

On the construction and topology of multi-type ancestral trees

Mariolys Rivas

A Thesis
In the Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy (Mathematics) at
Concordia University
Montreal, Quebec, Canada

September, 2014
©Mariolys Rivas, 2014

**CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES**

This is to certify that the thesis prepared

By: **Mariolys Rivas**

Entitled: **On the construction and topology of multi-type ancestral trees**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Approved by _____ Chair

Approved by Dr. J. Jaeger

Approved by _____ External Examiner

Approved by Dr. S. Lessard

Approved by _____ External to Program

Approved by Dr. J. Lessard

Approved by _____ Examiner

Approved by Dr. W. Sun

Approved by _____ Examiner

Approved by Dr. X. Zhou

Approved by _____ Thesis Supervisor

Approved by Dr. L. Popovic

Approved by _____

Approved by **Dr. J. Garrido, Graduate Program Director**

August, 2014

Dr. J. Locke, Dean
Faculty of Arts and Science

ABSTRACT

On the construction and topology of multi-type ancestral trees.

Mariolys Rivas, Ph.D.

Concordia University, 2014

Branching processes or Galton-Watson processes can be used to model the genealogy of a population of different species, where birth and death events represent speciation and extinction. In the more general context of multi-type branching processes, species are classified under phenotypical traits, and the probability of speciation and extinction is dependent on individual types. Since most accessible biological data concerns surviving species, it becomes necessary to extract information about the shape of genealogical trees from the available knowledge on the standing population, and to devise random models allowing backward reconstruction of ancestry under the rules of a particular branching process. We present two investigations on the topology of ancestral multi-type branching trees, generalizing several known results from the single-type case, and obtaining some new results that can only be formulated in the multi-type setting.

In the first part of the thesis, we present a backward construction algorithm for the ancestral tree of a planar embedding of a multi-type Galton-Watson tree assumed to be quasi-stationary, and we derive formulae for the conditional distribution of the time to the most recent common ancestor of two consecutive individuals at the present time, and of two individuals of the same type. We specialize these formulae to multi-type linear-fractional branching processes, and observe some effects of the symmetry of the

parameters in the two-type case.

In the second part of the thesis, we extend the concepts of cherries and pendant edges from rooted binary trees to the multi-type setting, and compute expressions and asymptotic properties for mean numbers and variances of these structures under the neutral two-type Yule model.

We explain how type mutations appear naturally in ancestral trees of multi-type birth-death processes, and show that these ancestral trees are Markovian and behave as pure-birth processes, by giving explicit time-dependent rates. We derive formulae and asymptotic properties for the mean number of cherries and pendant edges of each type in a multi-type pure-birth process with mutations. We show that sometimes it is possible to recover the defining rates of this process from the asymptotic proportion of leaves, cherries and pendant edges of each type.

This thesis is dedicated to the memory of my dad, Roberto Rivas.

Acknowledgements

I would like to acknowledge my supervisor for proposing the problems for our research and for all her help in tackling them. I would also like to thank the members of the committee for reading this work and for their useful suggestions.

I owe especially big big thanks to my forever boyfriend (a.k.a husband) Adolfo. Thanks for sharing with me all your mathematical advices, for motivating me and for your patience during these years. Thanks for being there in every moment of my career and thanks for making it more enjoyable.

I am very grateful and thankful for all the support from my mom and dad during this journey, for their advice and daily calls. They always encourage me to continue despite the distance and the tough moments.

Thanks also to all my friends from Montreal. The ones that are still here and the ones that left, for our lunch breaks on exhausting days and the adventures we have shared.

Gracias a Dios y a la Virgen del Valle.



Contents

List of Figures	ix
Introduction	1
1 Background	8
1.1 Multi-type branching processes	8
1.1.1 Multi-type linear-fractional branching processes	13
1.1.2 Single-type Galton-Watson trees	15
1.1.3 Single-type coalescent point processes	16
1.2 Birth-death processes	22
1.2.1 Birth-death processes	23
1.2.2 Multi-type birth-death processes	25
1.2.3 Yule model	26
1.3 Extended Pólya urn models	28
2 The coalescent point process of multi-type branching trees	32
2.1 Multi-type coalescent point processes	32
2.2 Special case: Multi-type linear-fractional branching processes	55
2.2.1 Coalescent times in linear-fractional ancestral tree	56
2.2.2 Comparison of ancestral trees in two-type models	63

3	Distribution of k-type cherries	68
3.1	Distribution of two-type cherries in neutral two-type binary trees	69
3.1.1	Probability distribution of two-type cherries	71
3.1.2	Asymptotics on two-type cherries	75
3.1.3	Pólya urn representation	77
3.1.4	Robust solvability of the reconstruction problem for neutral multi-type Yule trees	83
3.2	Distribution of k-type cherries in multi-type Yule trees	86
3.2.1	Multi-type Yule trees with mutations	94
3.2.2	Predictive accuracy of root type estimation in continuous neutral two-type Yule models	108
3.2.3	Monotonicity of the asymptotic proportion of cherries in certain two-type models	111
4	Conclusion and outlook	117

List of Figures

1	Sample tree of a branching process and sample tree of a birth-death process.	2
2	Phylogenetic trees of Bembidions and Phasmids..	3
1.1	Planar embedding of a single-type branching process (left), and the coalescent times along the ancestral lineages A_i (right)	18
1.2	Values of A_i , $C_{i,j}$ and $D_i(n)$ on a planar embedding of a branching process.	20
2.1	Values of \mathbf{A}_i and $\mathbf{D}_i(n)$ on a planar embedding of a branching process. .	35
2.2	Planar embedding of a two-type branching process (left), and types along the ancestral lineages \mathbf{A}_i (right)..	36
2.3	Sequences $(\mathbf{A}_i, \mathbf{D}_i(\cdot))_{i \geq 1}$ of surviving offspring types along the lineages of individuals $((0, i))_{i \geq 1}$ corresponding to the two-type ancestral tree given in Figure 2.2.	37
2.4	Sequences $\boldsymbol{\eta}_{A_i-1, \mathbf{D}_i(A_i)_{[2]}}(n)$ of surviving subtrees $\mathbf{D}_{i+1}(n)$ ($1 \leq n \leq A_i - 1$) are illustrated for the case $i = 5$	39
3.1	Two-type cherries and pendant edges.	70
3.2	Example of a two-type tree.	70
3.3	Sample trees of birth-death process.	87
3.4	The ancestral tree of Figure 3.3 and the corresponding two-type Yule tree with mutations	94

Introduction

Branching processes comprise a popular topic in probability because of their applications in a wide range of areas, including biology, physics and areas of pure mathematics. A *branching process* is a homogeneous Markov chain $(Z_n)_{n \geq 0}$ describing the size of a population in time. The main defining feature of a branching process is that the sub-process spanned by each individual from the time it was born, is independent of the individual's ancestry, and of the processes spanned by the other individuals existing at that time. Moreover, all of these sub-processes are equally distributed. The time scale of a branching process is usually *discrete*, so that individuals exist for a unit of time, after which they may produce a random number of offsprings. In many applications it is assumed that branching processes start with a single individual, and thus they may be visualized as planar rooted trees, where each node represents an individual which gives birth to as many offsprings as the node's out-degree. Figure 1.(a) shows part of a branching process' sample tree, started with a single individual at time 0. The time levels $0, 1, 2 \dots$ represent the *generations* of the process, and the number of individuals at each generation are the values Z_0, Z_1, \dots of the process itself.

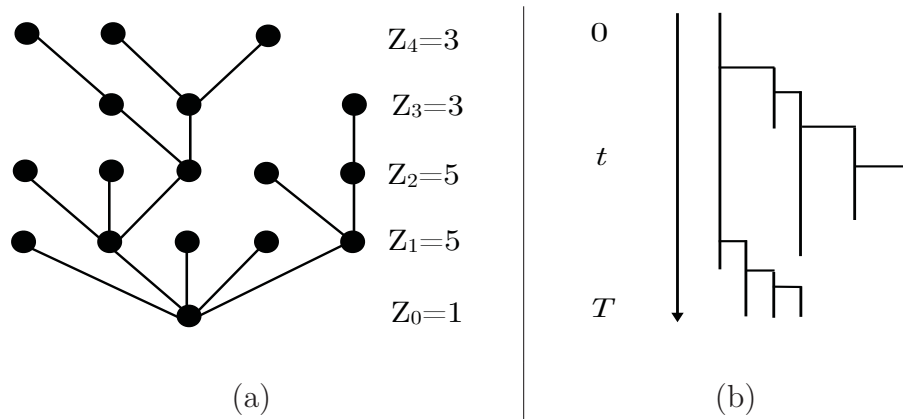


Figure 1: (a) Sample tree of a branching process. (b) Sample tree of a birth-death process.

Continuous-time generalizations of branching processes may be defined by allowing birth and death events to happen at any point in time. The most common extension is given by the continuous-time *birth-death process*, described in Section 1.2, where events occur randomly according to certain finite *rates*. A birth-death process Z_t is indexed by real numbers $t \geq 0$, and may be visualized as a binary tree with vertical edges which occasionally branch out to produce new offsprings. Figure 1.(b) displays a sample tree of a birth-death process with $Z_T = 4$, started by a single individual at time 0 ($Z_0 = 1$).

Branching processes may be used to model, for instance, a population of bacteria which reproduce asexually, since each offspring has a single parent. A crucial application of branching processes is the modelling of *phylogenetic trees*. These are trees that display the genetic ancestry of a population of different species containing information about their speciation and extinction. In a phylogenetic tree, each node represents a different species, and nodes may be further classified from phenotypic traits to obtain *multi-type* phylogenetic trees. A phylogenetic tree is often estimated from the information of extant species at the present time. In a branching process, the population at the present time is called the *standing population*. Its ancestry forms a subtree of the original branching tree, from which extinct lineages have been pruned. This subtree is often referred to as

the *reduced, ancestral or reconstructed tree* of the process. Figure 2 shows two examples of phylogenetic trees with their corresponding reduced trees. One of them is a *single-type* tree, in the sense that its nodes are not classified in any way, and the other one is a two-type phylogenetic tree, whose nodes are classified according to a specific trait (winged insects versus non-winged ones).

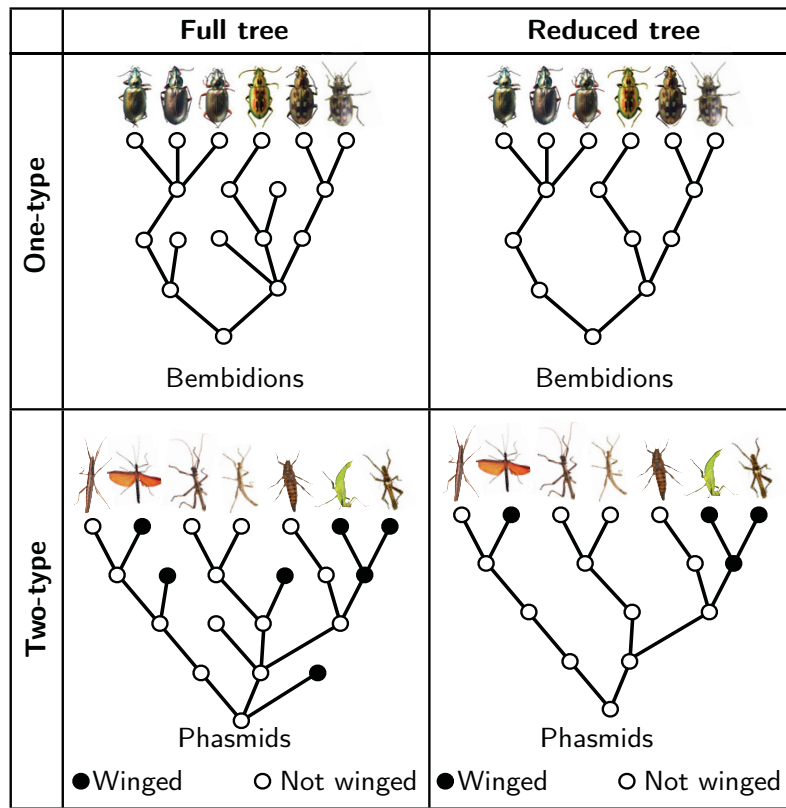


Figure 2: Phylogenetic trees of Bembidions and Phasmids. The first row shows the phylogenetic tree and the corresponding reduced tree of a genus of beetles known as Bembidions, and the second one shows a two-type phylogenetic tree of an order of insect known as Phasmida, classified by the presence of wings. The information used to draw these trees can be found in [45] and [9] respectively.

