, binary-oriented tree[11] with edge lengths given by the marks.

11: Which has a unique planar embedding.

The natural choice of the marks in our case is of course the (censored) lifetime of each individual, and this will be our primary example, but for now, we just work with a generic mark. Wanting to represent the reduced branching tree as a marked Neveu tree, we could also have chosen to consider only marks such that along each line of descend they would sum to $T - \tau$ (for a population initiated at time τ and observed at time T), but we will not make this restriction on the space either, leaving this to be a restriction on the measures on the space.

The space of all marked trees is defined as,

$$\mathcal{N}^* := \bigcup_{\gamma\in\mathcal{N}} \left\{\left(\gamma, \{\ell_x\}_{x\in\gamma}\right) : \{\ell_x\}_{x\in\gamma} \in \mathbb{R}_+^\gamma\right\} = \bigcup_{\gamma\in\mathcal{N}} \{\gamma\} \times \mathbb{R}_+^\gamma$$

If only finite trees are of interest, we make the analog definition of $\mathcal{N}_0^*$,

$$\mathcal{N}_0^* := \bigcup_{\gamma\in\mathcal{N}_0} \left\{\left(\gamma, \{\ell_x\}_{x\in\gamma}\right) : \{\ell_x\}_{x\in\gamma} \in \mathbb{R}_+^\gamma\right\} = \bigcup_{\gamma\in\mathcal{N}_0} \{\gamma\} \times \mathbb{R}_+^\gamma$$

Note that the Neveu tree operations defined in Subsection 4.1.3 all make sense for marked trees as well, by doing the operation on the underlying Neveu tree and only considering the relevant marks. In particular, we have the fundamental decomposition at generation $n \in \mathbb{N}_0$ of marked trees,

$$\gamma^* = \gamma^*_{|n} \circledast_{x\in\gamma_{n+1}} \mathcal{S}_x(\gamma^*) \tag{4.2}$$

In the following, we will only consider finite marked trees as this space should be vast enough to contain any binary random tree evolved in finite time under our assumptions that each individual only has a finite number (often binary) of offspring and that each individual has a strictly

positive lifetime, and there hence cannot be any *explosions* of births at any single moment.

4.2.1 A σ-algebra of finite marked trees

To treat random finite trees we must equip the space $\mathcal{N}_0^*$ with a σ-algebra, and to study distributions with density on this space, we will furthermore equip it with a suitable dominating measure. As a marked tree contains a discrete part and a continuous part, some apparent trouble arises in that the dimension of the continuous part, i.e. the marks, are dependent on the discrete part, so a simple product σ-algebra of all subsets of Neveu trees and some finite-dimensional Borel algebras are not sufficient, and a more intricate combination of the discrete and continuous part is necessary.

We propose the following collection of sets as an appropriate σ-algebra,

$$\mathcal{E} := \left\{ \bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma : B_\gamma \in \mathcal{B}_\gamma \text{ for each } \gamma \in \mathcal{N}_0 \right\}$$

where $\mathcal{B}_\gamma$ is the Borel σ-algebra on $\mathbb{R}_+^\gamma$. Elements $E \in \mathcal{E}$ are then of the form

$$E = \bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma = \bigcup_{\gamma \in \Gamma} \{\gamma\} \times B_\gamma$$

for some chosen B_γ and $\Gamma \subseteq \mathcal{N}_0$ such that $B_\gamma = \emptyset$ for all $\gamma \notin N$. Hence even though a general element of $\mathcal{E}$ might seem more complicated than necessary, simple subsets of $\mathcal{N}_0^*$ like a singleton fixed tree with fixed marks are in the collection if we fix a $\gamma \in \mathcal{N}_0$ and choose $\Gamma = \{\gamma\}$ with $B_\gamma = \{\ell_x\}_{x \in \gamma} \in \mathcal{B}_\gamma$. It is important to note that each element $E \in \mathcal{E}$ is a disjoint union of elements $E_\gamma = \{\gamma\} \times B_\gamma$ for all $\gamma \in \mathcal{N}_0$ as for two $\gamma, \gamma' \in \mathcal{N}_0$ the intersection $E_\gamma \cap E_{\gamma'}$ is only non-empty if $\gamma = \gamma'$.

It is not entirely evident that $\mathcal{E}$ actually constitutes a σ-algebra on $\mathcal{N}_0^*$, hence we make the argument,

Proposition 4.2.1 *$\mathcal{E}$ is a σ-algebra on $\mathcal{N}_0^*$.*

[47]: Klenke (2020), *Probability Theory*

Proof. We check the three criteria for a σ-algebra [47, Definition 1.2], namely that it contains $\mathcal{N}_0^*$, is closed under complements and under countable unions.

It is evident that $\mathcal{N}_0^* \in \mathcal{E}$ by choosing $B_\gamma = \mathbb{R}^\gamma$ for each $\gamma \in \mathcal{N}_0$. To show that $\mathcal{E}$ is closed under complements we take a set $E \in \mathcal{E}$, i.e. we take arbitrary $B_\gamma \in \mathcal{B}_\gamma$ for each $\gamma \in \mathcal{N}_0$ and look at the complement of this E in $\mathcal{N}_0^*$,

$$E^\complement = \mathcal{N}_0^* \setminus E = \left(\bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times \mathbb{R}^\gamma \right) \setminus \left(\bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma \right) = \bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma^\complement \in \mathcal{E}$$

as $B_\gamma^\complement \in \mathcal{B}_\gamma$ by $\mathcal{B}_\gamma$ being a σ-algebra for all $\gamma \in \mathcal{N}_0$.

For the last criterion, we choose a sequence $(E_n)_{n\in\mathbb{N}}$ of sets in $\mathscr{E}$, and by doing that also a sequence $(B^n_\gamma)_{(\gamma,n)\in\mathcal{N}_0\times\mathbb{N}}$ such that for each $\gamma\in\mathcal{N}_0$ we have $B^n_\gamma\in\mathscr{B}_\gamma$ for all $n\in\mathbb{N}$. Let us thus look at the union of the E_n's

$$\begin{aligned}\bigcup_{n\in\mathbb{N}} E_n &= \bigcup_{n\in\mathbb{N}}\left(\bigcup_{\gamma\in\mathcal{N}_0}\{\gamma\}\times B^n_\gamma\right)\\ &= \bigcup_{\gamma\in\mathcal{N}_0}\left(\bigcup_{n\in\mathbb{N}}\{\gamma\}\times B^n_\gamma\right)\\ &= \bigcup_{\gamma\in\mathcal{N}_0}\{\gamma\}\times\underbrace{\bigcup_{n\in\mathbb{N}} B^n_\gamma}_{\in\mathscr{B}_\gamma}\in\mathscr{E}\end{aligned}$$

again using that $\mathscr{B}_\gamma$ is a σ-algebra for each $\gamma\in\mathcal{N}_0$. We have thus shown that $\mathscr{E}$ is a σ-algebra on $\mathcal{N}_0^*$. □

As so often with σ-algebras, their elements are not quite as accessible as we would like, but for many arguments, it suffices to consider elements of a generator of the σ-algebra. As $\mathscr{B}_\gamma$ is a Borel σ-algebra, the following quite natural class of sets comprises a generator for $\mathscr{E}$

Lemma 4.2.2 *Consider the class of sets with the Borel set for each $\gamma\in\mathcal{N}_0$ being a Borel rectangle,*

$$\mathscr{G}_{\mathscr{E}} := \left\{\bigcup_{\gamma\in\mathcal{N}_0}\{\gamma\}\times\bigtimes_{x\in\gamma} B^x_\gamma : B^x_\gamma\in\mathscr{B}\text{ for each }\gamma\in\mathcal{N}_0, x\in\gamma\right\}$$

It holds that $\mathscr{G}_{\mathscr{E}}$ is $\cap$-stable, contains $\mathcal{N}_0$ and that $\sigma(\mathscr{G}_{\mathscr{E}})=\mathscr{E}$.

4.2.2 A dominating measure on marked tree space

With a measurable space at hand, the natural next step is of course to define measures on it. Certainly, if Υ^* is a random variable with values in $(\mathcal{N}_0^*,\mathscr{E})$ then the distribution of Υ^* must be such a measure. We will in the coming sections, when we have defined the reduced branching tree, find the distribution of this tree and in particular characterize it through a density function, where the density should of course be with respect to some dominating measure on $(\mathcal{N}_0^*,\mathscr{E})$. This measure should take into account the specific temporal features of the reduced branching tree as discussed, in particular, must the marks along each line of descent sum to $T-\tau$ (for a population initiated at time τ and observed at time T), as we will treat them as being (censored) lifetimes.

Initially, we will define a $\mathcal{N}_0$-kernel on $(\mathcal{N}_0^*,\mathscr{E})$, $(\mu_\gamma)_{\gamma\in\mathcal{N}_0}$, which for a given $E\in\mathscr{E}$ only measures the branch length sets (i.e. the continuous part) for the part of E that has to do with a specific Neveu tree. From this kernel we will eventually define the measure μ on $(\mathcal{N}_0^*,\mathscr{E})$ as the integration measure of the kernel with respect to the counting measure $c_{\mathcal{N}_0}$ on the set of finite trees, i.e.

$$\mu(E) := c_{\mathcal{N}_0}\otimes(\mu_\gamma)_{\gamma\in\mathcal{N}_0} = \sum_{\gamma\in\mathcal{N}_0}\mu_\gamma(E).$$

Remark 4.2.1 One should maybe for clarity super- or subscript these kernels and measures by the origin and observation times, as they are heavily dependent hereon, in fact, one could furthermore consider them as kernels indexed by these times as well. We will however consider $\tau \in \mathbb{R}$ and $T \geq \tau$ as fixed whenever we are working with this measure, so for notational brevity, we leave them being implicit.

Initially fix $\gamma \in \mathcal{N}_0$ and let us consider how to construct the measure μ_γ. Each mark/branch length of $\gamma^* = (\gamma, \ell_\gamma)$ is certainly very dependent as described, so also for the individual measure on each branch will we construct a kernel, taking into account the individual's ancestors' branch lengths, akin to our fundamental construction of the measures on $\left(\Omega^I, \mathscr{A}^I\right)$ in Proposition 2.3.1.

We will hence for each $x \in \gamma$ define a $(\mathbb{R}_+^{\mathrm{An}\, x}, \mathscr{B}_{\mathrm{An}\, x})$-kernel κ_γ^x on $\left((0, T-\tau], \mathscr{B}_{(0,T-\tau]}\right)$. Note that κ_γ^0 is always a measure, not a kernel, for given $\gamma \in \mathcal{N}_0$ as 0 has no ancestors.[12] The definition of each is given in the following, but certainly, the kernel on each internal branch should be some restricted Lebesgue measure and for the leaf branches be a Dirac measure[13] with the appropriate point of concentration determined by the branch lengths of the ancestors of that branch.