From the biological point of view, phylogenetic trees may be inferred backward in

time by observing the genetic makeup and phenotypic traits of extant species. For example, two very similar species are likely to have a more recent common ancestor than two very distinct ones. On the other hand, a sample tree of a discrete branching process may be constructed recursively, forward in time, by allowing each subsequent individual to produce as many offsprings as an independent instance of a particular random variable described by a probability distribution. From the distribution of a branching process' sample tree, one may compute formulae for certain statistics relating to the standing population, such as the times to the most recent common ancestors of consecutive individuals (See [29]). Conversely, from these formulae, it is sometimes possible to reconstruct the defining probabilities of the branching process by observing the topology of a sufficiently large sample tree. In the biological context, this means being able to understand the parameters that give way to a branching process modelling a particular phylogenetic tree which was inferred from biological features. This highlights the importance of devising formulae for statistics describing the relationships between individuals of the standing population in a branching process, which is one of the main motivations of our research.

Furthermore, since current biological data exists only for surviving species, it is productive to be able to construct a random branching process backward in time from the standing population. Typical methods for doing these reconstructions use simulations approaches, more specifically distance-based methods, maximum parsimony, maximum likelihood, and Markov chain Monte Carlo (MCMC) methods. For achieving this, it is necessary to make assumptions about the diversification rates (and their time-dependency in the continuous case), the root types or the stationarity of the processes (see Maddison et al. [32] and Igc & Goldberg [17]). Geiger [14], and Lambert & Popovic [29] have described backward constructions of infinite single-type sample trees. Specifically, Lambert & Popovic [29] use the *coalescent point process* which is a process describing the times to the most recent common ancestors of consecutive individuals from the standing

population. This is not to be confused with the coalescent process used in genetics by which all alleles of a gene shared by all members of a population are traced to a single ancestral copy. Recent renewed interest in multi-type branching processes motivates us to generalize these constructions to the multi-type setting. We tackle this problem in Chapter 2. These generalizations are not straightforward, mainly because in a multi-type process, the number and types of the offsprings are dependent on the type of the parent, and thus a sample tree of a multi-type branching process may not be constructed by randomly assigning types to a single-type sample tree.

Birth-death processes are often employed as analysis and prediction tools for real biological systems. They may serve as continuous phylogenetic models, and as discrete branching processes, their parameters may be deduced or adjusted from empirical biological data by observing the relationships between individuals of the standing population. It turns out that the ancestral tree of a birth-death process is itself a pure-birth process (see [39]), where the birth rate is dependent on time.

Birth-death processes are intimately related to the Yule model for the construction of random rooted binary trees. Indeed, the shape of the ancestral tree of a single-type birth-death process is distributed as a Yule tree (a rooted binary tree constructed under the Yule model). For a proof of this fact see Aldous [1]. A particularly interesting statistic on these shapes is the number of *cherries* (pairs of leaves adjacent to a common ancestor) in a rooted binary tree. McKenzie & Steel [34] obtained the distribution of the number of cherries in a Yule tree. Our main motivation for Chapter 3 is to extend some results from [34] onto the case when the nodes of the random trees have different types. In particular this means that there are different types of cherries as well. In the particular setting of the neutral two-type Yule model, we derive some results by solving classical recurrences on the generating functions of cherry counts of different types (Sections 3.1.1 and 3.1.2). Additional results on the asymptotics of these counts

arise from an *extended Pólya urn process* (Section 3.1.3). If we restrict ourselves to asymptotics, the results obtained from the Pólya urn approach are much stronger than what can be computed from individual generating functions, since from this approach we also obtain eigenvalues/correlated growth of various quantities in terms of the number n of leaves. Furthermore, the Pólya urn approach proves convergence in distribution, which is stronger than simple limits of means and variances. On the other hand the generating functions approach provides exact results for means and variances for finite values of n .

In Section 3.2 we show that the ancestral tree of a multi-type birth-death process is itself a pure-birth process with certain time-dependent rates, thus generalizing the single-type result from [39]. The ancestral tree of a multi-type birth-death process turns out to be a lot more complex than that of a single-type process, as the former allows individuals to *mutate* to different types at certain rates. This motivates the study of multi-type Yule trees which allow mutations, and where the probability of choosing an individual of type i at every step is weighted by a parameter which is dependent on i . This weight constraint makes the generating function and the extended Pólya urn approaches untreatable. We thus use a different method that considers a continuous model which embeds the discrete Yule model. We present explicit relations between probabilities and rates of the two models as a motivation for the asymptotic results we obtain next concerning the mean numbers of cherries and *pendant edges* (these are edges which are not part of a cherry) in the multi-type pure-birth process with mutations. When rates are time-independent, we illustrate how it is possible to obtain the rates defining the process from those asymptotic values.

This thesis is organized as follows. In Chapter 1 we present the classical definitions and terminology we use throughout this document as well as some motivation for our results. In Chapter 2, Section 2.1 we introduce the multi-type coalescent point process and describe an explicit and algorithmic way to construct an ancestral tree of the standing

population of a (quasi-stationary) multi-type branching process in terms of a Markov chain. We also provide explicit formulae for calculating the basic statistical features that describe the ancestral tree (the law of coalescence times together with the types on the ancestral lineages), as well as statistical features that link types in the standing population with the shape of the tree (the law of same-type coalescence times). In Section 2.2 we consider the special case of a multi-type branching process with linear-fractional offspring distribution, and we obtain very simple formulae for these two sets of statistical features. These formulae are then used to assess the differences in the ancestral trees of two different linear-fractional offspring distributions: one ‘symmetrical’ in its treatment of different offspring types, and the other completely ‘asymmetrical’ in that sense. In Chapter 3, Sections 3.1.1 and 3.1.2, we derive explicit formulae for the mean and variance of the number of two-type cherries in random binary two-type trees under a neutral Yule model, as well as the asymptotic distribution of those same statistics. In Section 3.2 we present the long time behaviour of the mean number of cherries and pendant edges of different types in k -type Yule trees allowing mutations, by embedding this multi-type Yule model in a multi-type pure-birth process. We also examine this continuous multi-type pure-birth process under time-dependent rates. At the end of this section we illustrate some reconstruction properties of these statistics to obtain information about the birth and death rates of the process.

The results presented in Chapters 2 and 3 may be found in the following two joint works with Dr. Lea Popovic:

L. Popovic, M. Rivas. The coalescent point process of multi-type branching trees. *Stochastic Processes and Applications*, Vol 124(12), 29 pp, 2014

L. Popovic, M. Rivas. Multi-type Yule trees. *In Preparation*.

Chapter 1

Background

In this chapter we present some theoretical foundations regarding the theory of branching processes and birth-death processes. We also summarize some existing results that motivate our work. We may occasionally refer to branching processes as *Galton-Watson processes* or *G-W processes*, since the formal definition we use below for branching processes arose originally from H. W. Watson and F. Galton's study on the extinction of families [12].

1.1 Multi-type branching processes

We start by introducing some notation on multi-type branching processes. Let $\{1, 2, 3, \dots\}$ denote a countable space of *types* of a population.

Definition 1.1. *A multi-type branching process or multi-type Galton-Watson process (G-W process) is a discrete-time Markov process $\mathbf{Z} = (\mathbf{Z}^{(n)})_{n \geq 0}$, where $\mathbf{Z}^{(n)} = (Z_1^{(n)}, Z_2^{(n)}, \dots)$ is an infinite-dimensional random vector whose ℓ -th coordinate is the number of individuals of type ℓ at generation n in a population. The offspring vector of each individual of type ℓ follows a fixed distribution $(\mathbf{Z}^{(1)} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$ (where \mathbf{e}_ℓ denotes the unit vector with a 1 in the ℓ -th coordinate), and this distribution is independent of the individual's*

ancestors and their siblings.

It is assumed that only a finite number of entries of the vector $\mathbf{Z}^{(n)}$ are non-zero. This may be guaranteed as long as the random vector $(\mathbf{Z}^{(1)} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$ has finite sum of entries for all ℓ . Generations are indexed by $n \in \mathbb{N}_0$ in the superscript, and types are indexed by $\{1, 2, \dots\}$ in the subscript. All the arguments in this Chapter and Chapter 2 are valid as well for a finite number of types.

For any $\mathbf{z} = (z_1, z_2, \dots) \in \mathbb{N}_0^\infty$ with $z_1 + z_2 + \dots < \infty$, the transition probabilities and the n -th iterations are denoted by

$$P_\ell(\mathbf{z}) = \mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell), \quad P_\ell^{(n)}(\mathbf{z}) = \mathbb{P}(\mathbf{Z}^{(n)} = \mathbf{z} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell).$$

For $\mathbf{s} = (s_1, s_2, \dots)$, the probability generating function of the offspring distribution is denoted by $\mathbf{f}(\mathbf{s}) := (f_1, f_2, \dots)(\mathbf{s})$ where

$$f_\ell(\mathbf{s}) = \mathbb{E}(\mathbf{s}^{\mathbf{Z}^{(1)}} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \sum_{\mathbf{z}=(z_1, z_2, \dots) \in \mathbb{N}_0^\infty} P_\ell(\mathbf{z}) s_1^{z_1} s_2^{z_2} \dots, \quad \text{for } |s_1|, |s_2|, \dots \leq 1,$$

and the probability generating function of the n -th generation population, the n -fold composition of $\mathbf{f}(\mathbf{s})$, is denoted by $\mathbf{f}^{(n)}(\mathbf{s})$ where $f_\ell^{(n)}(\mathbf{s}) = \mathbb{E}(\mathbf{s}^{\mathbf{Z}^{(n)}} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$. For $n = 0$ let $f^{(0)}(\mathbf{s}) = \mathbf{s}$, and note that $f^{(1)}(\mathbf{s}) = \mathbf{f}(\mathbf{s})$.

We let $\mathbf{M} = (m_{\ell\ell'})_{1 \leq \ell, \ell' \leq \infty}$ be the matrix of the expected number of offspring of each type from parents of different types:

$$m_{\ell\ell'} = \mathbb{E}(Z_{\ell'}^{(1)} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \left. \frac{\partial f_\ell(\mathbf{s})}{\partial s_{\ell'}} \right|_{\mathbf{s}=\mathbf{1}}, \quad \text{for } \ell, \ell' = 1, 2, \dots$$

where $\mathbf{1} = (1, 1, \dots)$ and we assume all $m_{\ell\ell'} < \infty$. This is also called the *mean matrix* of the process. It is immediate that the mean matrix of the n -th generation population \mathbf{Z}_n is \mathbf{M}^n .

Definition 1.2. A multi-type Galton-Watson process is called *positive regular* (or *irreducible*) if for some $n > 0$ the mean matrix of its n -th generation population \mathbf{M}^n is positive (all of its entries $m_{\ell\ell'}^{(n)} > 0$ are strictly positive entries).

Definition 1.3. A process is called *singular* if each individual has exactly one offspring.

We assume that the multi-type G-W process is non-singular and irreducible throughout this work.