12: It is certainly a kernel if we let γ vary, but this consideration is saved for later.

13: We denote the Dirac measure on $(\mathbb{R}, \mathscr{B})$ with concentration at $t \in \mathbb{R}$ by δ_x. I.e. for all $B \in \mathscr{B}$ we have

$$\delta_x(B) = \mathbf{1}_B(x).$$

Consider first the special case of $\gamma = \{0\}$, here we only have one mark, and thus no dependencies to take into account, but we do have to restrict what the length of the branch can be, as it must be exactly $T - \tau$. Hence we define the measure $\kappa_{\{0\}}^0$ for all $B \in \mathscr{B}_+$

$$\kappa_{\{0\}}^0(B) := \delta_{T-\tau}(B).$$

we will thus identify $\mu_{\{0\}}(E) := \kappa_{\{0\}}^0(B_{\{0\}})$ with $E = \bigcup_{\gamma' \in \mathcal{N}_0} \{\gamma'\} \times B_{\gamma'}$.

For any other $\gamma \neq \{0\}$, we define for the ancestor 0 the measure for all $B \in \mathscr{B}_{(0,T-\tau]}$,

$$\kappa_0^\gamma(B) := m\left(B \cap (0, T-\tau)\right)$$

and for all other internal branches $x \in \mathring{\gamma}$ we let

$$\kappa_x^\gamma(\ell_{\mathrm{An}\, x}, B) := m\left(B \cap \left(0, \sum_{y<x} \ell_y\right]\right)$$

and for all the leaf branches $x \in \check{\gamma}$ we give it as

$$\kappa_x^\gamma(\ell_{\mathrm{An}\, x}, B) := \delta_{T-\sum_{y<x} \ell_y}(B).$$

We will then define μ_γ as the integration measure of the product kernel of all of these ordered under the linear order $<$, let us call it κ_γ, evaluated in the branch length sets associated with γ. The order of the product kernels is important, as we want all the necessary branch length information available to the right kernels, i.e. if we denote the maximal element of γ under the linear order $<$ by v, and the maximal element of $\mathring{\gamma}$ by u, we

have

$$\begin{aligned}\mu_\gamma(E) &:= \kappa_\gamma(B_\gamma) := \kappa_\gamma^0 \otimes \left(\bigotimes_{x=1}^{v} \kappa_\gamma^x\right)(B_\gamma) \\ &= \int \int \cdots \int \mathbf{1}_{B_\gamma}(\ell_\gamma)\, \mathrm{d}\kappa_\gamma^v(\ell_{\mathrm{An}\, v}, \ell_v) \cdots \mathrm{d}\kappa_\gamma^1(\ell_0, \ell_1)\, \mathrm{d}\kappa_\gamma^0(\ell_0) \\ &= \int_{(0,T-\tau)} \int_{(0,T-\tau-\ell_0)} \cdots \int \mathbf{1}_{B_\gamma}(\ell_\gamma)\, \mathrm{d}\delta_{T-\sum_{y<v}\ell_y}(\ell_v) \cdots \mathrm{d}m(\ell_1)\, \mathrm{d}m(\ell_0) \\ &= \int_{(0,T-\tau)} \cdots \int_{(0,T-\tau-\sum_{y<u}\ell_y)} \mathbf{1}_{B_\gamma}(\tilde{\ell}_\gamma)\, \mathrm{d}m(\ell_u) \cdots \mathrm{d}m(\ell_0)\end{aligned}$$

where $\tilde{\ell}_x = \ell_x$ if $x \in \mathring{\gamma}$ and $\tilde{\ell}_x = T - \tau - \sum_{y<x} \ell_y$ if $x \in \breve{\gamma}$, and where we have used that integrating with respect to a Dirac measure is the same as evaluating the integrand in the point of concentration. By [47, Corollary 14.26] this actually defines a measure for given $\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}$.

[47]: Klenke (2020), *Probability Theory*

With μ_γ given for all $\gamma \in \mathcal{N}_0$, we define the integration measure of the kernel with respect to $c_{\mathcal{N}_0}$, to get our final measure on $(\mathcal{N}_0^*, \mathcal{E})$,

$$\mu(E) := c_{\mathcal{N}_0} \otimes (\mu_\gamma)_{\gamma \in \mathcal{N}_0} = \sum_{\gamma \in \mathcal{N}_0} \mu_\gamma(E)$$

and this is actually a measure by [47, Corollary 14.26].

As our goal with the measure μ is to use it as a dominating measure for a distribution, and a density with respect to μ for that distribution, it will be convenient to know how to integrate with respect to μ.

Proposition 4.2.3 *Let $f : (\mathcal{N}_0^*, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}_+)$ be a non-negative function. The integral of f with respect to μ is*

$$\int f\, \mathrm{d}\mu = \sum_{\gamma \in \mathcal{N}_0} \int f\, \mathrm{d}\mu_\gamma = \sum_{\gamma \in \mathcal{N}_0} \int f(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma)$$

where both sides can equal ∞ and the integration with respect to κ_γ^T for $\gamma \neq \{0\}$ is given by

$$\begin{aligned}&\int f(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\ &\quad = \int_{(0,T-\tau)} \cdots \int_{(0,T-\sum_{y<\max\gamma}\ell_y)} f(\gamma, \tilde{\ell}_\gamma)\, \mathrm{d}m(\ell_{\max\gamma}) \cdots \mathrm{d}m(\ell_0)\end{aligned}$$

with $\tilde{\ell}_x = \ell_x$ if $x \in \mathring{\gamma}$ and $\tilde{\ell}_x = T - \sum_{y<x} \ell_y$ for $x \in \breve{\gamma}$, while for $\gamma = \{0\}$ it is given by

$$\int f(\{0\}, \ell_0)\, \mathrm{d}\kappa_{\{0\}}^0(\ell_0) = f(\{0\}, T - \tau).$$

Proof. Initially, we see how to integrate indicator functions $\mathbf{1}_E$ for some

$E = \bigcup_{\gamma' \in \mathcal{N}_0} \{\gamma'\} \times B_{\gamma'} \in \mathcal{C}$ with $B_{\gamma'} \in \mathcal{B}_{\gamma'}$ for each $\gamma' \in \mathcal{N}_0$, we then have

$$\begin{aligned}
\int \mathbf{1}_E \, \mathrm{d}\mu = \mu(E) &= \sum_{\gamma \in \mathcal{N}_0} \mu_\gamma(E) = \sum_{\gamma \in \mathcal{N}_0} \kappa_\gamma\left(B_\gamma\right) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \mathbf{1}_{B_\gamma}(\ell_\gamma) \, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \mathbf{1}_{\{\gamma\} \times B_\gamma}(\gamma, \ell_\gamma) \, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \mathbf{1}_{\bigcup_{\gamma' \in \mathcal{N}_0} \{\gamma'\} \times B_{\gamma'}}(\gamma, \ell_\gamma) \, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \mathbf{1}_E(\gamma, \ell_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma)
\end{aligned}$$

[89]: Williams (1991), *Probability with Martingales*

Now the result for general, non-negative f follows by the *standard machine* [89, Chapter 5.12]. For non-negative simple functions, i.e. if $f_n = \sum_{k=1}^n e_k \, \mathbf{1}_{E_k}$ for $e_k \geq 0$, $E_k \in \mathcal{C}$ for all $k \leq n \in \mathbb{N}$, we have

$$\begin{aligned}
\int f_n \, \mathrm{d}\mu &= \int \sum_{k=1}^n e_k \, \mathbf{1}_{E_k} \, \mathrm{d}\mu \\
&= \sum_{k=1}^n e_k \int \mathbf{1}_{E_k} \, \mathrm{d}\mu \\
&= \sum_{k=1}^n e_k \sum_{\gamma \in \mathcal{N}_0} \int \mathbf{1}_{E_k}(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(s_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \sum_{k=1}^n e_k \, \mathbf{1}_{E_k}(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int f_n(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma)
\end{aligned}$$

where the sums can be exchanged since we only work with non-negative terms.

And finally, any non-negative measurable function can be approximated by an increasing sequence of non-negative simple functions, i.e. there exists a sequence $(f_n)_{n \in \mathbb{N}}$ of non-negative simple functions such that $f_n \uparrow f$, so by the monotone convergence theorem we have

$$\begin{aligned}
\int f \, \mathrm{d}\mu &= \int \lim_{n \to \infty} f_n \, \mathrm{d}\mu \\
&= \lim_{n \to \infty} \int f_n \, \mathrm{d}\mu \\
&= \lim_{n \to \infty} \sum_{\gamma \in \mathcal{N}_0} \int f_n(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \lim_{n \to \infty} \int f_n(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int \lim_{n \to \infty} f_n(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma) \\
&= \sum_{\gamma \in \mathcal{N}_0} \int f(\gamma, t_\gamma) \, \mathrm{d}\kappa_\gamma(t_\gamma)
\end{aligned}$$

We have used monotone convergence several times above, once to exchange the limit and integral, then to exchange the limit and the summation (integrals are monotone, so the integrals of an increasing sequence are increasing), and finally to exchange the limit and integral back into order.

The integration formula for each κ_γ follows from Tonelli's theorem for kernels [47, Theorem 14.32] and that integration with respect to the Dirac measure is evaluation of the integrand in the point of support of the measure. □

[47]: Klenke (2020), *Probability Theory*

With the measure space of finite marked trees at hand, we are ready to define the reduced branching tree whose distribution our goal is to characterize.

4.3 The reduced branching tree

Constructing the reduced branching tree is as described in Subsection 4.1.1, not as simple as letting it be the random set $\operatorname{An} \mathfrak{X}_T$, endowed with some marks. On the contrary, we have to start all over in our branching population construction as to define a new branching population, where we instead of thinking of each individual as a species, we think of it as a least common ancestor of some subset of the extant individuals, that is, we essentially take the set

$$\{\operatorname{lca} M : M \subset \mathfrak{X}_T\}$$

as our starting point and define a new branching population over all such possible sets. The important point here is that the set of least common ancestors will evolve in exactly the same form as our original branching population, where the least common ancestor of all the extant individuals will be labeled by 0, and this ancestor produce exactly two new ancestors labeled by 1 and 2 which are ancestors of each their subset of the extant individuals, and so on, independently of how a given realization of $\mathfrak{X}_T$ looks, as long as it is not the empty set. Thus we identify the set of all *potential ancestors* with the binary Ulam-Harris space. For clarity's sake, we denote this set of potential ancestors by $\check{I}$, even though it is completely identical to the set I, if it represents the binary Ulam-Harris space.

Now we can, just as done in Section 2.2, define the *common ancestor life space* $(\check{\Omega}, \check{\mathscr{A}})$ the *common ancestor population space* $(\check{\Omega}^{\check{I}}, \check{\mathscr{A}}^{\check{I}})$ and the *common ancestor time-population* space $(\underline{\check{\Omega}}, \underline{\check{\mathscr{A}}})$. On the common ancestor life space we can define the $(\mathbb{R}, \mathcal{B})$-Markov kernel[14] $(\Pi_\tau^T)_{\tau\in\mathbb{R}}$, which by Proposition 2.3.1 defines the $(\mathbb{R}, \mathcal{B})$-Markov kernel $(\mathbb{Q}_\tau^T)_{\tau\in\mathbb{R}}$ on the common ancestor population space, which by the arguments in Remark 2.3.1 can also be treated as a kernel on the time-population space. It is essential to note that as this construction is fully analogous to that of Section 2.3, the branching property of Theorem 2.4.1 will also hold for $(\mathbb{Q}_\tau^T)_{\tau\in\mathbb{R}}$.

14: See that Π_τ^T is also indexed by the observation time. One could see it as a kernel indexed both by τ and T, but we will not let T vary at any point, so this is not considered.

We will in general endow any variable or process that has already been defined on either the life, population, or time-population space of our original population, with a $\check{}$ to signify its version as defined on any of the common ancestor sample spaces, and if writing words about them we will prepose them with *ancestral* or *common ancestor*.

We will throughout this section only consider the original population which is initiated at time $\tau \in \mathbb{R}$, is observed at time $T \geq \tau$, and for which we assume that the population is not a.s. going extinct, i.e. $\mathbb{P}_\tau(Z(T) > 0) \neq 0$. If the original population did go extinct a.s., there would a.s. be no extant individuals, and hence no reduced tree to study, so this is not as such a restriction.

4.3.1 Ancestral lifetime distribution

Initially, we can consider the ancestral lifetime variable $\check{L}\colon (\check{\Omega}, \check{\mathscr{A}}) \to (\mathbb{R}_+, \mathscr{B}_+)$ and the ancestral offspring process $\check{N}(t)\colon (\check{\Omega}, \check{\mathscr{A}}) \to (\mathbb{N}_0, 2^{\mathbb{N}_0})$ for all $t \in \mathbb{R}$. Under the general assumptions above, we know that the least common ancestor of all extant individuals in the original population must die at time $\sigma_{\mathrm{lca}(\mathscr{X}_T)}$, so this should certainly also be the splitting time of the ancestors' ancestor 0 in $\check{I}$, which is also, by definition, born at time τ. Now as the splitting time $\check{\sigma}\colon (\check{\Omega}, \check{\mathscr{A}}) \to (\mathbb{R}, \mathscr{B})$ of an ancestral individual can be expressed from the common ancestor life space as $\check{\sigma} = \tau + \check{L}$ we have for the ancestral life law that

$$\check{L}(\Pi_\tau^T)(B) = \Pi_\tau^T(\check{L} \in B) := \mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} - \tau \in B \,\middle|\, Z(T) > 0\right)$$

for all $B \in \mathscr{B}_+$. From a measure-theoretic point of view, this makes sense as both $\check{L}(\Pi_\tau^T)$ and $(\sigma_{\mathrm{lca}(\mathscr{X}_T)} - \tau)(\mathbb{P}_\tau^{(Z(T)>0)})$ are measures on $(\mathbb{R}_+, \mathscr{B}_+)$. The distribution $\sigma_{\mathrm{lca}(\mathscr{X}_T)}(\mathbb{P}_\tau^{(Z(T)>0)})$ restricted to $\left([\tau, T), \mathscr{B}([\tau, T))\right)$ has been shown in Theorem 3.5.2 to have m-density f_τ^T over $[\tau, T)$.

Now $\sigma_{\mathrm{lca}(\mathscr{X}_T)}$ can possibly take values larger than or equal to T, but the specific value it takes will certainly not affect any of our random objects as these are only defined over $[\tau, T]$. Hence we can extend f_τ^T to all of $[\tau, \infty)$ by letting it have be some arbitrary integrable non-negative function over $[T, \infty)$ as long as it satisfies

$$\int_{[T,\infty)} f_\tau^T \,\mathrm{d}m = \mathbb{P}_\tau\left(\sigma_{\mathrm{lca}(\mathscr{X}_T)} \geq T \,\middle|\, Z(T) > 0\right) = \mathbb{P}_\tau\left(\sigma^T = T \,\middle|\, Z(T) > 0\right)$$

such that the full m-integral of f_τ^T is one, and it thus becomes a probability density. With this extension at hand, we can without any problems assume that the splitting time of a common ancestor born at time τ will have m-density f_τ^T and thus that the lifetime distribution has an m-density $f_\tau^T(\tau + \cdot)$ over $\mathbb{R}_+$, and is thus given by

$$\check{L}(\Pi_\tau^T)(B) = \int_B f_\tau^T(\tau + \ell) \,\mathrm{d}m(\ell)$$

for all $B \in \mathscr{B}_+$.

The distribution of the offspring process is much simpler. We know with certainty that any common ancestor at their splitting time produce exactly two offspring, otherwise, it would not be a least common ancestor of any extant individuals. Hence we simply have that $\check{q}_2(t) = \Pi_\tau^T(\check{N}(t) = 2) = 1$, and thus also that $\check{q}_0(t) = 0$ for all $t \in \mathbb{R}$. Consequently, the population will never go extinct, and any common ancestor $x \in \check{I}$ with a birth time $\check{\tau}_x \leq T$ will be realized, and we have no need for the *realized* version of any random variables or random sets as we had in Subsection 2.2.1.

4.3.2 Defining the reduced branching tree

With the ancestral branching population fully characterized, we can finally define the *reduced binary branching tree shape* $\check{\Upsilon}_T : (\underline{\check{\Omega}}, \underline{\check{\mathcal{A}}}) \to (\mathcal{N}_0, 2^{\mathcal{N}_0})$ being the set of (realized) common ancestors in $\check{I}$ up to time T, i.e. for all $\varpi \in \underline{\check{\Omega}}$

$$\check{\Upsilon}(\varpi) := \{x \in \check{I} : \check{\tau}_x(\varpi) \leq T\}$$

and we get the *reduced binary branching tree* with edge lengths $\check{\Upsilon}^* : (\underline{\check{\Omega}}, \underline{\check{\mathcal{A}}}) \to (\mathcal{N}_0^*, \mathscr{E})$ being the tree shape marked with the (censored) lifetimes of each common ancestor

$$\check{\Upsilon}^*(\varpi) := \left(\check{\Upsilon}(\varpi), \left\{\check{L}_x(\varpi) \wedge \left(T - \check{\tau}_0(\varpi) - \sum_{y \prec x} \check{L}_y(\varpi)\right)\right\}_{x \in \check{\Upsilon}(\varpi)}\right)$$

for all $\varpi \in \underline{\check{\Omega}}$. Note that we do not explicitly write these trees as being dependent on T, which they most certainly are, but once again, to save on the number of super- and subscripts, we leave it out.

Now this definition is much simpler than the faulty ones described in Subsection 4.1.1, and it contains exactly the right amount of individuals each having the correct labeling, and the correct branch length. Finally, if we want to consider its distribution, we can make use of the branching process, conditioning on the children of the least common ancestor of all extant individuals, by conditioning on the σ-algebra $\check{\mathcal{F}}_{\{1,2\}}$, and this is exactly what we will do in the next section.

4.3.3 The reduced tree distribution

We are now, at long last, ready to state the first characterization of the distribution of the reduced binary branching tree. This characterization is, however insightful it is into the actual dependency structure of the reduced tree, only an intermediate recursive formulation, which we will in the next section use to derive the μ-density of the reduced tree distribution.

Loosely, the following lemma shows that the distribution of the reduced branching tree is given as the distribution of the two descendant reduced branching trees. We then have to condition on the splitting time of the ancestor 0 in $\check{I}$, and integrate over it again, to take into account the difference in the distribution of the subtrees depending on their time of birth. There is of course also a contribution from the possibility that the ancestor has not died by observation, which is the only non-recursive part, acting like a *stopping condition*, so we only need to apply the recursion a finite number of times to get the full characterization of the distribution. The proof is a simple application of the basic tree decomposition of Equation 4.1, the branching property, and that probability measures are uniquely determined by their values on a $\cap$-stable generator.

Lemma 4.3.1 (The tree recursion) [15] *Let $\tau \in \mathbb{R}$ and $T \geq \tau$. Consider $E = \bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma \in \mathscr{G}_{\mathscr{E}}$ i.e. with $B_\gamma = \times_{x \in \gamma} B_\gamma^x$ for $B_\gamma^x \in \mathscr{B}_+$ for all $\gamma \in \mathcal{N}_0$ and $x \in \gamma$. The distribution $\check{\Upsilon}^*(\mathbb{Q}_\tau^T)$ is the unique probability measure*

15: This is a much more general result than as stated here; Any (marked) Neveu tree conditioned on any stopping line σ-algebra has a very similar recursive characterization of its distribution.

on $(\mathcal{N}_0^*, \mathscr{E})$ *satisfying the recursion*

$$\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) = \mathbb{Q}_\tau^T\left(\check{L}_0 > T - \tau\right)\mathbf{1}_{B^0_{\{0\}}}(T-\tau) + \sum_{\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}} \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0 \in B^0_\gamma \cap (0, T-\tau))}\mathbb{Q}^T_{\tau+\check{L}_0}\left(\check{\Upsilon}^* \in E^1_\gamma\right)\mathbb{Q}^T_{\tau+\check{L}_0}\left(\check{\Upsilon}^* \in E^2_\gamma\right)\right]$$

where $E^x_\gamma = \{\mathcal{S}_x(\gamma)\} \times \bigtimes_{y \in \mathcal{S}_x(\gamma)} B^y_\gamma$ *for all* $\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}$ *and* $x \in \gamma$ *is the set of shifted marked trees.*

Proof. We start by noting that E is a disjoint union of $\mathscr{E}$ measurable sets, so by σ-additivity of measures we initially have,

$$\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in \bigcup_{\gamma \in \mathcal{N}_0} \{\gamma\} \times B_\gamma\right) = \sum_{\gamma \in \mathcal{N}_0} \mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in \{\gamma\} \times B_\gamma\right) \tag{4.3}$$

The case in which $\gamma = \{0\}$ is special as the tree would contain only a single branch and the least common ancestor must then not have died by observation. We see that

$$\begin{aligned}\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in \{\{0\}\} \times B^0_{\{0\}}\right) &= \mathbb{Q}_\tau^T\left(\check{\Upsilon} = \{0\}, \tau \in B^0_{\{0\}}\right) \\ &= \mathbb{Q}_\tau^T\left(\check{\Upsilon} = \{0\}\right)\mathbf{1}_{B^0_{\{0\}}}(T-\tau) \\ &= \mathbb{Q}_\tau^T\left(\check{L}_0 > T - \tau\right)\mathbf{1}_{B^0_{\{0\}}}(T-\tau)\end{aligned}$$

where we use that $\check{\Upsilon}(\omega) = \{0\}$ can happen if and only if $\check{\tau}_x(\omega) > T$ for all $x \neq 0$ or equivalently if $\check{L}_0 > T - \tau$.