Definition 1.4. A matrix \mathbf{M} is said to be *aperiodic* if for all ℓ , the greatest common divisor of all natural numbers n such that $m_{\ell\ell}^{(n)} > 0$, is equal to 1.

Theorem 1.1 (Seneta [43]). All the power series from the entries of $\mathbf{M}(s) = \sum_{n \geq 0} s^n \mathbf{M}^n$ have a common convergence radius R , for $0 \leq R < \infty$.

Definition 1.5. The matrix \mathbf{M} is called *R-transient* if $\sum_{n \geq 0} m_{ii}^{(n)} R^n < \infty$ for $i \geq 1$ and *R-recurrent* if $\sum_{n \geq 0} m_{ii}^{(n)} R^n = \infty$ for $i \geq 1$, where R is the common convergence ratio from the previous theorem.

The following theorem follows as an extension of the well known Perron-Frobenius theorem.

Theorem 1.2 (Seneta [43]). For an *R-recurrent* matrix \mathbf{M} there exist unique (up to constant multipliers) positive vectors \mathbf{u} and \mathbf{v} such that

$$R\mathbf{M}\mathbf{v}^\top = \mathbf{v}^\top \text{ and } R\mathbf{u}\mathbf{M} = \mathbf{u}.$$

where \mathbf{v}^\top represents the transpose of the vector \mathbf{v} .

Definition 1.6. In the *R-recurrent* case the matrix \mathbf{M} is called *R-null* when $\mathbf{u}\mathbf{v}^\top = \infty$ and *R-positive* when $\mathbf{u}\mathbf{v}^\top < \infty$.

Theorem 1.3 (Seneta [43]). *For an R -positive matrix \mathbf{M} after scaling the vectors \mathbf{u}, \mathbf{v} such that $\mathbf{v} \cdot \mathbf{u} = \mathbf{v}\mathbf{u}^\top = 1$,*

$$R^n \mathbf{M}^n \rightarrow \mathbf{u}^\top \mathbf{v}, \text{ as } n \rightarrow \infty.$$

In our work we assume that \mathbf{M} is aperiodic, R -recurrent and R -positive. In this case, using the previous results we ensure the existence of R, \mathbf{u} and \mathbf{v} .

We say that a multi-type Galton-Watson process is *transient (recurrent, positive)* if the mean matrix \mathbf{M} is R -transient (R -recurrent, R -positive).

We scale the vectors \mathbf{u} and \mathbf{v} so that $\mathbf{u} \cdot \mathbf{v} = 1$ and $\mathbf{u} \cdot \mathbf{1} = 1$.

Definition 1.7. *Set $\rho = 1/R$. The process \mathbf{Z} is subcritical, critical or supercritical when $\rho < 1$, $\rho = 1$ or $\rho > 1$, respectively.*

Since $\rho = 1/R$, equivalently we say that the process is *subcritical, critical or supercritical* if $R > 1$, $R = 1$ or $R < 1$, respectively. Thus, the role that ρ plays in the countably many types setting is similar to the role of $\mu = \mathbb{E}(Z^{(1)} | Z^{(0)} = 1)$ in the one type case.

The next three definitions involve Markov processes in general.

Definition 1.8. *Let $\mathbf{Z} = (\mathbf{Z}_t)_{t \geq 0}$ be a Markov process on $[0, \infty)$. A quasi-stationary distribution of \mathbf{Z} is a positive measure ν such that*

$$\nu(A) = \mathbb{P}(\mathbf{Z}_t \in A | \mathbf{Z}_t \neq \mathbf{0}), \quad t \geq 0.$$

Definition 1.9. *The Yaglom limit \mathcal{Y} of a Markov process \mathbf{Z} is a random variable whose law is a quasi-stationary distribution defined by:*

$$\mathbb{P}(\mathcal{Y} \in A) := \lim_{t \rightarrow \infty} \mathbb{P}_z(\mathbf{Z}_t \in A | \mathbf{Z}_t \neq \mathbf{0}) = \lim_{t \rightarrow \infty} \mathbb{P}(\mathbf{Z}_t \in A | \mathbf{Z}_t \neq \mathbf{0}, \mathbf{Z}_0 = z), \quad z \neq \mathbf{0}.$$

Definition 1.10. *The Q-process $\widehat{\mathbf{Z}}$ is a Markov process such that for any $\Theta \in \mathcal{F}_t$, where*

\mathcal{F}_t is the sigma algebra generated by $\mathbf{Z}_{t'}$ ($0 \leq t' \leq t$),

$$\widehat{\mathbb{P}}_z(\Theta) := \mathbb{P}(\widehat{\mathbf{Z}} \in \Theta \mid \widehat{\mathbf{Z}}_0 = z) = \lim_{s \rightarrow \infty} \mathbb{P}_z(\Theta \mid \mathbf{Z}_{t+s} \neq 0).$$

It can be shown (See Dallaporta & Joffe [6]), as in the one-playsdimensional case of a Galton-Watson process (see [14]), that the transition probabilities of the Q -process are given by

$$\widehat{\mathbb{P}}(\mathbf{Z}^{(n)} = z \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \frac{P_\ell(z)}{\rho} \frac{z \cdot \mathbf{v}}{\mathbf{e}_\ell \cdot \mathbf{v}},$$

where $P_\ell(z) = \mathbb{P}(\mathbf{Z}^{(1)} = z \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$ represents the transition probability of a Galton-Watson process at generation n . This representation of the probabilities of the Q -process plays an important role in Chapter 2 when reconstructing the spine of the process.

Theorem 1.4 (Seneta [43]). *Suppose that $(\mathbf{Z}^{(n)})_{n \geq 0}$ is irreducible, aperiodic, positive with the mean matrix \mathbf{M} being R -positive for some $R > 1$, and if the left eigenvector of \mathbf{M} satisfies $\sum_i u_i < \infty$, then the limits*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\mathbf{Z}^{(n)} = z \mid \mathbf{Z}^{(n)} \neq \mathbf{0}), \text{ and}$$

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}_\ell(\mathbf{Z}^{(n)} = z \mid \mathbf{Z}^{(n+m)} \neq \mathbf{0}),$$

exist and define honest probability distributions, meaning that their sums over all values of z add up to 1.

The previous theorem ensures the existence of the Yaglom limit, and of the limit of the Q -process for a multi-type Galton-Watson process in the subcritical case.

The existence of the limit of the first quasi-stationary distribution is needed for the assumption in Chapter 2 that the infinite embeddings of multi-type trees are quasi-stationary. The second limit is a key ingredient in a construction of the spine as it is explained in Remark 2.2.

1.1.1 Multi-type linear-fractional branching processes

Linear-fractional branching processes form an important family of particular Galton-Watson processes. They satisfy the convenient property that their iterations are again linear-fractional.

We use the same notation as [41] for ease of drawing on known results and making comparisons. Let \mathbf{H} be an infinite square sub-stochastic matrix, that is, each row \mathbf{h}_ℓ of \mathbf{H} is a non-negative vector with $\sum_{\ell' \geq 1} h_{\ell\ell'} \leq 1$, and let $h_{\ell 0} = 1 - \sum_{\ell' \geq 1} h_{\ell\ell'}$. Let \mathbf{g} be a non-negative vector such that $\mathbf{g}\mathbf{1}^\top = \sum_{\ell' \geq 1} g_{\ell'} = 1$. Let $m > 0$. For any $\mathbf{z} = (z_1, z_2, \dots) \in \mathbb{N}_0^\infty$, let $|\mathbf{z}| = \sum_{\ell \geq 1} z_\ell$.

Definition 1.11. *A random vector $\boldsymbol{\xi}_\ell$ taking values in \mathbb{N}_0^∞ has a linear-fractional distribution LF $(\mathbf{h}_\ell, \mathbf{g}, m)$ if for any non-negative integer vector $\mathbf{z} = (z_1, z_2, \dots)$*

$$\mathbb{P}(|\boldsymbol{\xi}_\ell| = 0) = h_{\ell 0}, \quad \mathbb{P}(\boldsymbol{\xi}_\ell = \mathbf{e}_{\ell'} + \mathbf{z}) = h_{\ell\ell'} \frac{m^{|\mathbf{z}|}}{(1+m)^{|\mathbf{z}|+1}} \begin{pmatrix} |\mathbf{z}| \\ z_1, z_2, \dots \end{pmatrix} \mathbf{g}^{\mathbf{z}},$$

where $\mathbf{g}^{\mathbf{z}} = g_1^{z_1} g_2^{z_2} \dots$. The probability generating function of $\boldsymbol{\xi}_\ell$ has the linear fractional form

$$f_\ell(\mathbf{s}) = h_{\ell 0} + \frac{\sum_{\ell' \geq 1} h_{\ell\ell'} s_{\ell'}}{1 + m - m \sum_{\ell' \geq 1} g_{\ell'} s_{\ell'}}.$$

One can also represent the random vector $\boldsymbol{\xi}_\ell$ as a sequence of offsprings, where the first offspring has a type distribution given by \mathbf{h}_ℓ and the offsprings after the first one have geometric distribution with mean m and type distribution given by \mathbf{g} independently for each offspring. Moreover, the probability generating function of $(\boldsymbol{\xi}_\ell | \boldsymbol{\xi}_\ell \neq \mathbf{0})$ is that of a shifted multivariate-geometric distribution

$$\mathbb{E}(\mathbf{s}^{\boldsymbol{\xi}_\ell} | \boldsymbol{\xi}_\ell \neq \mathbf{0}) = \frac{(1 - h_{\ell 0})^{-1} \sum_{\ell' \geq 1} h_{\ell\ell'} s_{\ell'}}{1 + m - m \sum_{\ell' \geq 1} g_{\ell'} s_{\ell'}}.$$

Definition 1.12. A multi-type linear-fractional branching process $\text{LF}(\mathbf{H}, \mathbf{g}, m)$ is a Galton-Watson process in which each individual of type $\ell \in \{1, 2, \dots\}$ reproduces according to the $\text{LF}(\mathbf{h}_\ell, \mathbf{g}, m)$ offspring distribution $\boldsymbol{\xi}_\ell$.

In other words, the probability generating function of the offspring distribution of a multi-type linear-fractional branching process is $\mathbf{f}(\mathbf{s}) = (f_1, f_2, \dots)(\mathbf{s})$ with f_ℓ as above. Its mean matrix is given by $\mathbf{M} = \mathbf{H} + m\mathbf{H}\mathbf{1}^\top\mathbf{g}$.

The independence of the parameters \mathbf{g} and m from the parent's type ensures that the population size in each generation of this process also has a linear fractional distribution. This fact is detailed by the following known result.

Theorem 1.5 (Proposition 3 [22], Theorem 3 [41]). *The n -th generation population size vector $\mathbf{Z}^{(n)}$ of a multi-type linear-fractional branching process $\text{LF}(\mathbf{H}, \mathbf{g}, m)$ started with one individual $\mathbf{Z}^{(0)} = \mathbf{e}_\ell$ has a linear-fractional distribution $\text{LF}(\mathbf{h}_\ell^{(n)}, \mathbf{g}^{(n)}, m^{(n)})$ whose parameters are given by:*

$$\begin{aligned} m^{(n)} &= m\mathbf{g}(\mathbf{I} + \mathbf{M} + \dots + \mathbf{M}^{n-1})\mathbf{1}^\top, \\ \mathbf{g}^{(n)} &= \frac{m}{m^{(n)}}\mathbf{g}(\mathbf{I} + \mathbf{M} + \dots + \mathbf{M}^{n-1}), \\ \mathbf{H}^{(n)} &= \mathbf{M}^n - \frac{m^{(n)}}{1 + m^{(n)}}\mathbf{M}^n\mathbf{1}^\top\mathbf{g}^{(n)}, \end{aligned} \tag{1.1}$$

where the vector $\mathbf{h}_\ell^{(n)}$ is the ℓ -th row of the matrix $\mathbf{H}^{(n)}$, and $\mathbf{1}^\top$ is the transpose of $\mathbf{1} = (1, 1, \dots)$.

Note that, as a consequence, $(\mathbf{Z}^{(n)} | \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$ is distributed as a shifted multivariate-geometric distribution

$$\mathbb{E}(\mathbf{s}^{\mathbf{Z}^{(n)}} | \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \frac{(1 - h_{\ell 0}^{(n)})^{-1} \sum_{\ell' \geq 1} h_{\ell \ell'}^{(n)} s_{\ell'}}{1 + m^{(n)} - m^{(n)} \sum_{\ell' \geq 1} g_{\ell'}^{(n)} s_{\ell'}}.$$

This theorem was proved in [22] using an algebraic approach, while [41] provided a

different proof using the jumping contour representation of the branching process and its nice Markovian structure.

1.1.2 Single-type Galton-Watson trees

In this section we summarize some of the initial results of Geiger [14], since we aim to extend them to the multi-type setting in the following chapter. In his work, Geiger studies planar Galton-Watson trees conditioned on non-extinction at the present time, having started somewhere in the past. We use the notation from the previous sections and introduce some new one.

Let \mathcal{T} be the a random family tree associated with the Galton Watson process $(Z^{(n)})_{n \geq 0}$ started with one individual. Thus \mathcal{T} is a rooted planar tree. According to our planar embedding (where individuals are located at points of a discrete lattice), we distinguish the first individual from left to right which has descendants at generation n .

To avoid ambiguity, we sometimes denote the Galton Watson process associated to a particular tree \mathcal{T} by $(Z^{(n)}(\mathcal{T}))_{n \geq 0}$. Given a Galton Watson tree \mathcal{T} and a fixed value of $Z^{(1)}(\mathcal{T})$, denote by \mathcal{T}^i the subtree founded by the i -th offspring of the root for $1 \leq i \leq Z^{(1)}(\mathcal{T})$. Let us denote by $R_{n+1}(\mathcal{T})$ the *position* or *rank* of the leftmost individual of the root who has descendants at generation $n + 1$, i.e,

$$R_{n+1}(\mathcal{T}) = \min\{1 \leq i \leq Z^{(1)}(\mathcal{T}) : Z^{(n)}(\mathcal{T}^i) \neq 0\}, \quad n \geq 0,$$

with $\min(\emptyset) = \infty$. With this notation, we recall the first result from [14].

Lemma 1.1 (Geiger [14]). *The subtrees \mathcal{T}^i , $1 \leq i \leq Z^{(1)}$, for $1 \leq j \leq k < \infty$ satisfy:*

$$(\mathcal{T}^i | R_{n+1} = j, Z^{(1)} = k) \stackrel{d}{=} \begin{cases} (\mathcal{T} | Z^{(n)} = 0), & 1 \leq i \leq j - 1, \\ (\mathcal{T} | Z^{(n)} > 0), & i = j, \\ \mathcal{T}, & j + 1 \leq i \leq k, \end{cases}$$

where the notation $\stackrel{d}{=}$ means “equally distributed” or “with the same law”.

Furthermore, the conditional joint distribution of R_{n+1} and $Z^{(1)}$, $n \geq 0$, is given by

$$\mathbb{P}(R_{n+1} = j, Z^{(1)} = k \mid Z^{(n+1)} > 0) = \frac{p_k \mathbb{P}(Z^{(n)} > 0) \mathbb{P}(Z^{(n)} = 0)^{j-1}}{\mathbb{P}(Z^{(n+1)} > 0)}, \quad (1.2)$$

where $p_k = \mathbb{P}(Z^{(1)} = k \mid Z^{(0)} = 1)$ denotes the probability that an individual has exactly k offsprings.

This previous result states that the subtrees \mathcal{T}^i are conditionally independent given $\{R_{n+1} = j, Z^{(1)} = k\}$. Geiger observes that Lemma 1.1 may be used to define a reverse construction of the Galton-Watson tree conditioned on nonextinction at the present time.

1.1.3 Single-type coalescent point processes

The *coalescent point process* of a branching tree is a process describing the genealogy of the *standing population* (the population at the present time) backwards in time, directly displaying the *coalescence times* as a sequence running over the current population size. It constructs a set of points, each corresponding to the most recent common ancestor of two individuals in the current population, whose depth (or vertical height) corresponds to the time when the lineages of these two individuals branched off (separated) from each other. The coalescent point process has a bijective correspondence with the ancestral tree of the current population, and allows the full ancestral tree to be reconstructed from its values. It was introduced in [40] for the ancestral tree of a continuous-time single-type branching process conditioned on its current population size, and called the *genealogical point process*. The convenient property of that particular branching model is that its contour process is Markovian, which implied that the points in this point process are *simple* -that is, each branch point has degree two- and that they are independent samples from the same distribution of depths. This allows one to reconstruct the ancestral tree of a population of n current individuals based simply on a sample of size n from this

distribution (see Fig. 2 of [40]). This genealogical point process was used in [2] to obtain statistical information for the ancestral trees of a critical branching process, was extended to non-critical binary processes in [16] and to homogeneous binary Crump-Mode-Jagers processes in [28].

The original construction of the genealogical point process had to be extended to accommodate Galton-Watson branching processes with general offspring distribution when the contour process of the branching tree is no longer Markovian. In this case depths of points in the process are no longer sufficient in order to fully reconstruct the ancestral tree, as the most recent common ancestors were no longer distinct for every pair of current individuals. In other words, branch points in the ancestral tree no longer always have degree exactly equal to two, and it was necessary to keep track of the multiplicity of these points as given by their branching degree. In [29] a construction was made which, rather than having all simple points with mass one, has points with (positive) integer valued masses. Each point again corresponds to a most recent common ancestor of two individuals in the current population, and its depth records the time when the two individuals' lineages separated. The additional mass coordinate of this point records the number of current individuals with the same most recent common ancestor as these two which are embedded after (or horizontally to the right) of them. This process was called the *coalescent point process (with multiplicities)*.

Before we present our multi-type extension of this construction we first recall the notation from [29].

Consider an arbitrarily large population at the present time from a general quasi-stationary branching process originating at an unspecified arbitrarily large time in the past. In the planar embedding of this process, individuals are located at points of a discrete lattice $(n \in \mathbb{Z}, i \in \mathbb{N})$, where the first coordinate $-n$ denotes the generation and the second coordinate i denotes the position of the individual in the planar embedding

from left to right. The number of offspring of individual (n, i) is denoted by $\xi(n, i)$. See Figure 1.1, the doubly infinite embedding of the tree is shown in the figure on the left. Its coordinates (n, i) represent its location on the two dimensional grid. The *standing population* is the population at the present time (generation $n = 0$), and its *ancestral tree* is the subtree of the branching tree obtained by following only the branches that lead to an individual present in the standing population. The ancestral tree is more easily observed in the right figure in Figure 1.1.

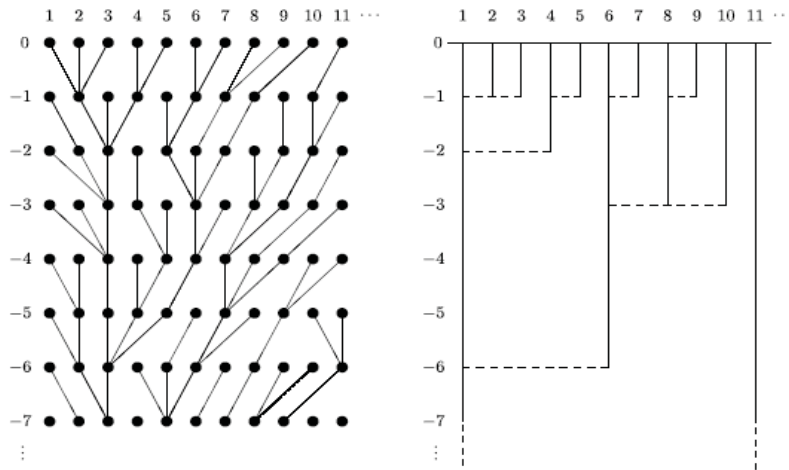


Figure 1.1: Planar embedding of a single-type branching process (left), and the coalescent times along the ancestral lineages A_i (right). Lambert & Popovic [29].

The ancestry of an individual from generation 0 can be traced backwards in time as follows. Define

$$a_i(n) := \text{index of the ancestor of individual } (0, i) \text{ at generation } -n.$$

Definition 1.13. *The coalescent time $C_{i,j}$ of individuals $(0, i)$ and $(0, j)$ is the time of*

the most recent common ancestor between these two, that is,

$$C_{i,j} := \min\{n \geq 1 : a_i(n) = a_j(n)\}, \text{ with } \min(\emptyset) = \infty.$$

In particular, define $A_i := C_{i,i+1}$ which identifies the coalescent time of individuals $(0, i)$ and $(0, i + 1)$. It can be easily shown that $C_{i,j} = \max\{A_i, A_{i+1}, \dots, A_{j-1}\}$.

Definition 1.14. *The sequence $(A_i)_{i \geq 1}$ is called the coalescent point process.*

The genealogy back in time of the present population, that is its ancestral tree, is then uniquely determined by the process $(A_i)_{i \geq 1}$. (This was sufficient information for the genealogical point process of binary branching processes in [40, 2, 16, 28].) The coalescent point-process (with multiplicities) can be read off from the right figure (in Figure 1.1) by observing the depth of the most recent common ancestor (MRCA) between two individuals and counting the multiplicities of its future appearance, e.g. $A_1 = 1$ with multiplicity 2, $A_2 = 1$ with multiplicity 1, $A_3 = 2$ with multiplicity 1, etc). Define an auxiliary process $(D_i)_{i \geq 1}$ of integer valued sequences $D_i = \{D_i(n), n \geq 1\}$ for each $i \geq 1$, which records future branch degrees along the ancestral lineage of individual $(0, i)$

$D_i(n) :=$ number of *surviving* offsprings of individual $(-n, a_i(n))$ embedded in the
ancestral tree to the right of the lineage of $(0, i)$ itself.

It turns out that the process $(D_i)_{i \geq 1}$ has all the nice properties needed to identify the law of the coalescent point-process. See Figure 1.2 for an illustration of these values on a sample tree.

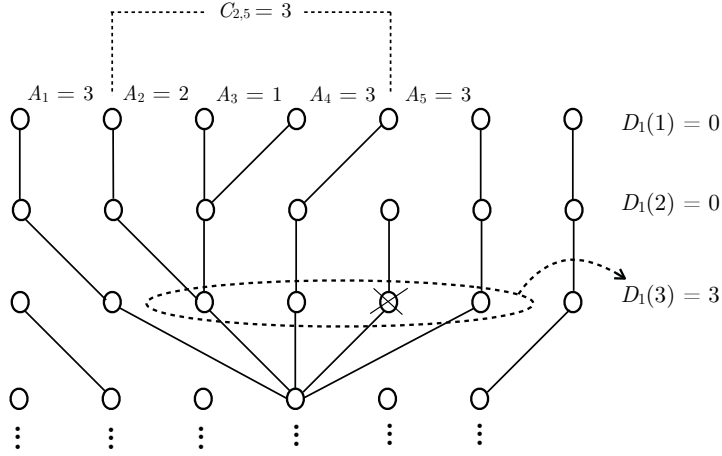


Figure 1.2: Some values of A_i , $C_{i,j}$ and $D_i(n)$ illustrated on a planar embedding of a branching process.