For all other Neveu trees $\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}$, we can use the basic tree decomposition of Equation 4.1 to see that for any $\omega \in (\check{\Upsilon}^* \in \{\gamma\} \times B_\gamma)$

$$\begin{aligned}\check{\Upsilon}^*(\omega) &= \check{\Upsilon}^*(\omega)_{|0} \circledast_{x \in \{1,2\}} \mathcal{S}_x\left(\check{\Upsilon}^*(\omega)\right) \\ &= \left\{\left(\{0\}, \{\check{L}_0(\omega)\}\right)\right\} \circledast_{x \in \{1,2\}} \check{\Upsilon}^*(\mathcal{S}_x(\omega))\end{aligned}$$

using that shifting the tree to have root at x must be the same as growing the tree from the lives of x's progeny, i.e. shifting ω. Using this decomposition, that the most recent common ancestor must have died before observation, and that any Neveu tree cut at generation 0 must be $\{0\}$ the event we are working on can be written as

$$(\check{\Upsilon}^* \in \{\gamma\} \times B_\gamma) = (\check{L}_0 \in B^0_\gamma \cap (0, T-\tau), \check{\Upsilon}^*(S_1) \in E_1, \check{\Upsilon}^*(S_2) \in E_2)$$

To get an expression for the probability of this event, we use the law of total expectation conditioning on $\mathcal{F}_{\{1,2\}}$, that $\check{L}_0$ is $\mathcal{F}_{\{1,2\}}$-measurable, and

the branching property (Theorem 2.4.1),

$$\begin{aligned}
&\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in \{\gamma\} \times B_\gamma\right) \\
&\quad = \mathbb{Q}_\tau^T\left[\mathbb{Q}_\tau^T\left(\check{L}_0 \in B_\gamma^0 \cap (0, T-\tau), \check{\Upsilon}^*(S_1) \in E_1, \check{\Upsilon}^*(S_2) \in E_2 \,\middle|\, \mathcal{F}_{\{1,2\}}\right)\right] \\
&\quad = \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0 \in B_\gamma^0 \cap (0,T-\tau))}\mathbb{Q}_\tau^T\left(\check{\Upsilon}^*(S_1) \in E_1, \check{\Upsilon}^*(S_2) \in E_2 \,\middle|\, \mathcal{F}_{\{1,2\}}\right)\right] \\
&\quad = \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0 \in B_\gamma^0 \cap (0,T-\tau))}\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\right]
\end{aligned}$$

so by Equation 4.3, summing over all Neveu trees we have

$$\begin{aligned}
\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) &= \mathbb{Q}_\tau^T\left(\check{L}_0 > T-\tau\right)\mathbf{1}_{B_{\{0\}}^0}(T-\tau) \\
&\quad + \sum_{\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}} \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0 \in B_\gamma^0 \cap (0,T-\tau))}\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\right]
\end{aligned}$$

By Lemma 4.2.2 we have characterized the measure $\check{\Upsilon}^*(\mathbb{Q}_\tau^T)$ on a $\cap$-stable generator of $\mathcal{E}$, so by uniqueness of probability measures on their generators [47, Lemma 1.42] the distribution is fully characterized by this recursion. □

[47]: Klenke (2020), *Probability Theory*

Example 4.3.1 Lemma 4.3.1 gives a very general recursion for a given $E \in \mathcal{G}_\mathcal{E}$, but choosing a fixed Neveu tree $\gamma \in \mathcal{N}_0$ and considering the set $E = \{\gamma\} \times B_\gamma$ for some Borel rectangle $B_\gamma \in \mathcal{B}_\gamma$, we get a somewhat simpler expression

$$\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) = \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0 \in B_\gamma^0 \cap (0,T-\tau))}\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\right]$$

for $\gamma \neq \{0\}$, and

$$\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) = \mathbb{Q}_\tau^T\left(\check{L}_0 > T-\tau\right)\mathbf{1}_{B_{\{0\}}^0}(T-\tau)$$

if $\gamma = \{0\}$.

With this recursive characterization, we are set up to prove that the distribution $\check{\Upsilon}^*(\mathbb{Q}_\tau^T)$ has a μ-density, given through quantities as described in Chapter 3. The density function evaluated in a given marked tree is quite intuitive, as it is simply a product of the conditional densities of each branch length of the given tree shape, given all the earlier branch lengths under $\prec$. The proof relies on a repeated application of Lemma 4.3.1, in particular, the form given in Example 4.3.1, together with a number of integral transformations to see that this is, in fact, a μ-density.

Theorem 4.3.2 *Let $\tau \in \mathbb{R}$ and $T \geq \tau$. The distribution $\check{\Upsilon}^*(\mathbb{Q}^T)$ has a μ-density $\pi_\tau^T : (\mathcal{N}_0^*, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}_+)$ given by*

$$\pi_\tau^T(\gamma, \ell_\gamma) = \prod_{x \in \check{\gamma}} f_{\tau+\sum_{y \prec x} \ell_y}^T\left(\tau + \sum_{y \preceq x} \ell_y\right) \prod_{x \in \check{\gamma}} \mathbb{Q}_{\tau+\sum_{y \prec x} \ell_y}^T\left(\check{L}_0 > \ell_x\right)$$

for $\gamma \neq \{0\}$ and $\ell_\gamma \in \{\{\ell_y\}_{y\in\gamma} : \sum_{z<x} \ell_z \text{ for all } x \in \check{\gamma}\} \subset \mathbb{R}^\gamma$ and by

$$\pi_\tau^T(\{0\}, T-\tau) = \mathbb{Q}_{t_0}^T(\check{L}_0 > T-\tau)$$

for $\gamma = \{0\}$, while it vanishes for all other $(\gamma, \ell_\gamma) \in \mathcal{N}_0^*$

Proof. To show that this is in fact a μ-density for the distribution we have to show that for all $E \in \mathcal{C}$ the following holds

$$\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) = \int_E \pi_\tau^T \,\mathrm{d}\mu = \sum_{\gamma\in\mathcal{N}_0} \int_{B_\gamma} \pi_\tau^T(\gamma, \ell_\gamma)\,\mathrm{d}\kappa_\gamma(\ell_\gamma)$$

As so often is suffices to check this on a $\cap$-stable generator of $\mathcal{C}$, so let $E \in \mathcal{G}_\mathcal{C}$, i.e. $E = \bigcup_{\gamma\in\mathcal{N}_0}\{\gamma\}\times B_\gamma$ where $B_\gamma = \bigtimes_{x\in\gamma} B_\gamma^x$ is a Borel rectangle and let $E_\gamma^x = \{\mathcal{S}_x(\gamma)\} \times \bigtimes_{y\in\mathcal{S}_x(\gamma)} B_\gamma^y$ for each $\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}$ and $x \in \gamma$.

From the recursion in Lemma 4.3.1 we have

$$\begin{aligned}\mathbb{Q}_\tau^T\left(\check{\Upsilon}^* \in E\right) &= \mathbb{Q}_\tau^T\left(\check{L}_0 > T-\tau\right)\mathbf{1}_{B_{\{0\}}^0}(T-\tau)\\ &\quad + \sum_{\gamma\in\mathcal{N}_0\setminus\{\{0\}\}} \mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0\in B_\gamma^0\cap(0,T-\tau))}\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\right]\end{aligned}$$

Let us, for now, consider each term in the sum above individually, i.e. take a $\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}$. We can write this expectation as an integral and see that the integrand is $\check{L}_0$-measurable, and thus the abstract change-of-variables [47, Theorem 4.10] can be used,

[47]: Klenke (2020), *Probability Theory*

$$\begin{aligned}&\mathbb{Q}_\tau^T\left[\mathbf{1}_{(\check{L}_0\in B_\gamma^0\cap(0,T-\tau))}\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\right]\\ &= \int_{(\check{L}_0\in B_\gamma^0\cap(0,T-\tau))} \mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\check{L}_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\mathrm{d}\mathbb{Q}_\tau^T\\ &= \int_{B_\gamma^0\cap(0,T-\tau)} \mathbb{Q}_{\tau+\ell_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\ell_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\mathrm{d}\check{L}_0(\mathbb{Q}_\tau^T)(\ell_0) = *\end{aligned}$$

The distribution $\check{L}_0(\mathbb{Q}_\tau^T)$, like the life space distribution $\check{L}(\Pi_\tau^T)$, has m-density $f_\tau^T(\tau+\cdot)$ as

$$\begin{aligned}\check{L}_0(\mathbb{Q}_\tau^T)(B) &= \mathbb{Q}_\tau^T\left((\check{L}\circ U_0)^{-1}(B)\right)\\ &= \mathbb{Q}_\tau^T\left(U_0^{-1}(\check{L}^{-1}(B))\right)\\ &= \int_{\check{L}^{-1}(B)} \mathrm{d}\Pi_\tau^T(\omega)\\ &= \Pi_\tau^T(\check{L}^{-1}(B))\\ &= \check{L}(\Pi_\tau^T)(B)\end{aligned}$$

where we have used Proposition 2.3.1 keeping in mind that 0 has no ancestors in $\check{I}$. Integrating with respect to this distribution with m-density then certainly gives us a Lebesgue integral,

$$* = \int_{B_\gamma^0\cap(0,T-\tau)} f_\tau^T(\tau+\ell_0)\,\mathbb{Q}_{\tau+\ell_0}^T\left(\check{\Upsilon}^* \in E_\gamma^1\right)\mathbb{Q}_{\tau+\ell_0}^T\left(\check{\Upsilon}^* \in E_\gamma^2\right)\mathrm{d}m(\ell_0) = *$$

Now if $\mathcal{S}_x(\gamma) = \{0\}$ for either $x \in \{1,2\}$, i.e. one of the remaining tree probabilities inside the integral, x must be a leaf, i.e. $x \in \check{\gamma}$, and by Example 4.3.1 we must have that

$$\mathbb{Q}^T_{\tau+\ell_0}\left(\check{\Upsilon}^* \in \{\{0\}\} \times B^x_\gamma\right) = \mathbb{Q}^T_{\tau+\ell_0}\left(\check{L}_0 > T - (\tau+\ell_0)\right) \mathbf{1}_{B^x_\gamma}(T-(\tau+\ell_0))$$

On the other hand, if $\mathcal{S}_x(\gamma) \neq \{0\}$, we can apply the simple recursion of Example 4.3.1 and our integral transformations from above to get

$$\begin{aligned}
&\mathbb{Q}^T_{\tau+\ell_0}\left(\check{\Upsilon}^* \in E^x_\gamma\right) \\
&= \int_{B^x_\gamma \cap (0, T-(\tau+\ell_0))} f^T_{\tau+\ell_0}(\tau+\ell_0+\ell_1)\mathbb{Q}^T_{\tau+\ell_0+\ell_1}\left(\check{\Upsilon}^* \in E^{x1}_\gamma\right)\mathbb{Q}^T_{\tau+\ell_0+\ell_1}\left(\check{\Upsilon}^* \in E^{x2}_\gamma\right) \mathrm{d}m(\ell_1).
\end{aligned}$$