Theorem 1.6 (Theorem 2.1 of [29]). A_i is a functional of D_i given by

$$A_i = \min\{n \geq 1 : D_i(n) \neq 0\},$$

and the law of the process $(D_i, i \geq 1)$ is determined by the fact that it is a sequence-valued Markov chain, started at the null sequence $D_0 = (0, 0, \dots)$, with transitions given as follows - for any sequence $d. = (d_n; n \geq 0) \in \mathbb{N}_0^{\mathbb{N}}$

$$(D_{i+1}(n) \mid D_i(\cdot) = d.) \stackrel{d}{=} \begin{cases} d_n & \text{for } n > A_i, \\ d_{A_i} - 1 & \text{for } n = A_i, \\ \zeta'_n & \text{for } 1 \leq n < A_i, \end{cases}$$

where the random variables $\zeta'_1, \zeta'_2, \dots, \zeta'_{A_i-1}$ are independent random variables.

The distributions of variables $\{\zeta'_n\}_{n \geq 1}$ are specified as follows. If ξ is the offspring distribution of this Galton-Watson branching process with probability generating function $f(s)$, the random variables $\xi(n, j)$, representing the number of offspring of individual (n, j)

for any indices $n, j \in \mathbb{N}$, are all independent identically distributed as ξ . The survival probability to generation 0 of each offspring of an individual at generation $-n$ is given by $p_{n-1} := 1 - f^{(n-1)}(0)$ where $f^{(n-1)}$ is the $(n-1)$ -fold composition of f . This, in particular, holds for the offspring of $(n, a_i(n))$, the ancestor of $(0, i)$ at generation $-n$. If we let $\{\epsilon_n^1, \epsilon_n^2, \dots\}$ be an independent sequence of i.i.d. Bernoulli variables with parameter $\mathbb{P}(\epsilon_n^m = 1) = p_{n-1}$ (we deviate slightly notation from [29] here), and use an independent variable ξ , we can define the random sum

$$\zeta_n := \sum_{m=1}^{\xi} \epsilon_n^m,$$

and, for each $n \geq 1$, the law of ζ'_n is defined by

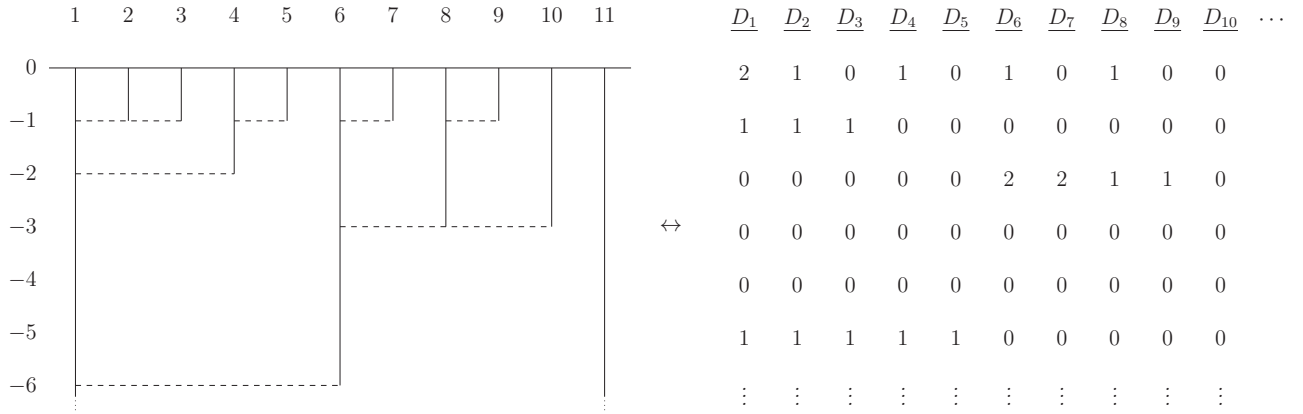
$$\zeta'_n := \stackrel{d}{=} (\zeta_n - 1 | \zeta_n \neq 0).$$

Example 1.1. *The previous Theorem shows how to obtain the sequence $(A_i)_{i \geq 0}$ from $(D_i(n))_{i \geq 0}$. The latter may also be encoded as the graph of the coalescent point process, where $D_i(n)$ is the number of dotted segments at generation n to the right of the point (n, i) as it is shown in the following figure.*

$$\begin{array}{cccccc}
C_{1,1} = 0 & C_{2,1} = 1 & C_{3,1} = 1 & \cdots & C_{6,1} = 6 & \cdots \\
C_{1,2} = 1 & C_{2,2} = 0 & C_{3,2} = 1 & \cdots & C_{6,2} = 6 & \cdots \\
C_{1,3} = 1 & C_{2,3} = 1 & C_{3,3} = 0 & \cdots & C_{6,3} = 6 & \cdots \\
C_{1,4} = 2 & C_{2,4} = 2 & C_{3,4} = 2 & \cdots & C_{6,4} = 6 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

$$A_1 = 1 \quad A_2 = 1 \quad A_3 = 2 \quad A_4 = 1 \quad A_5 = 6 \quad A_6 = 1 \quad A_7 = 3 \quad A_8 = 1 \quad A_9 = 3 \quad \cdots$$

↑



Remark 1.1. *The trees from [29] are closely related to the ones constructed in [14] (Section 1.1.2). In particular the offspring process of the ascendants of any element at generation 0 in [29] has the same distribution as the tree constructed in [14]. They are both planar Galton-Watson processes assuming non-extinction at the present time.*

1.2 Birth-death processes

Birth-death processes are used to simulate speciation in biology and other areas of science. In the next four subsections, some classical continuous-time birth and death processes are presented for both the single type case and the multi-type case.

1.2.1 Birth-death processes

A *continuous-time birth-death process* $(Y_t)_{t \geq 0}$ is a continuous-time Markov process on $[0, \infty)$ that is defined by parameters $\tilde{b}(i)$, $\tilde{d}(i)$, where $\tilde{b}(i)$ represents the rate at which one individual is born at time t given that $Y_t = i$ and $\tilde{d}(i)$ is the rate at which one individual dies out given that $Y_t = i$. It is assumed that the time interval $t \rightarrow t + \Delta t$ is infinitesimally small so that during this time interval at most one event occurs, either a birth or a death. Thus, the infinitesimal transition probabilities of a continuous-time birth-death process satisfy

$$P_{i+j,i}(\Delta t) = \mathbb{P}(Y_{t+\Delta t} = i + j \mid Y_t = i) = \begin{cases} \tilde{b}(i)\Delta t + O(\Delta t), & j = 1, \\ \tilde{d}(i)\Delta t + O(\Delta t), & j = -1, \\ 1 - (\tilde{b}(i) + \tilde{d}(i))\Delta t + O(\Delta t), & j = 0, \\ O(\Delta t), & j \neq -1, 0, 1. \end{cases}$$

After summing over these transition probabilities, dividing by Δt and taking limits we can obtain the forward Kolmogorov differential equation of this process; for $P_{j,i}(t) = \mathbb{P}(Y_t = j \mid Y_0 = i)$, we have

$$\frac{dP_{j,i}(t)}{dt} = \tilde{b}(j-1)P_{j-1,i}(t) - \left(\tilde{b}(j) + \tilde{d}(j)\right)P_{j,i}(t) + \tilde{d}(j+1)P_{j+1,i}(t),$$

$$\frac{dP_{0,i}(t)}{dt} = -\tilde{b}(0)P_{0,i}(t) + \tilde{d}(1)P_{1,i}(t),$$

for $i \geq 1$ and $j \geq 1$.

A *simple birth-death process* $(Y_t)_{t \geq 0}$ is a birth-death process where the rates are constant for all individuals. In other words, each individual, independently of all the others, produces an offspring at a constant \tilde{b} and dies at a constant rate \tilde{d} . In this case we have $\tilde{b}(i) = i\tilde{b}$ and $\tilde{d}(i) = i\tilde{d}$. Assuming that $Y_0 = 1$, the differential equation of the process

satisfies, for $P_i(t) = \mathbb{P}(Y_t = i \mid Y_0 = 1)$, is

$$\frac{dP_i(t)}{dt} = (i-1)\tilde{b}P_{i-1}(t) - i(\tilde{b} + \tilde{d})P_i(t) + (i+1)\tilde{d}P_{i+1}(t),$$

$$\frac{dP_0(t)}{dt} = \tilde{d}P_1(t).$$

We often refer to the event of an individual producing an offspring as a *split*, and we may say that \tilde{b} is the *birth rate* or *splitting rate* of the process.

A *pure birth process* or *birth process* is a birth-death process whose death rates are 0. A *simple pure birth process* is a simple birth-death process with $\tilde{d} = 0$. Notice that these definitions can be stated on a discrete-time setting where Y_n is defined for $n \in \{0, 1, 2, \dots\}$ and where the corresponding transition probabilities will satisfy recurrence relations instead of differential equations.

Simple birth-death processes may be extended by allowing the birth and death rate parameters to depend on time; \tilde{b}_t, \tilde{d}_t . These time-dependent extensions appear naturally when considering conditional versions of time-independent birth-death processes. We also refer to time-independent simple birth-death processes as *constant-rate* birth-death processes. Nee et al [39] consider four related processes;

1. A simple birth-death process $(Y_t)_{t \geq 0}$ with constant rates starting at time 0.
2. The same process conditioned on surviving to time t ; $(Y_t \mid Y_t \neq 0)$.
3. The same process conditioned on non-extinction at the present time T ; $(Y_t \mid Y_T \neq 0)$.
4. The *ancestral* or *reconstructed* process, resulting from pruning all progeny which does not survive to time T , conditioned on non-extinction at time T . The random tree representing this process is known as the *ancestral tree* or *reconstructed tree*.

The authors show that the processes 1, 2 and 4 have a geometric distribution when observed at a fixed time $t \in [0, T]$. The distribution of the third process is that of the sum

of two independent geometric distributions. The parameters of these geometric distributions are in terms of t and of the birth and death rates \tilde{b} and \tilde{d} . Furthermore, they cleverly pick a time-dependent parameter and show that the ancestral process (process 4) has the distribution of a time-dependent pure birth process with that parameter. We aim to generalize this particular result to the multi-type setting as a motivation for our research on some of the asymptotic properties of multi-type Yule trees (see Chapter 3).

1.2.2 Multi-type birth-death processes

A *multi-type birth-death process* $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{k,t})$ for $t \geq 0$ is a vector-valued Markov process whose entry $Y_{i,t}$ represents the number of individuals of type i in a population at time t . Parents produce offsprings at rates which depend only on the parent's and offspring's types. Denote by \tilde{b}_i^{ij} the rate at which individual of type i produces an offspring of type j , and denote by \tilde{d}_i the death rate of an individual of type i . Thus $\tilde{b}_i = \sum_{j=1}^k \tilde{b}_i^{ij}$ is the birth rate of a type- i individual. Thus, the infinitesimal transition probabilities satisfy:

$$P_{\mathbf{i}+\mathbf{j},\mathbf{i}}(\Delta t) = \mathbb{P}(\mathbf{Y}_{t+\Delta t} = \mathbf{i}+\mathbf{j} \mid \mathbf{Y}_t = \mathbf{i}) = \begin{cases} \sum_{\ell=1}^k i_\ell \tilde{b}_\ell^{\ell a} \Delta t + O(\Delta t), & \mathbf{j} = \mathbf{e}_a, \\ i_a \tilde{d}_a \Delta t + O(\Delta t), & \mathbf{j} = -\mathbf{e}_a, \\ 1 - \sum_{a=1}^k \left(\sum_{\ell=1}^k i_\ell \tilde{b}_\ell^{\ell a} + i_a \tilde{d}_a \right) \Delta t + O(\Delta t), & \mathbf{j} = \mathbf{0}, \\ O(\Delta t), & \text{else,} \end{cases}$$

for $a \in \{1, 2, \dots, k\}$. Hence, it may be shown that these probabilities satisfy the system of differential equations

$$\frac{dP_{\mathbf{j},\mathbf{i}}(t)}{dt} = \sum_{a=1}^k \sum_{\ell=1}^k i_\ell \tilde{b}_\ell^{\ell a} P_{\mathbf{j}-\mathbf{e}_a,\mathbf{i}}(t) + \sum_{a=1}^k i_a \tilde{d}_a P_{\mathbf{j}+\mathbf{e}_a,\mathbf{i}}(t) - \sum_{a=1}^k \left(\sum_{\ell=1}^k i_\ell \tilde{b}_\ell^{\ell a} + i_a \tilde{d}_a \right) P_{\mathbf{j},\mathbf{i}}(t).$$