Applying this argument recursively down each line of descend from 0 to each $x \in \gamma$, we get an analog integral for each internal branch and a probability for each leaf, nested according to the partial order $\prec$ on $\check{I}$. Then moving all the integrands into the innermost integral, such that all the integrals are nested according to the linear order $<$, we have from Proposition 4.2.3 that this is an integral with respect to κ_γ, and we particularly see that our proposed density function π^T_τ appears,

$$\begin{aligned}
* &= \int_{B_\gamma} \prod_{x \in \check{\gamma}} f^T_{\tau+\sum_{y<x}\ell_y}\left(\tau + \sum_{y \preceq x} \ell_y\right) \prod_{x \in \check{\gamma}} \mathbb{Q}^T_{\tau+\sum_{y<x}\ell_y}\left(\check{L}_0 > \ell_x\right) \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&= \int_{B_\gamma} \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma)
\end{aligned}$$

Plugging this into the general recursion we get

$$\begin{aligned}
&\mathbb{Q}^T_\tau\left(\check{\Upsilon}^* \in E\right) \\
&\quad = \mathbb{Q}^T_\tau\left(\check{L}_0 > T - \tau\right) \mathbf{1}_{B^0_{\{0\}}}(T-\tau) + \sum_{\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}} \int_{B_\gamma} \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&\quad = \int_{B_{\{0\}}} \pi^T_\tau(\{0\}, \ell_0)\, \mathrm{d}\kappa_{\{0\}}(\ell_0) + \sum_{\gamma \in \mathcal{N}_0 \setminus \{\{0\}\}} \int_{B_\gamma} \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&\quad = \sum_{\gamma \in \mathcal{N}_0} \int_{B_\gamma} \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\kappa_\gamma(\ell_\gamma) \\
&\quad = \sum_{\gamma \in \mathcal{N}_0} \int_E \pi^T_\tau(\gamma', \ell_{\gamma'})\, \mathrm{d}\mu_\gamma(\gamma', \ell_{\gamma'}) \\
&\quad = \int_E \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\mu
\end{aligned}$$

again using Proposition 4.2.3.

Now as we have seen that $\check{\Upsilon}^*(\mathbb{Q}^T_\tau)$ and $\pi^T_\tau \cdot \mu$ agrees on the $\cap$-stable generator $\mathcal{G}_{\mathcal{E}}$, we must by uniqueness of measures on a $\cap$-stable generator have for any $E \in \mathcal{E}$ that

$$\mathbb{Q}^T_\tau\left(\check{\Upsilon}^* \in E\right) = \int_E \pi^T_\tau(\gamma, \ell_\gamma)\, \mathrm{d}\mu$$

i.e. $\check{\Upsilon}^*(\mathbb{Q}^T_\tau)$ has μ-density π^T_τ. □

[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'
[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

Example 4.3.2 This result is very close to those of [22, 41]. To simplify to that of [22] we choose the offspring process on the original population spaces to be almost surely equal to 2, i.e. we have no extinctions, and we choose an arbitrary time-constant lifetime distribution with m-density. In this case, the density found in Theorem 3.5.2 simplifies to just being the lifetime distribution itself, and almost all of the extra constructions in the current chapter become superfluous.

To simplify to the one-type situation of [41], we choose the offspring distribution as time-constant, but allowing for both extinctions and speciations. We also here choose an arbitrary time-constant lifetime distribution with m-density. Our integral equation for h_τ^T is slightly different from that which [41] calls b. This is likely due to differences in what we condition on, as it is not entirely clear whether [41] conditions on non-extinction of the population, but it is mentioned briefly in Section 2 (p. 39) that b is not always a density, which we speculate accounts for the situation where the population has gone extinct before observation. As the arguments for the integral equations in [41] are completely based on heuristic argument, there might also be some terms that were not considered as relevant, but that we have shown them to be.

Both of these results have an obvious difference from ours in that the densities are multiplied by a factor depending on the Neveu tree. This factor arises as we work directly with Neveu trees, being oriented, while the two papers ignore the labeling of the tree, and thus we should apply the counting result for the number of oriented trees having a particular tree shape from Subsection 4.1.2.

Conclusion and outlook 5

We have in this thesis explored the theory of branching processes and branching trees and we have developed a novel framework for using branching processes and trees for applications in phylogenetics. The application part has been deemphasized quite a lot, besides inspiring us into which directions to pursue. But the technical contents of this thesis are certainly only justifiable if they can provide some insight into the (at least theoretical[1]) sciences, they are not a product of *math for math's sake*. So, to qualify our efforts in rigor and formalization, we will here summarize our results, see how they might directly provide some insights, where there is still some work to do, where they by themselves might be insufficient, and which directions to look further into.

5.1 Conclusion 73
5.2 Estimating evolutionary dynamics 74
5.3 Implementation 76
5.4 Further work and limitations 77

1: By this I mean the systematic study of a field of science, for example, biology, using e.g. probabilistic models. What is often called *applied probability/mathematics*, which, as argued in [38], is not a very well chosen name.

5.1 Conclusion

Stochastic models for describing phylogenetic trees have been around for exactly one hundred years. Since then a flurry of scientific insights and advances has been based on such models, but the models themselves have not evolved remarkably since Yule's original process and the later birth-death processes. In particular, do a large class of such models, which we would otherwise have hoped for to capture some defining features of the evolutionary process, provide an unsatisfactory fit to empirical reduced trees [11, 49]. Considering models outside of this class could thus prove important to gain a better understanding of the evolutionary dynamics of the past.

[11]: Blum et al. (2006), 'Which Random Processes Describe the Tree of Life?'
[49]: Lambert et al. (2013), 'Birth–death models and coalescent point processes'

One such deviation is to let the distribution of speciation times be dependent of the age of the species. As this breaks the species-speciation exchangeability of the previously mentioned class of models, such a model would most likely not produce the same distribution over tree shapes as this class does. Simulation studies, and simple analytic cases [22, 26, 41, 42], have proved that this is in fact the case for the classical Bellman-Harris process.

[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'
[26]: Hagen et al. (2015), 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'
[41]: Jones (2011), 'Calculations for multitype age-dependent binary branching processes'
[42]: Jones (2011), 'Tree Models for Macroevolution and Phylogenetic Analysis'

Besides in these small initial studies have the class of non-Markovian branching process, including the Bellman-Harris process, not been examined as a model for phylogenetic trees. In particular, has the literature seen no thorough mathematical extension of the processes to accommodate the special features of phylogenetic data such as seen for the birth-death processes [48, 59]. We provide a starting point for such further endeavors.

[48]: Lambert (2010), 'The contour of splitting trees is a Lévy process'
[59]: Nee et al. (1994), 'The Reconstructed Evolutionary Process'

Initially, we built a strong mathematical foundation directly from the concept of a branching population, on which we have seen the Markov-like branching property holds. The branching property is not a new concept, as first proposed in [12, 60] and with a form close to the current proved in [37]. Meanwhile, the author knows of no easily accessible proofs in the literature, hence a modern and thoroughly explained version has been presented. With the branching property at hand, several results, which

[12]: Chauvin (1986), 'Arbres et processus de Bellman-Harris'
[60]: Neveu (1986), 'Arbres et processus de Galton-Watson'
[37]: Jagers (1989), 'General branching processes as Markov fields'

would otherwise be stated as obvious or just not given an argument, can be proved stringently.

From the foundation, the reduced time-varying Bellman-Harris branching process was defined under the unifying framework of random characteristics, and using the branching property, a complete characterization of the distribution of this process was found. Defining the process through random characteristics meant that it was easy to see the deep connection between the classical and the reduced branching process. Specializing to the binary offspring case, we showed integral equations for a number of interesting quantities akin to [41]. In particular, we found the density of the first splitting time of the reduced branching process, conditioned on non-extinction before observation, a very important variable for our further work, but also in and of itself as it is closely related to the *time since least common ancestor*.

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

Lastly, we proposed a framework for modeling reduced branching trees rigorously, in particular, we defined a natural measure space on which such trees can take values and a measure on this space which was later shown to be the dominating measure for the density of the reduced branching tree. Using this framework we defined the reduced branching tree and derived a recursive characterization of the trees distribution, which was finally used to obtain the density function of a reduced branching tree, thus formalizing, and extending, the statements in [41].

5.2 Estimating evolutionary dynamics

From the very first section, has phylogenetics been our main source of inspiration in developing probabilistic methods for branching populations and trees. In particular, the probabilistic models for the evolution of reduced phylogenetic trees have been the central driver of our progress. Now, our methods should then also be applicable in phylogenetics, so let us look at a few possible applications, and how our methods might provide new insights.

Our reduced tree density can serve as a likelihood function of the evolutionary dynamics for a given observation of a reduced phylogenetic tree. In particular, if we have parametrized the offspring process and the lifetime distribution, for example by assuming it to be a gamma or Weibull distribution,[2] we can, at least theoretically, estimate these parameters by maximizing the likelihood function. As our density, and thus likelihood, function is the most general and practically useful form of the likelihood of a non-Markovian model of evolution, it might provide more accurate and robust estimates than traditional methods, this is certainly only a hypothesis, and further studies are very much needed.

2: Many different distributions can be used naturally as a time-to-death distribution, some might have natural biological interpretations, while others are used due to mathematical tractability.

However, whether such parameters are even identifiable is an open question. As noted, a recent study showed that general time-varying rate parameters in a birth-death process are not identifiable [53], while it becomes identifiable when restricting the parameter space to either piecewise constant functions [50] or piecewise polynomial functions [51]. General estimation theory for continuous-time non-Markovian branching processes has seen little developments since [7], and to the best knowledge of the author, no estimation theory exists for when the full branching

[53]: Louca et al. (2020), 'Extant timetrees are consistent with a myriad of diversification histories'

[50]: Legried et al. (2022), 'A class of identifiable phylogenetic birth–death models'

[51]: Legried et al. (2023), 'Identifiability and inference of phylogenetic birth–death models'

[7]: Athreya et al. (1977), 'Estimation Theory for Continuous-Time Branching Processes'

tree is observed, not only the counting process, so this might prove an interesting further field of study.

When inferring evolutionary dynamics from stochastic processes of phylogenetics trees we have to be careful with what we condition the likelihood function on. [80] has made a thorough review of seven different likelihood functions for the rate parameters in a simple birth-death model, each of which is conditioned on slightly different events, and has done a simulation study comparing the accuracy of the estimators from each likelihood. Our likelihood function as presented here is conditioned on the origin time of the process as well as non-extinction of the population before observation, but conditioning on further events, such as the process having a certain number of extant individuals is a trivial extension.

[80]: Stadler (2013), 'How Can We Improve Accuracy of Macroevolutionary Rate Estimates?'