As in the single-type case, these processes may also be extended to time-dependent rates. Furthermore, one may allow for *mutations*, so that an individual of type i may switch its type to j at time t at a rate $\tilde{b}_i^j(t)$. In this more general setting, the transition rates become

$$P_{i+j,i}(\Delta t) = \begin{cases} \sum_{j=1}^k i_j \tilde{b}_j^{\ell}(t) \Delta t + O(\Delta t), & \mathbf{j} = \mathbf{e}_\ell, \\ i_\ell \tilde{d}_\ell(t) \Delta t + O(\Delta t), & \mathbf{j} = -\mathbf{e}_\ell, \\ i_{\ell_1} \tilde{b}_{\ell_1}^{\ell_2}(t) \Delta t + O(\Delta t), & \mathbf{j} = \mathbf{e}_{\ell_2} - \mathbf{e}_{\ell_1}, \\ 1 - \sum_{\ell=1}^k \left(\sum_{a=1}^k i_j \tilde{b}_a^{\ell}(t) + i_\ell \tilde{d}_\ell(t) + i_\ell \tilde{b}_\ell^a(t) \right) \Delta t + O(\Delta t), & \mathbf{j} = \mathbf{0}, \\ O(\Delta t), & \text{else,} \end{cases}$$

for $\ell, \ell_1, \ell_2 \in \{1, 2, \dots, k\}$. Multi-type pure birth processes with mutations and time-dependent rates reappear in Section 3.2 when we consider the ancestral tree of a constant-rate multi-type birth-death process.

1.2.3 Yule model

The *Yule model* is a discrete-time branching process where at each step, an individual is chosen uniformly at random and splits into two individuals (or equivalently, it dies off and produces two offsprings). Assuming that this process starts with one individual at time 0, its n -th iteration may be represented as a rooted binary tree with $n + 1$ leaves (each step increments the number of leaves by one) and n ordered internal nodes, where the order is given by the splitting order and is thus consistent with the shape of the tree, that is, the first node is the root, and all parent nodes precede their offsprings.

A random tree obtained from the Yule model is called a *Yule tree*. Yule trees appear naturally from constant-rate birth-death processes as defined in Section 1.2.1, indeed:

Lemma 1.2 (Aldous [1], Lemma 1). *Let \mathcal{T} be a birth-death tree with constant rates started*

at time 0 with one individual and conditioned on having $n + 1 > 0$ living individuals at the present time T . Let $\mathcal{Y}(\mathcal{T})$ be the random rooted binary tree representing the topology of \mathcal{T} , with its n internal nodes ordered according to the time their corresponding splits occurred in \mathcal{T} . Then $\mathcal{Y}(\mathcal{T})$ is distributed as a random rooted binary tree under the Yule model.

Some statistics on Yule trees are of special importance to our research. A *cherry* is a pair of leaves with a common parent (only one edge away from each other), and a *pendant edge* is a leaf which does not belong to a cherry.

Equivalently the Yule model may be seen as a process to generate random rooted binary trees, where leaves are added successively by uniformly choosing a leaf, and splitting it to form a cherry. We may assume that the tree starts at generation 1 with one single cherry. A random tree of n leaves formed this way is said to follow the *Yule distribution*. McKenzie & Steel [34] investigated some properties of the number of cherries under the Yule model of tree generation:

Theorem 1.7 ([34]). *Under the Yule distribution, let μ_n denote the mean number of cherries of a random rooted binary tree on n leaves, and σ_n^2 the variance of the number of cherries. We have*

$$\mu_n = \frac{n}{3} \quad (n \geq 3); \quad \sigma_n^2 = \frac{2n}{45} \quad (n \geq 5).$$

Proposition 1.1 ([34]). *Let C_n be the number of cherries of a random rooted binary tree on n leaves. For the Yule model we have;*

$$\frac{C_n - n/3}{\sqrt{2n/45}} \rightarrow N(0, 1) \text{ in distribution.}$$

This Proposition is a consequence of regarding the splitting process from the Yule model as particular extended Pólya urn process, as described by Smythe [44] and Janson [20]. In the next section we give a brief introduction of these processes. In Section

3.1.2, we use Pólya urns to extend some of McKenzie & Steel [34]’s results to the neutral two-type Yule model.

1.3 Extended Pólya urn models

An *extended Pólya urn process* $(\mathbf{X}_n)_{n \geq 0}$ is a discrete process where the entry $X_{ni} \geq 0$ of the random vector $\mathbf{X}_n = (X_{n1}, \dots, X_{nl})$ represents the number of balls of type i (for $i = 1, \dots, l$) in an urn at time n . This urn starts with a given vector \mathbf{X}_0 that may or may not be random, and at each step, balls of different types are added or removed from it. Some parameters need to be introduced to understand this step: a positive weight $a_i \geq 0$ and a random l -dimensional vector $\vartheta_i = (\vartheta_{i1}, \dots, \vartheta_{il})$ with integer coordinates are assigned to each type $i \in \{1, \dots, l\}$. It is assumed that;

$$\vartheta_{ij} \geq 0, \text{ for } j \neq i, \quad \text{and} \quad \vartheta_{ii} \geq -1, \tag{1.3}$$

$$\text{with } \mathbb{E}(\vartheta_{ij}^2) < \infty. \tag{1.4}$$

At each step, a ball is randomly selected so that the probability of selecting a ball of type i is proportional to its weight a_i . Subsequently, if a ball of type i was selected, then ϑ_{ij} balls of type j are added to the urn for $j = 1, \dots, l$.

In particular, the condition $\vartheta_{ii} \geq -1$ means the selected ball is drawn and may or may not be replaced. The probability of drawing a ball of type i at time $n \geq 1$ is given by the fraction

$$\frac{a_i X_{(n-1)i}}{\sum_j a_j X_{(n-1)j}}.$$

In the work of Janson [20], it is assumed that the urn never becomes empty, that is

$$|\mathbf{X}_n| = \sum_{i=1}^l X_{n,i} > 0 \text{ for all } n. \quad (1.5)$$

The author uses the known fact that Pólya urn models are embedded in a multi-type continuous-time Markov branching process (introduced first by Athreya & Karlin [4]) to obtain asymptotic results associated with the covariances and means of the process entries. Next we introduce some notation and state some of the results of Janson, which we use in Section 3.1.2 to generalize Theorem 1.7 and Proposition 1.1 to the two-type setting. We do not state these results in their original form, but rather specialize them to a simpler form which still serves our applications.

A type i is said to *dominate* another type j , denoted $i \succ j$, if starting with one single ball of type i it is possible to eventually obtain at least one ball of type j . This type is *dominating* if it dominates every other type. The process is said to be *irreducible* if all types are dominating. In a process started with a single ball, we may ignore all types that never occur and ensure irreducibility by requiring that all types that may appear dominate each other.

Let $\mathbf{A} := (a_j \mathbb{E}(\vartheta_{ji}))_{i,j=1}^l$ be the *generating matrix* of the Pólya urn model as defined in Janson [20]. This matrix is said to be irreducible if the process is irreducible. The eigenvalues of this matrix are ordered according to decreasing real parts: $\lambda_1 > \text{Re } \lambda_2 \geq \text{Re } \lambda_3 \geq \dots$. In fact, it is known by the standard Perron-Frobenius theory that the largest eigenvalue λ_1 is real and strictly greater than the real part of any of the other eigenvalues.

For our specialized version of Janson's theorems, we work under the following assumptions, which together imply assumptions A1-A6 from Janson [20].

(α_1) Conditions 1.3, 1.4 and 1.5 hold.

(α_2) The largest real eigenvalue λ_1 of \mathbf{A} is positive.

(α_3) The model starts with one single ball, and all types that may appear afterwards dominate each other.

(α_4) The eigenvalues λ_1 and λ_2 are simple, with left and right eigenvectors $\mathbf{u}_1, \mathbf{v}_1$ and $\mathbf{u}_2, \mathbf{v}_2$ respectively, satisfying

$$\mathbf{u}_1 \cdot \mathbf{v}_1 = \mathbf{u}_2 \cdot \mathbf{v}_2 = 1,$$

$$\mathbf{a} \cdot \mathbf{v}_1 = 1,$$

with $\mathbf{a} = (a_1, \dots, a_l)$. The following additional assumption allows to simplify Janson's results further:

(α_5) $\operatorname{Re} \lambda_2 > \operatorname{Re} \lambda_3$.

In particular this condition implies that the set Λ_{II} of eigenvalues λ satisfying $\operatorname{Re} \lambda = \lambda_1/2$ is either the empty set or the set $\{\lambda_2\}$, which simplifies some of Janson's formulae.

Theorem 1.8 ([20]. Theorem 3.21). *Under conditions (α_1)-(α_5) above, we have*

$$\frac{\mathbf{X}_n}{n} \rightarrow \lambda_1 \mathbf{v}_1 \text{ a.s. as } n \rightarrow \infty.$$

Theorem 1.9 ([20]. Theorems 3.22 and 3.23). *Assume conditions (α_1)-(α_5) above.*

1. *Suppose $\operatorname{Re} \lambda_2 < \lambda_1/2$. We have as $n \rightarrow \infty$*

$$\frac{\mathbf{X}_n - n\lambda_1 \mathbf{v}_1}{\sqrt{n}} \rightarrow N(0, \Sigma) \text{ in distribution,}$$

where

$$\boldsymbol{\Sigma} := \int_0^\infty \boldsymbol{\psi}(s, \mathbf{A}) \mathbf{B} \boldsymbol{\psi}(s, \mathbf{A})' e^{-\lambda_1 s} \lambda_1 ds - \lambda_1^2 \mathbf{v}_1 \mathbf{v}_1^\top,$$

$$\mathbf{B} := \sum_{i=1}^l \mathbf{v}_{1i} a_i \mathbb{E}(\boldsymbol{\vartheta}_i \boldsymbol{\vartheta}_i^\top),$$

$$\boldsymbol{\psi}(s, \mathbf{A}) := e^{s\mathbf{A}} - \lambda_1 \mathbf{v}_1 \mathbf{a}^\top \phi(s, \mathbf{A}),$$

$$\phi(s, \mathbf{A}) := \int_0^s e^{t\mathbf{A}} dt.$$

2. Suppose $\operatorname{Re} \lambda_2 = \lambda_1/2$. We have as $n \rightarrow \infty$

$$\frac{\mathbf{X}_n - n\lambda_1 \mathbf{v}_1}{n \ln(n)} \rightarrow N(0, \boldsymbol{\Sigma}) \text{ in distribution,}$$

where the covariance matrix is given by;

$$\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{T}) \boldsymbol{\Sigma}_{II} (\mathbf{I} - \mathbf{T}^\top),$$

with $\mathbf{T} := \lambda_2^{-1} \lambda_1 \mathbf{v}_1 \mathbf{a}^\top \mathbf{v}_2 \mathbf{u}_2^\top$, and

$$\boldsymbol{\Sigma}_{II} := \mathbf{v}_2 \mathbf{u}_2^\top \mathbf{B} (\mathbf{v}_2 \mathbf{u}_2^\top)^*.$$

Extended Pólya urn models are useful when obtaining asymptotic information about a specific discrete Markov process. We use them in Section 3.1.3 to derive some asymptotic results on the distribution of two-type cherries in neutral two-type binary trees.

Chapter 2

The coalescent point process of multi-type branching trees

In this chapter we extend the coalescent point process construction of Section 1.1.3 (see [29]) to the case of multi-type branching processes. Our goal is to exploit the Markovian features of the coalescent point process in order to derive features of multi-type phylogenetic trees, and identify the statistics in multi-type phylogenetic trees that are not present in single type trees. We first derive the distribution of the most recent common ancestor of two species from the standing population. We then derive the time of the most recent common ancestor of two species of the same type, and its dependence on the species type. Finally we apply these results to the special case of multi-type linear fractional branching processes (Section 1.1.1).

2.1 Multi-type coalescent point processes

Our construction of the coalescent point process for a multi-type Galton-Watson branching tree is a natural generalization of the single-type coalescent point process from the previous Section 1.1.3. For the sake of completeness, we describe its construction, with-

out relying on previous concepts. Consider a multi-type Galton-Watson process, assumed to be quasi-stationary. The distribution of the number and types of offsprings of each individual depends on its type, and for an individual of type ℓ will be denoted by $\mathbb{P}(\mathbf{Z}^{(1)} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell)$. We consider a doubly-infinite embedding of its genealogical tree as in Figure 2.2 (left), infinite in the number of standing individuals as well as in the number of past generations. Embedding of the genealogical tree in a two dimensional grid is made in such a way that empty spaces and intersections between lineages are avoided. We also specify a convention for the *order of embedding* an individual's offspring. We assume that the order in which they are embedded is chosen *uniformly* at random from all possible ways to order them. In a later section we will assume a more *specific ordering* in the case where the offspring distribution is linear-fractional.

Unlike the single-type case, a Markov process from which the multi-type coalescent point process can be reconstructed will have to contain information on the individuals' types as well. This, unfortunately, also makes notation for the multi-type process lengthier. Throughout this work we reserve boldface symbols for vectors and matrices.

Each individual in the genealogical tree is defined by its location coordinates, where (n, i) identifies the i -th individual from the left at generation $-n$. Let

$$t(n, i) := \text{type of the individual } (n, i).$$

Let

$$a_i(n) := \text{index (from left to right) of the ancestor of individual } (0, i) \text{ at generation } -n.$$

Recall that coalescence times between individuals $(0, i)$ and $(0, i + 1)$ at generation 0 are defined as $A_i := \min\{n \geq 1 : a_i(n) = a_{i+1}(n)\}$ for $i \geq 1$, and by convention $A_0 = +\infty$. Furthermore, the type type-enriched *ancestral lineage* of individual $(0, i + 1)$

back to its most recent common ancestor with individual $(0, i)$ is, for $i \geq 1$, denoted by $\mathbf{A}_i \in \{1, 2, \dots\}^{\mathbb{N}_0}$, and including a special 0-th entry is defined as:

$$\mathbf{A}_i := (t(0, a_{i+1}(0)), t(-1, a_{i+1}(1)), \dots, t(-A_i + 1, a_{i+1}(A_i - 1)), t(-A_i, a_{i+1}(A_i))).$$

The type-enriched ancestral lineages can be seen in Figure 2.2 (right) the type-enriched ancestral lineage of individual $(0, 2)$ until depth $A_1 = 1$ has two types: $\mathbf{A}_1 = (2, 1)$; of individual $(0, 3)$ until depth $A_2 = 1$ has two types: $\mathbf{A}_2 = (2, 1)$; of individual $(0, 4)$ until depth $A_3 = 2$ has three types: $\mathbf{A}_3 = (1, 2, 1)$; of individual $(0, 5)$ until depth $A_4 = 1$ has two types: $\mathbf{A}_4 = (2, 2)$, etc.

For a vector \mathbf{v} let $\mathbf{v}_{[j]}$ denote its j -th entry and $\|\mathbf{v}\|$ denote its number of entries, with the convention that $\|\mathbf{v}\| = 0$ if $\mathbf{v} = \emptyset$. Note that $A_i = \|\mathbf{A}_i\| - 1$. Since $a_{i+1}(0) = i + 1$, the 0-th entry $\mathbf{A}_{i[0]}$ of the vector \mathbf{A}_i is the type of the individual $(0, i + 1)$. Also $A_0 = \infty$ and the first ancestral lineage \mathbf{A}_0 consists of types of all individuals on the left most infinite (back into the past) spine of the ancestral tree. This special left most lineage in the example in Figure 2.2 (right) has the sequence of types $\mathbf{A}_0 = (2, 1, 1, 2, 2, 1, 2, \dots)$.

Definition 2.1. We call the process $(\mathbf{A}_i)_{i \geq 1}$ the multi-type coalescent point process.

We define the process $(\mathbf{D}_i)_{i \geq 1}$ of vector valued sequences $\mathbf{D}_i = \{\mathbf{D}_i(n), n \geq 1\}$ in such a way that each $\mathbf{D}_i(n) \in \{1, 2, \dots\}^{\mathbb{N}}$ is a vector of types of offspring of the ancestor $a_i(n)$ at generation $-n$ embedded to the right of the lineage of $(0, i)$ that are ‘survivors’ (meaning that they have progeny that are alive at generation 0):

$\mathbf{D}_i(n) :=$ vector of *types of surviving* offspring of individual $(-n, a_i(n))$ embedded in the *ancestral tree* to the right of and *including* the lineage of $(0, i)$.

Clearly $\|\mathbf{D}_i(n)\| \geq 1$, and note that $D_i(n) := \|\mathbf{D}_i(n)\| - 1$ is the number of surviving offspring of individual $(-n, a_i(n))$ embedded to the right of (and excluding) the lineage

of $(0, i)$, as in the single type process of Section 1.1.3. The values of this process for the tree in Figure 2.2 are given in Figure 2.3, the ancestor of individual $(0, 1)$ at generation -1 has three surviving offspring all of type 2: $\mathbf{D}_1(1) = (2, 2, 2)$, the ancestor of individual $(0, 1)$ at generation -2 has only two offspring with surviving progeny of types 1 and 2: $\mathbf{D}_1(2) = (1, 2)$, the ancestor of individual $(0, 1)$ at generation -3 has only one offspring with surviving progeny of type 1: $\mathbf{D}_1(3) = (1)$, etc. Note that the labelling of the sequences is such that \mathbf{A}_{i-1} and \mathbf{D}_i are sequences that describe the ancestral lineage of individual $(0, i)$. In particular, for each $i \geq 1$ the value of A_i is equal to the first depth n which $D_i(n) = \|\mathbf{D}_i(n)\| - 1 \neq 0$, and the n -th entry of \mathbf{A}_{i-1} (which we denote $\mathbf{A}_{(i-1)[n-1]}$ since \mathbf{A}_{i-1} has a 0-th entry) is equal to the first entry of $\mathbf{D}_i(n)$ for $1 \leq n \leq A_{i-1} + 1$. See Figure 2.1 for an illustration of these values on a sample two-type tree.

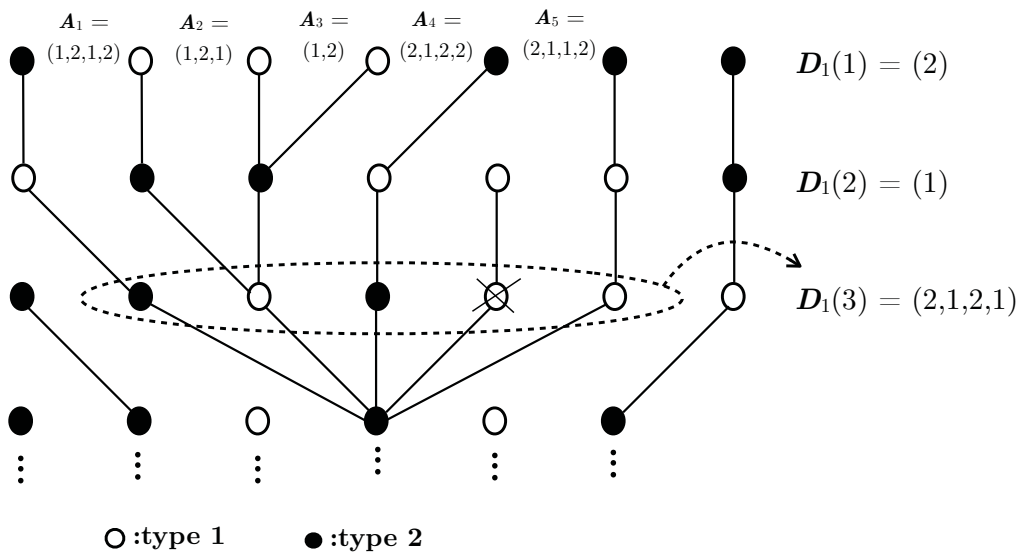


Figure 2.1: Some values of \mathbf{A}_i and $\mathbf{D}_i(n)$ illustrated on a planar embedding of a two-type branching process.

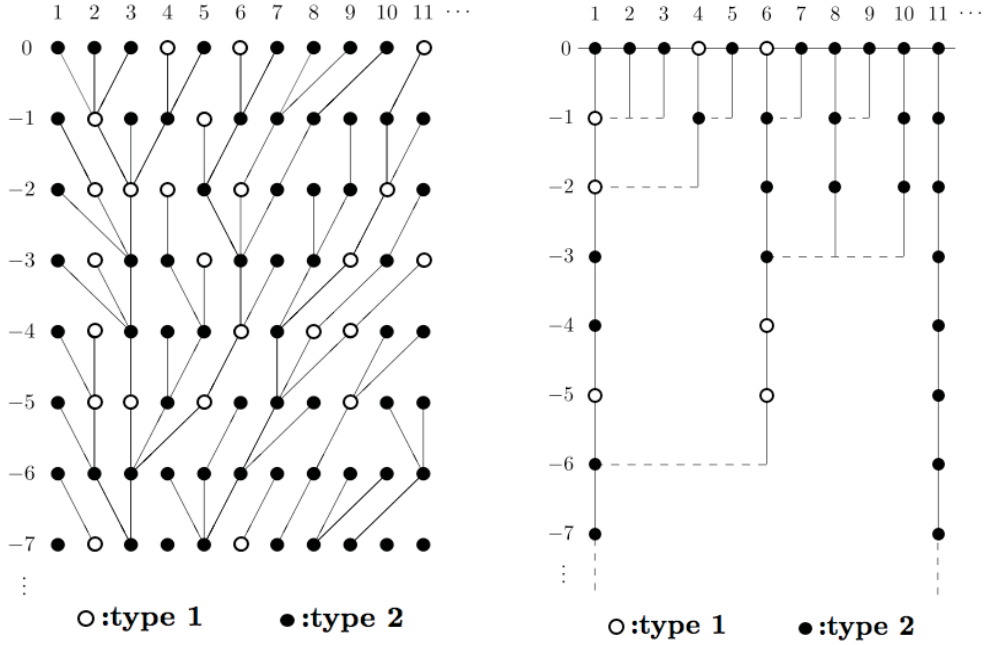


Figure 2.2: Planar embedding of a two-type branching process (left), and types along the ancestral lineages \mathbf{A}_i (right): the special left most lineage here has types $\mathbf{A}_0 = (2, 1, 1, 2, 2, 1, 2, 2, \dots)$, the coalescent times here are $A_1 = 1, A_2 = 1, A_3 = 2, \dots$ with types along ancestral lineages $\mathbf{A}_1 = (2, 1), \mathbf{A}_2 = (2, 1), \mathbf{A}_3 = (1, 2, 1), \mathbf{A}_4 = (2, 2), \dots$

In order to describe the law of \mathbf{D} we need to provide notation for surviving lineages. Let $\boldsymbol{\xi}_\ell$ be the offspring distribution of an individual of type ℓ with probability generating function $f_\ell(\mathbf{s})$. Denote by $\mathbf{f}(\mathbf{s})$ the vector with entries $f_\ell(\mathbf{s})$ for $\ell = 1, 2, \dots$. For individual (n, i) the law of the number of its offspring, given that its type is $t(n, i) = \ell$, is that of $\boldsymbol{\xi}_\ell$. The survival probability of the progeny in generation 0 of an offspring of some type ℓ' individual in some generation $-n'$ is given by $p_{n'-1, \ell'} := 1 - f_{\ell'}^{(n'-1)}(0, 0, \dots)$ where $f_{\ell'}^{(n'-1)}$ is the ℓ' -th entry of the $(n' - 1)$ -fold composition of \mathbf{f} . We consider all the survivor progeny of a generation $-n$ ancestor of some individual from the standing population, and suppose that the type of this generation $-n$ ancestor is ℓ . For different

$\ell' \in \{1, 2, \dots\}$, let $\{\epsilon_{n,\ell'}^1, \epsilon_{n,\ell'}^2, \dots\}$ be independent sequences of i.i.d. Bernoulli variables with parameters $\mathbb{P}(\epsilon_{n,\ell'}^m = 1) = p_{n-1,\ell'}$. Start with an independent variable ξ_ℓ , which takes values in \mathbb{N}_0^∞ and has $\xi_{\ell,\ell'}$ offspring of type ℓ' , and define the vector of random sums:

$$\zeta_{n,\ell} := \left(\sum_{m=1}^{\xi_{\ell,1}} \epsilon_{n,1}^m, \sum_{m=1}^{\xi_{\ell,2}} \epsilon_{n,2}^m, \dots \right),$$

whose ℓ' coordinate is denoted by $\zeta_{n,\ell,\ell'}$.