Too much conditioning is bad as each event we condition on will "remove some information" present in the data, making estimators less accurate. For example, when conditioning on the number of extant individuals, we do not anymore have that as a "data point" which can inform us about the dynamics of the process. In general, [80] gives a series of recommendations on what to condition for accurate phylogenetic inference with birth-death models. It is argued to always condition on non-extinction of the population, as we would in practice never observe an extinct tree, and extinction rates are thus underestimated. Conditioning on both the origin time of the process and the number of species is not recommended as this increases the variance of the estimators too much.

Knowing the origin time of a phylogenetic tree is not very common, and most phylogenetic analyses only have data on the tree from the least common ancestors splitting time to observation. This might seem damning for our approach as we directly condition on the origin time, but we can deal with this situation in two ways. First, we can simply remove the least common ancestor vertex in the tree and consider the resulting subtrees as independent branching populations due to the independence structure given in the branching property. In this way, we only need to know the splitting time of the least common ancestor, which is then the origin time of each process. Another approach that is often taken, is to condition on the number of extant species and assume a prior distribution over the origin time (this is what we discussed just after Proposition 2.3.1) over which we integrate. As we seldom know anything about the time of origin, is is often done using the (improper) uniform prior over $(-\infty, T)$. For more details on this approach see [6, 22].

[6]: Andersen (2023), 'Models for the evolution of a phylogenetic tree'
[22]: Gernhard et al. (2008), 'Stochastic properties of generalised Yule models, with biodiversity applications'

Finally, it is worth noting that while inferring evolutionary dynamics from observed phylogenetic trees is in itself a difficult problem, the validity of the methods will always heavily rely on the accuracy of the phylogenetic trees used as data, and reconstructing such trees from genetic data on extant individuals is a very complex and demanding task, which needs its methodological developments to become more robust [43]. Our distribution over reduced phylogenetic trees might even prove an interesting tool in this endeavor, as a more sophisticated prior tree distribution in a Bayesian phylogenetic reconstruction framework, which today usually uses Yule or birth-death process priors [18].

[43]: Kapli et al. (2020), 'Phylogenetic tree building in the genomic age'

[18]: Drummond et al. (2007), 'BEAST'

5.3 Implementation

Our general methods contain essentially no closed-form solution for anything, and solving the integral equations analytically is not very feasible for anything but already known cases such as the Yule process and the birth-death process. Thus applying our methods will require computational tools, and implementing these is not completely trivial, in particular doing so efficiently. Making readily available and efficient implementations is of utmost importance if these methods should ever be used as an alternative to the already well-implemented and used models. The methods developed in [41] for example have seen no published applications, which might simply be due to its intricate mathematical formulation which is not trivial to implement.

[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'

For a similar renewal-like integral equation, for the mean of a Bellman-Harris process, in the context of infectious disease epidemiology, has [61] proposed a simple Riemann-sum discretization which can be elegantly implemented in a vectorized approach with an interpreted programming language such as `python` or `R`. Using a vectorized approach means getting an efficiency close to that of a compiled language while keeping the code readable and easily modified. In [5] an extension of this approach to an integral equation around the same level of complexity as ours is proposed and implemented, hence a similar approach seems feasible for our equations. A vectorized approach was in these cases crucial for the applicability of the methods, as a person without experience with compiled languages should be able to extend the method to their particular situation. Whether such extensibility of the implementation is as important for phylogenetic application, or the efficiency of a possible implementation in a compiled language is of higher priority is not certain.

[61]: Pakkanen et al. (2023), 'Unifying incidence and prevalence under a time-varying general branching process'

[5]: Andersen (2023), 'Efficient computation of the probability generating function in a general branching process model for infectious diseases'

Another important computational tool for processes like ours is a simulation algorithm. Due to the non-Markovian nature of the process, no efficient Doob-Gillespie-type algorithms [23] exist. Instead, one must turn to direct simulation of each individual's lifetime and offspring process, after which to simulate each of its offspring in the same way. If we simply want a simulation of a population originating at time $\tau \in \mathbb{R}$ and observed at $T \geq \tau$, this is relatively straightforward, while tedious and inefficient. On the other hand, if we wanted to stop the simulation when it has a fixed number of extant individuals, trouble arises for processes with extinction as there might be several natural time points at which to stop the process. A general sampling approach for phylogenetic trees dealing with this exact problem, which will also fit our methods, is proposed in [29]. Such simulation methods, both for stopping at a given time and with a given number of extant species are implemented for the simple non-time varying (reduced) Bellman-Harris process in [25], and this package is easily extendable to our framework. Further work might lie in more efficient implementations in a compiled language, which is certainly necessary if we need to simulate a large number of trees, for example in a Bayesian analysis using Markov chain Monte Carlo.

[23]: Gillespie (1976), 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions'

[29]: Hartmann et al. (2010), 'Sampling Trees from Evolutionary Models'

[25]: Hagen et al. (2018), 'TreeSimGM: Simulating phylogenetic trees under general Bellman–Harris models with lineage-specific shifts of speciation and extinction in R'

5.4 Further work and limitations

As our main motivation for working with age-dependent models for phylogenetic trees was the observed deviation between expected and observed tree shape for a large class of models, a characterization of the tree shape distribution induced by our reduced time-varying Bellman-Harris process should be investigated. Our model is probably too broad and intractable for a full and clear description of the distribution, but further analytic results as those hidden in the supplementary materials for [26] might be possible for other lifetime distributions and possibly including extinctions.

[26]: Hagen et al. (2015), 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'

Another point from [26], is that our model for age-dependent speciation is not necessarily close to actual speciations in nature. In particular is it very unlikely that speciations occur instantaneously at the extinction of another species, we would more likely see that an offspring species originates sometime during the life of the maternal species and that this species might go on to exist alongside its descendant. [26] suggests an asymmetric speciation mode, where one offspring of an individual inherits the individual's age, and the other is endowed with age 0. This does however not fit very well into the Bellman-Harris framework,[3] but would be more appropriately modeled using a simple Crump-Mode-Jagers process. Whether to pursue such a direction more is uncertain as the simulation study in [26] shows that using symmetric contra asymmetric speciation has little effect on either tree shape or branch lengths.

3: Although Harris [28, Chapter VI, 29.] actually discusses this model superficially.

A few natural extensions of our process inspired by [33, 41, 76] is to include multiple types of individuals and consider incomplete sampling of extant individuals. The first extension is to make our branching process a *multi-type branching process* [54], this will certainly complicate many of our equations, but they should still be similar to the multi-type approach in [41]. Extending our foundational approach to take multiple types into account amounts only to indexing our basic Markov kernel by both the birth time of an individual and its type, if the type of an individual is determined from its ancestor alone, the branching property and hence all of our results should follow. A multi-type approach might be useful in a phylogenetic application as different species can then have different heritable traits determining their life law. Another extension, that is often considered in the modeling of phylogenetic trees, is the incomplete sampling of extant individuals. In [41, 76] a simple random sampling scheme is considered and in [33] a larger array of different and more intricate sampling schemes are studied. In particular, has it been shown that sampling of extant individuals can have a large effect on the accuracy of the estimators of the evolutionary dynamics. Our model should thus be extended to allow for, at least, a simple random sampling of extant individuals. We conjecture that an adjustment of the random characteristic defining the reduced process could do the work, but haven't been successful in working out the details, from this, we should then again obtain similar but extended versions of the results in [41].

[33]: Hohna et al. (2011), 'Inferring Speciation and Extinction Rates under Different Sampling Schemes'
[41]: Jones (2011), 'Calculations for multi-type age-dependent binary branching processes'
[76]: Stadler (2009), 'On incomplete sampling under birth–death models and connections to the sampling-based coalescent'
[54]: Mode (1971), *Multitype Branching Processes*

Finally, our theories are only as good as the impact they can have on actual scientific advancement, and for such to happen we need actual (theoretical) studies using our methods, and putting them into a

sound biological theory of evolution. As for now, we have only taken age-dependency into account as it could solve a statistical problem, without any biological motivation as such for this. There are a few biological interpretations of particular choices of non-exponential lifetime distributions as given in [26, 42, 83] for distributions with a very high speciation rate at a low age, while the opposite case, where the speciation rates are lower for very young species, is argued for in [52]. Thus, it is evident that we still lack the unifying theory of stochastic evolution of phylogenetic trees.

[26]: Hagen et al. (2015), 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'

[42]: Jones (2011), 'Tree Models for Macroevolution and Phylogenetic Analysis'

[83]: Steel et al. (2001), 'Properties of phylogenetic trees generated by Yule-type speciation models'

[52]: Losos et al. (1995), 'Stumped by Trees?'

Bibliography

Here are the references in alphabetical order.

[1] Romain Abraham and Jean-François Delmas. *An introduction to Galton-Watson trees and their local limits*. Sept. 18, 2020. URL: http://arxiv.org/abs/1506.05571 (visited on 11/27/2023) (cited on page 53).

[2] David Aldous. 'Probability Distributions on Cladograms'. In: *Random Discrete Structures*. Vol. 76. Series Title: The IMA Volumes in Mathematics and its Applications. New York, NY: Springer New York, 1996, pp. 1–18. DOI: 10.1007/978-1-4612-0719-1_1. (Visited on 08/13/2023) (cited on page 5).

[3] David Aldous and Lea Popovic. 'A Critical Branching Process Model for Biodiversity'. In: *Advances in Applied Probability* 37 (2005), pp. 1094–1115 (cited on page 4).

[4] Michael E. Alfaro et al. 'Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates'. In: *Proceedings of the National Academy of Sciences* 106.32 (Aug. 11, 2009). Publisher: Proceedings of the National Academy of Sciences, pp. 13410–13414. DOI: 10.1073/pnas.0811087106. (Visited on 12/18/2023) (cited on page 4).

[5] Frederik Mølkjær Andersen. 'Efficient computation of the probability generating function in a general branching process model for infectious diseases'. Project outside of coursescope. University of Copenhagen, May 26, 2023 (cited on page 76).

[6] Frederik Mølkjær Andersen. 'Models for the evolution of a phylogenetic tree'. Project outside of coursescop. University of Copenhagen, Aug. 2023 (cited on pages 53, 57, 75).

[7] K. B. Athreya and Niels Keiding. 'Estimation Theory for Continuous-Time Branching Processes'. In: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 39.2 (1977). Publisher: Springer, pp. 101–123. (Visited on 12/19/2023) (cited on page 74).

[8] Krishna B. Athreya and Peter E. Ney. *Branching Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1972. (Visited on 05/20/2023) (cited on page 40).

[9] Richard Bellman and Theodore Harris. 'On Age-Dependent Binary Branching Processes'. In: *The Annals of Mathematics* 55.2 (Mar. 1952), p. 280. DOI: 10.2307/1969779. (Visited on 08/03/2023) (cited on page 3).

[10] Irénée-Jules Bienaymé. 'De la loi de multiplication et de la duree des familles'. In: *Soc. Philomath. Paris Extraits*. 5th ser. (1845), pp. 37–39 (cited on page 3).

[11] Michael G. B. Blum and Olivier François. 'Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance'. In: *Systematic Biology* 55.4 (Aug. 1, 2006). Ed. by Mike Steel, pp. 685–691. DOI: 10.1080/10635150600889625. (Visited on 08/02/2023) (cited on pages 5, 73).

[12] B Chauvin. 'Arbres et processus de Bellman-Harris'. In: *Annales de l'IHP Probabilités et statistiques* 22.2 (1986), pp. 209–232 (cited on pages 17, 28, 73).

[13] Brigitte Chauvin. 'Product Martingales and Stopping Lines for Branching Brownian Motion'. In: *The Annals of Probability* 19.3 (1991). Publisher: Institute of Mathematical Statistics, pp. 1195–1205. (Visited on 12/12/2023) (cited on page 28).

[14] Kenny Crump and Charles J Mode. 'A general age-dependent branching process. II'. In: *Journal of Mathematical Analysis and Applications* .25.1 (Jan. 1969), pp. 8–17. DOI: 10.1016/0022-247X(69)90210-8. (Visited on 03/19/2023) (cited on page 3).

[15] Kenny S Crump and Charles J Mode. 'A general age-dependent branching process. I'. In: *Journal of Mathematical Analysis and Applications* .24.3 (Dec. 1968), pp. 494–508. DOI: 10.1016/0022-247X(68)90005-X. (Visited on 03/19/2023) (cited on page 3).

[16] Charles Darwin. *Darwin's Notebook B: Transmutation of species*. 1837 (cited on page iii).

[17] Charles Darwin. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859 (cited on page 1).

[18] Alexei J. Drummond and Andrew Rambaut. 'BEAST: Bayesian evolutionary analysis by sampling trees'. In: *BMC Evolutionary Biology* 7.1 (Nov. 8, 2007), p. 214. DOI: 10.1186/1471-2148-7-214. (Visited on 12/18/2023) (cited on pages 5, 75).

[19] Willy Feller. 'Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung'. In: *Acta Biotheoretica* 5.1 (May 1, 1939), pp. 11–40. DOI: 10.1007/BF01602932. (Visited on 12/16/2023) (cited on page 3).

[20] Mareike Fischer et al. *Tree Balance Indices: A Comprehensive Survey*. 1st ed. 2023 edition. Springer, Nov. 1, 2023. 419 pp. (cited on page 5).

[21] Tanja Gernhard. 'The conditioned reconstructed process'. In: *Journal of Theoretical Biology* 253.4 (Aug. 2008), pp. 769–778. DOI: 10.1016/j.jtbi.2008.04.005. (Visited on 07/19/2023) (cited on pages 4, 35).

[22] Tanja Gernhard, Klaas Hartmann, and Mike Steel. 'Stochastic properties of generalised Yule models, with biodiversity applications'. In: *Journal of Mathematical Biology* 57.5 (Nov. 2008), pp. 713–735. DOI: 10.1007/s00285-008-0186-y. (Visited on 07/04/2023) (cited on pages 4, 6–8, 53, 72, 73, 75).

[23] Daniel T Gillespie. 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions'. In: *Journal of Computational Physics* 22.4 (Dec. 1, 1976), pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3. (Visited on 12/19/2023) (cited on page 76).

[24] Peter Guttorp. *Three papers on the history of branching processes*. Technical report 242. Department of Statistics, GN-22: University of Washington, Nov. 1992 (cited on page 3).

[25] Oskar Hagen and Tanja Stadler. 'TreeSimGM: Simulating phylogenetic trees under general Bellman–Harris models with lineage-specific shifts of speciation and extinction in R'. In: *Methods in Ecology and Evolution* 9.3 (Mar. 2018). Ed. by Samantha Price, pp. 754–760. DOI: 10.1111/2041-210X.12917. (Visited on 07/05/2023) (cited on page 76).

[26] Oskar Hagen et al. 'Age-Dependent Speciation Can Explain the Shape of Empirical Phylogenies'. In: *Systematic Biology* 64.3 (May 1, 2015), pp. 432–440. DOI: 10.1093/sysbio/syv001. (Visited on 05/20/2023) (cited on pages 4, 6–8, 73, 77, 78).

[27] E F Harding. 'The Probabilities of Rooted Tree-Shapes Generated by Random Bifurcation'. In: *Advances in Applied Probability* 3.1 (1971), pp. 44–77 (cited on page 5).

[28] Theodore Edward Harris. *The theory of branching processes*. OCLC: 931719544. Berlin: Springer-Verlag, 1963 (cited on pages 3, 17, 30, 77).

[29] Klaas Hartmann, Dennis Wong, and Tanja Stadler. 'Sampling Trees from Evolutionary Models'. In: *Systematic Biology* 59.4 (July 1, 2010), pp. 465–476. DOI: 10.1093/sysbio/syq026. (Visited on 07/11/2023) (cited on page 76).

[30] Paul H. Harvey, Robert M. May, and Sean Nee. 'Phylogenies Without Fossils'. In: *Evolution* 48.3 (1994). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.1994.tb01341.x, pp. 523–529. DOI: 10.1111/j.1558-5646.1994.tb01341.x. (Visited on 12/18/2023) (cited on page 4).

[31] S. Blair Hedges and Sudhir Kumar, eds. *The timetree of life*. Oxford biology. OCLC: ocn268792961. Oxford ; New York: Oxford University Press, 2009. 551 pp. (cited on page 54).

[32] Jody Hey. 'Using Phylogenetic Trees to study Speciation and Extinction'. In: *Evolution* 46.3 (June 1992), pp. 627–640. DOI: 10.1111/j.1558-5646.1992.tb02071.x. (Visited on 08/19/2023) (cited on page 4).

[33] S. Hohna et al. 'Inferring Speciation and Extinction Rates under Different Sampling Schemes'. In: *Molecular Biology and Evolution* 28.9 (Sept. 1, 2011), pp. 2577–2589. DOI: 10.1093/molbev/msr095. (Visited on 09/04/2023) (cited on pages 4, 77).

[34] Jean Jacod and Philip Protter. *Probability Essentials*. Universitext. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. (Visited on 12/19/2023) (cited on page 24).

[35] Peter Jagers. 'A general stochastic model for population development'. In: *Scandinavian Actuarial Journal* .1969.1 (Jan. 1969), pp. 84–103. DOI: 10.1080/03461238.1969.10405220. (Visited on 03/19/2023) (cited on page 3).

[36] Peter Jagers. *Branching processes with biological applications*. London ; New York: Wiley, 1975. 268 pp. (cited on pages 11, 29, 30).

[37] Peter Jagers. 'General branching processes as Markov fields'. In: *Stochastic Processes and their Applications* 32.2 (Aug. 1989), pp. 183–212. DOI: 10.1016/0304-4149(89)90075-6. (Visited on 07/19/2023) (cited on pages 8, 12, 28, 31, 73).

[38] Peter Jagers. 'Applied probability as theoretical science: 50 years in the applied probability community'. In: *Advances in Applied Probability* 50 (Dec. 1, 2018), pp. 1–2. DOI: 10.1017/apr.2018.64 (cited on page 73).

[39] Peter Jagers and Olle Nerman. 'The Growth and Composition of Branching Populations'. In: *Advances in Applied Probability* 16.2 (1984), pp. 221–259 (cited on pages 28, 29).

[40] Peter Jagers and Olle Nerman. 'Branching Processes in Periodically Varying Environment'. In: *The Annals of Probability* 13.1 (Feb. 1, 1985). DOI: 10.1214/aop/1176993079. (Visited on 10/06/2023) (cited on page 8).

[41] Graham Jones. 'Calculations for multi-type age-dependent binary branching processes'. In: *Journal of Mathematical Biology* 63.1 (July 2011), pp. 33–56. DOI: 10.1007/s00285-010-0362-8. (Visited on 05/20/2023) (cited on pages 4, 6–8, 40, 44, 45, 53, 72–74, 76, 77).

[42] Graham R. Jones. 'Tree Models for Macroevolution and Phylogenetic Analysis'. In: *Systematic Biology* 60.6 (Dec. 1, 2011), pp. 735–746. DOI: 10.1093/sysbio/syr086. (Visited on 07/10/2023) (cited on pages 6, 7, 73, 78).

[43] Paschalia Kapli, Ziheng Yang, and Maximilian J. Telford. 'Phylogenetic tree building in the genomic age'. In: *Nature Reviews Genetics* 21.7 (July 2020). Number: 7 Publisher: Nature Publishing Group, pp. 428–444. DOI: 10.1038/s41576-020-0233-0. (Visited on 12/19/2023) (cited on page 75).

[44] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes*. 2d ed. New York: Academic Press, 1975. 557 pp. (cited on page 46).

[45] David G. Kendall. 'On the Generalized "Birth-and-Death" Process'. In: *The Annals of Mathematical Statistics* 19.1 (Mar. 1948), pp. 1–15. DOI: 10.1214/aoms/1177730285. (Visited on 10/01/2022) (cited on page 3).

[46] Marek Kimmel. 'The Point-Process Approach to Age- and Time-Dependent Branching Processes'. In: *Advances in Applied Probability* 15.1 (1983). Publisher: Applied Probability Trust, pp. 1–20. DOI: 10.2307/1426979. (Visited on 01/30/2023) (cited on pages 8, 40).

[47] Achim Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Cham: Springer International Publishing, 2020. (Visited on 02/03/2023) (cited on pages 19–21, 23, 24, 27, 34, 37, 39, 48–50, 60, 63, 65, 69, 70).

[48] Amaury Lambert. 'The contour of splitting trees is a Lévy process'. In: *The Annals of Probability* 38.1 (Jan. 1, 2010). DOI: 10.1214/09-AOP485. (Visited on 07/05/2023) (cited on pages 5, 73).

[49] Amaury Lambert and Tanja Stadler. 'Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies'. In: *Theoretical Population Biology* 90 (Dec. 2013), pp. 113–128. DOI: 10.1016/j.tpb.2013.10.002. (Visited on 07/19/2023) (cited on pages 4, 5, 73).

[50] Brandon Legried and Jonathan Terhorst. 'A class of identifiable phylogenetic birth–death models'. In: *Proceedings of the National Academy of Sciences* 119.35 (Aug. 30, 2022). Publisher: Proceedings of the National Academy of Sciences, e2119513119. DOI: 10.1073/pnas.2119513119. (Visited on 12/18/2023) (cited on pages 7, 74).

[51] Brandon Legried and Jonathan Terhorst. 'Identifiability and inference of phylogenetic birth–death models'. In: *Journal of Theoretical Biology* 568 (July 2023), p. 111520. DOI: 10.1016/j.jtbi.2023.111520. (Visited on 07/12/2023) (cited on pages 7, 74).

[52] Jonathan B. Losos and Frederick R. Adler. 'Stumped by Trees? A Generalized Null Model for Patterns of Organismal Diversity'. In: *The American Naturalist* 145.3 (Mar. 1995). Publisher: The University of Chicago Press, pp. 329–342. DOI: 10.1086/285743. (Visited on 12/19/2023) (cited on page 78).

[53] Stilianos Louca and Matthew W. Pennell. 'Extant timetrees are consistent with a myriad of diversification histories'. In: *Nature* 580.7804 (Apr. 23, 2020), pp. 502–505. DOI: 10.1038/s41586-020-2176-1. (Visited on 07/12/2023) (cited on pages 7, 74).

[54] Charles J. Mode. *Multitype Branching Processes: Theory and Applications*. Google-Books-ID: ghvvAAAAMAAJ. American Elsevier Publishing Company, 1971. 358 pp. (cited on page 77).

[55] Hélène Morlon, Todd L. Parsons, and Joshua B. Plotkin. 'Reconciling molecular phylogenies with the fossil record'. In: *Proceedings of the National Academy of Sciences* 108.39 (Sept. 27, 2011), pp. 16327–16332. DOI: 10.1073/pnas.1102543108. (Visited on 09/04/2023) (cited on page 4).

[56] Sean Nee. 'Inferring Speciation Rates from Phylogenies'. In: *Evolution* 55.4 (2001). Publisher: [Society for the Study of Evolution, Wiley], pp. 661–668. (Visited on 12/18/2023) (cited on page 4).

[57] Sean Nee. 'Birth-Death Models in Macroevolution'. In: *Annual Review of Ecology, Evolution, and Systematics* 37.1 (Dec. 1, 2006), pp. 1–17. DOI: 10.1146/annurev.ecolsys.37.091305.110035. (Visited on 07/12/2023) (cited on page 4).

[58] Sean Nee, Robert M May, and Paul H Harvey. 'Extinction rates can be estimated from molecular phylogenies'. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 344.1307 (Apr. 29, 1994), pp. 77–82. DOI: 10.1098/rstb.1994.0054. (Visited on 07/12/2023) (cited on page 4).

[59] Sean Nee, Robert M May, and Paul H Harvey. 'The Reconstructed Evolutionary Process'. In: *Philosophical Transactions: Biological Sciences* 344.1309 (1994), pp. 305–311 (cited on pages 4, 8, 35, 46, 73).

[60] Jacques Neveu. 'Arbres et processus de Galton-Watson'. In: *Annales de l'IHP Probabilités et statistiques* 22.2 (1986), pp. 199–207 (cited on pages 12, 17, 73).

[61] Mikko S. Pakkanen et al. 'Unifying incidence and prevalence under a time-varying general branching process'. In: *Journal of Mathematical Biology* 87.2 (Aug. 1, 2023), p. 35. DOI: 10.1007/s00285-023-01958-w. (Visited on 08/28/2023) (cited on pages 8, 15, 40, 76).

[62] Matthew J. Penn et al. 'Intrinsic randomness in epidemic modelling beyond statistical uncertainty'. In: *Communications Physics* 6.1 (June 20, 2023). Number: 1 Publisher: Nature Publishing Group, pp. 1–9. DOI: 10.1038/s42005-023-01265-2. (Visited on 08/28/2023) (cited on pages 8, 15).

[63] David Pollard. *A User's Guide to Measure Theoretic Probability*. 1st ed. Cambridge University Press, Dec. 10, 2001. (Visited on 01/20/2023) (cited on page 9).

[64] Lea Popovic. 'Asymptotic genealogy of a critical branching process'. In: *The Annals of Applied Probability* 14.4 (Nov. 2004). Publisher: Institute of Mathematical Statistics, pp. 2120–2148. DOI: 10.1214/105051604000000486. (Visited on 12/18/2023) (cited on page 4).

[65] Daniel L Rabosky and Irby J Lovette. 'Density-dependent diversification in North American wood warblers'. In: *Proceedings of the Royal Society B: Biological Sciences* 275.1649 (July 8, 2008). Publisher: Royal Society, pp. 2363–2371. DOI: 10.1098/rspb.2008.0630. (Visited on 12/18/2023) (cited on page 4).

[66] Daniel L. Rabosky. 'Likeligood methods for detecting Temorak Shifts in Diversification rates'. In: *Evolution* 60.6 (June 2006), pp. 1152–1164. DOI: 10.1111/j.0014-3820.2006.tb01194.x. (Visited on 08/19/2023) (cited on page 4).

[67] Mark A Ragan. 'Trees and networks before and after Darwin'. In: *Biology Direct* 4.1 (2009), p. 43. DOI: 10.1186/1745-6150-4-43. (Visited on 12/01/2023) (cited on pages 1, 2).

[68] Bruce Rannala and Ziheng Yang. 'Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference'. In: *Journal of Molecular Evolution* 43 (1996), pp. 304–311 (cited on pages 4, 7).

[69] David M. Raup et al. 'Stochastic Models of Phylogeny and the Evolution of Diversity'. In: *The Journal of Geology* 81.5 (1973). Publisher: The University of Chicago Press, pp. 525–542. (Visited on 12/16/2023) (cited on page 3).

[70] Fredrik Ronquist et al. 'MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space'. In: *Systematic Biology* 61.3 (May 2012), pp. 539–542. DOI: 10.1093/sysbio/sys029. (Visited on 12/18/2023) (cited on page 5).

[71] S. M. Sagitov. 'Reduced Critical Bellman–Harris Branching Process with Several Types of Particles'. In: *Theory of Probability & Its Applications* 30.4 (Dec. 1986), pp. 783–796. DOI: 10.1137/1130097. (Visited on 11/10/2023) (cited on pages 8, 32).

[72] B. A. Sevast'yanov. 'Age-dependent branching processes'. In: *Teor. Veroyat. Primen.* IX.4 (1964), pp. 577–594 (cited on page 3).

[73] Daniele Silvestro, Nicolas Salamin, and Jan Schnitzler. 'PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data'. In: *Methods in Ecology and Evolution* 5.10 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12263, pp. 1126–1131. DOI: 10.1111/2041-210X.12263. (Visited on 12/18/2023) (cited on page 5).

[74] T. Stadler. 'Recovering speciation and extinction dynamics based on phylogenies'. In: *Journal of Evolutionary Biology* 26.6 (June 2013), pp. 1203–1219. DOI: 10.1111/jeb.12139. (Visited on 12/18/2023) (cited on pages 4, 6).

[75] Tanja Stadler. 'Lineages-through-time plots of neutral models for speciation'. In: *Mathematical Biosciences* 216.2 (Dec. 1, 2008), pp. 163–171. DOI: 10.1016/j.mbs.2008.09.006. (Visited on 12/19/2023) (cited on page 46).

[76] Tanja Stadler. 'On incomplete sampling under birth–death models and connections to the sampling-based coalescent'. In: *Journal of Theoretical Biology* 261.1 (Nov. 2009), pp. 58–66. DOI: 10.1016/j.jtbi.2009.07.018. (Visited on 08/10/2023) (cited on pages 4, 77).

[77] Tanja Stadler. 'Sampling-through-time in birth–death trees'. In: *Journal of Theoretical Biology* 267.3 (Dec. 2010), pp. 396–404. DOI: 10.1016/j.jtbi.2010.09.010. (Visited on 07/19/2023) (cited on page 4).

[78] Tanja Stadler. 'Inferring speciation and extinction processes from extant species data'. In: *Proceedings of the National Academy of Sciences* 108.39 (Sept. 27, 2011). Publisher: Proceedings of the National Academy of Sciences, pp. 16145–16146. DOI: 10.1073/pnas.1113242108. (Visited on 12/18/2023) (cited on page 7).

[79] Tanja Stadler. 'Mammalian phylogeny reveals recent diversification rate shifts'. In: *Proceedings of the National Academy of Sciences* 108.15 (Apr. 12, 2011). Publisher: Proceedings of the National Academy of Sciences, pp. 6187–6192. DOI: 10.1073/pnas.1016876108. (Visited on 12/18/2023) (cited on pages 4, 5).

[80] Tanja Stadler. 'How Can We Improve Accuracy of Macroevolutionary Rate Estimates?' In: *Systematic Biology* 62.2 (Mar. 1, 2013), pp. 321–329. DOI: 10.1093/sysbio/sys073. (Visited on 09/12/2023) (cited on pages 4, 75).

[81] Richard P. Stanley. *Enumerative Combinatorics*. 1st ed. Cambridge University Press, Jan. 13, 1999. (Visited on 08/08/2023) (cited on page 56).

[82] M. A. Steel. *Phylogeny: discrete and random processes in evolution*. CBMS-NSF regional conference series in applied mathematics 89. Philadelphia: Society for Industrial and Applied Mathematics, 2016. 293 pp. (cited on pages 53, 54, 57).

[83] Mike Steel and Andy McKenzie. 'Properties of phylogenetic trees generated by Yule-type speciation models'. In: *Mathematical Biosciences* (2001) (cited on page 78).

[84] J. F. Steffensen. 'Om Sandsynligheden for at Afkommet uddør'. In: *Matematisk tidsskrift. B* (1930). Publisher: Mathematica Scandinavica, pp. 19–23. (Visited on 12/18/2023) (cited on page 3).

[85] E. A. Thompson. *Human evolutionary trees*. Cambridge [Eng.] ; New York: Cambridge University Press, 1975. 158 pp. (cited on page 4).

[86] V. A. Vatutin. 'Distance to the nearest common ancestor in bellman-harris branching processes'. In: *Mathematical notes of the Academy of Sciences of the USSR* 25.5 (May 1, 1979), pp. 378–382. DOI: 10.1007/BF01224843. (Visited on 12/13/2023) (cited on pages 8, 32).

[87] Chris Venditti, Andrew Meade, and Mark Pagel. 'Phylogenies reveal new interpretation of speciation and the Red Queen'. In: *Nature* 463.7279 (Jan. 2010), pp. 349–352. DOI: 10.1038/nature08630. (Visited on 09/05/2023) (cited on page 7).

[88] H. W. Watson and Francis Galton. 'On the Probability of the Extinction of Families'. In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1875). Publisher: [Royal Anthropological Institute of Great Britain and Ireland, Wiley], pp. 138–144. DOI: 10.2307/2841222. (Visited on 12/18/2023) (cited on page 3).

[89] David Williams. *Probability with Martingales*. 1st ed. Cambridge University Press, Feb. 14, 1991. (Visited on 11/07/2023) (cited on page 64).

[90] Z. Yang and B. Rannala. 'Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method'. In: *Molecular Biology and Evolution* 14.7 (July 1, 1997), pp. 717–724. DOI: 10.1093/oxfordjournals.molbev.a025811. (Visited on 07/09/2023) (cited on page 4).

[91] G. Udny Yule. 'A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.' In: *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 21 (1925), pp. 21–87 (cited on pages 2, 46).