$A_0 = \infty$	$A_1 = \mathbf{1}$	$A_2 = \mathbf{1}$	$A_3 = \mathbf{2} \dots$
$\mathbf{A}_0 = (\mathbf{2}, \mathbf{1}, \mathbf{1}, \dots)$	$\mathbf{A}_1 = (\mathbf{2}, \mathbf{1})$	$\mathbf{A}_2 = (\mathbf{2}, \mathbf{1})$	$\mathbf{A}_3 = (\mathbf{1}, \mathbf{2}, \mathbf{1}) \dots$
$D_1(\mathbf{1}) = (\mathbf{2}, \mathbf{2}, \mathbf{2})$	$D_2(\mathbf{1}) = (\mathbf{2}, \mathbf{2})$	$D_3(\mathbf{1}) = (\mathbf{2})$	$D_4(\mathbf{1}) = (\mathbf{1}, \mathbf{2})$
$D_1(\mathbf{2}) = (\mathbf{1}, \mathbf{2})$	$D_2(\mathbf{2}) = (\mathbf{1}, \mathbf{2})$	$D_3(\mathbf{2}) = (\mathbf{1}, \mathbf{2})$	$D_4(\mathbf{2}) = (\mathbf{2})$
$D_1(\mathbf{3}) = (\mathbf{1})$	$D_2(\mathbf{3}) = (\mathbf{1})$	$D_3(\mathbf{3}) = (\mathbf{1})$	$D_4(\mathbf{3}) = (\mathbf{1})$
$D_1(\mathbf{4}) = (\mathbf{2})$	$D_2(\mathbf{4}) = (\mathbf{2})$	$D_3(\mathbf{4}) = (\mathbf{2})$	$D_4(\mathbf{4}) = (\mathbf{2})$
$D_1(\mathbf{5}) = (\mathbf{2})$	$D_2(\mathbf{5}) = (\mathbf{2})$	$D_3(\mathbf{5}) = (\mathbf{2})$	$D_4(\mathbf{5}) = (\mathbf{2})$
$D_1(\mathbf{6}) = (\mathbf{1}, \mathbf{1})$	$D_2(\mathbf{6}) = (\mathbf{1}, \mathbf{1})$	$D_3(\mathbf{6}) = (\mathbf{1}, \mathbf{1})$	$D_4(\mathbf{6}) = (\mathbf{1}, \mathbf{1})$
$D_1(\mathbf{7}) = (\mathbf{2})$	$D_2(\mathbf{7}) = (\mathbf{2})$	$D_3(\mathbf{7}) = (\mathbf{2})$	$D_4(\mathbf{7}) = (\mathbf{2})$
\vdots	\vdots	\vdots	\vdots

Figure 2.3: Sequences $(\mathbf{A}_i, \mathbf{D}_i(\cdot))_{i \geq 1}$ of surviving offspring types along the lineages of individuals $((0, i))_{i \geq 1}$ corresponding to the two-type ancestral tree given in Figure 2.2: note that for each $i \geq 1$ and $1 \leq n \leq A_{i-1} + 1$ we have $\mathbf{A}_{i-1}(n-1) = \mathbf{D}_i(n)_{[1]}$ (including $i = 1$ when $A_0 = +\infty$).

Then, the law of $\zeta'_{n,\ell}$, which represents the number of surviving offspring of different types at generation $-(n-1)$ of the initiating generation $-n$ type ℓ ancestor conditioned

on having at least one surviving offspring, is given by:

$$\zeta'_{n,\ell} := \stackrel{d}{=} (\zeta_{n,\ell} \mid \sum_{\ell'=1}^{\infty} \zeta_{n,\ell,\ell'} \neq 0).$$

Let $\mathbf{d}(\zeta'_{n,\ell}) \in \{1, 2, \dots\}^{\mathbb{N}}$ be an ordering of all the offspring counted by $\zeta'_{n,\ell}$ chosen uniformly at random from all possible orderings (or in some specific way, as in the next section).

Recall that $\mathbf{v}_{[i]}$ denotes the i -th coordinate of a vector $\mathbf{v} \in \{1, 2, \dots\}^{\mathbb{N}_0}$. Then $j := \mathbf{d}(\zeta'_{n,\ell})_{[1]}$ is the type of the left most surviving offspring at generation $-(n-1)$ of the type ℓ ancestor from generation $-n$. Again, for different $\ell' \in \{1, 2, \dots\}$, let $\{\epsilon_{n-1,\ell'}^1, \epsilon_{n-1,\ell'}^2, \dots\}$ be independent sequences of i.i.d. Bernoulli variables with parameters $\mathbb{P}(\epsilon_{n-1,\ell'}^m = 1) = p_{n-2,\ell'}$ (independent of all earlier sequences of Bernoulli variables). Proceed with an independent variable ξ_j , and define the vector of random sums:

$$\zeta_{n-1,j} := \left(\sum_{m=1}^{\xi_{j,1}} \epsilon_{n-1,1}^m, \sum_{m=1}^{\xi_{j,2}} \epsilon_{n-1,2}^m, \dots \right), \quad \zeta'_{n-1,j} := \stackrel{d}{=} (\zeta_{n-1,j} \mid \sum_{\ell'=1}^{\infty} \zeta_{n-1,j,\ell'} \neq 0),$$

and let $\mathbf{d}(\zeta'_{n-1,j}) \in \{1, 2, \dots\}^{\mathbb{N}}$ be the ordering of these surviving offspring. Then $\kappa := \mathbf{d}(\zeta'_{n-1,j})_{[1]}$ is the type of the left most surviving progeny at generation $-(n-2)$ of the initiating individual from generation $-n$. Following the left most surviving progeny of an individual of type 1 can be seen in Figure 2.4, where the ancestor at generation -5 of the individual $(0, 6)$ is followed.

We proceed in this way recursively until generation -1 when we obtain the set of offspring $\mathbf{d}(\zeta'_{1,i})$. In order to collect all types of the left most surviving progeny (and their siblings) in different generations $0, -1, \dots, -(n-2)$, and $-(n-1)$ in one vector, we define an n long sequence of vectors in type space initiated by individual of type ℓ in generation $-n$:

$$\eta_{n,\ell} := (\mathbf{d}(\zeta'_{1,i}), \dots, \mathbf{d}(\zeta'_{n-2,\kappa}), \mathbf{d}(\zeta'_{n-1,j}), \mathbf{d}(\zeta'_{n,\ell})),$$

whose coordinates are then the vectors of surviving offspring types in different generations
 $\eta_{n,\ell}(1) := \mathbf{d}(\zeta'_{1,i}), \dots, \eta_{n,\ell}(n-1) := \mathbf{d}(\zeta'_{n-1,j}), \eta_{n,\ell}(n) := \mathbf{d}(\zeta'_{n,\ell})$.

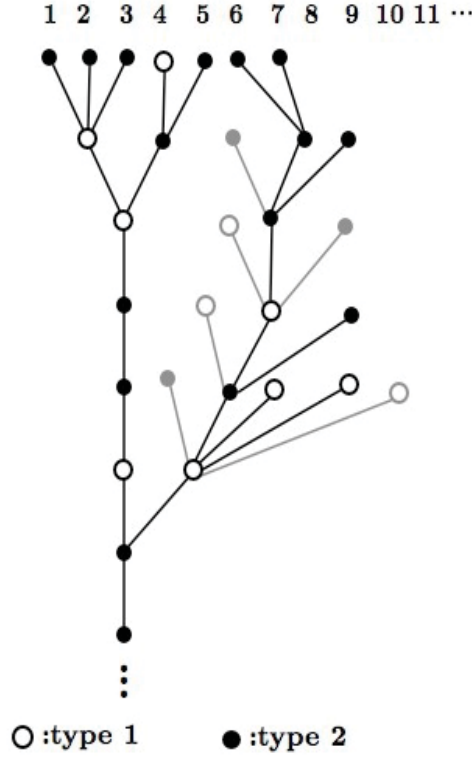


Figure 2.4: Sequences $\eta_{A_i-1, \mathcal{D}_i(A_i)_{[2]}}(n)$ of surviving subtrees $\mathcal{D}_{i+1}(n)$ ($1 \leq n \leq A_i - 1$) are illustrated for the case $i = 5$ where $A_5 = 6$: the collection of offspring on ancestral lineage of individual $(0, 6)$ originating from its common ancestor at generation -6 with individual $(0, 5)$ are marked with either dark edges (surviving to generation 0) or light edges (non-surviving). Specifically: $\eta_{5,1}(5) = (2, 1, 1)$, $\eta_{5,2}(4) = (1, 2)$, $\eta_{5,1}(3) = (2)$, $\eta_{5,2}(2) = (2, 2)$, $\eta_{5,2}(1) = (2, 2)$.

Having defined the random variables $\eta_{n,\ell}$, for arbitrary n and ℓ , the reconstruction of the ancestral tree from the auxiliary process is possible as in the single type case.

Theorem 2.1. *The coalescent times $(A_i)_{i \geq 1}$ and types along the ancestral tree $(\mathbf{A}_i)_{i \geq 1}$ are both functionals of $(\mathbf{D}_i)_{i \geq 1}$ given by*

$$A_i = \min\{n \geq 1 : \|\mathbf{D}_i(n)\| - 1 \neq 0\}, \quad \mathbf{A}_i = (\mathbf{D}_{i+1}(1)_{[1]}, \dots, \mathbf{D}_{i+1}(A_i)_{[1]}, \mathbf{D}_{i+1}(A_i+1)_{[1]}).$$

The sequence $(\mathbf{D}_i)_{i \geq 1}$ is a Markov chain with transition probabilities given by:

$$(\mathbf{D}_{i+1}(n) \mid \mathbf{D}_i) \stackrel{d}{=} \begin{cases} \mathbf{D}_i(n) & \text{for } n > A_i, \\ (\mathbf{D}_i(n)_{[2]}, \mathbf{D}_i(n)_{[3]}, \dots) & \text{for } n = A_i, \\ \boldsymbol{\eta}_{A_i-1, \mathbf{D}_i(A_i)_{[2]}}(n) & \text{for } 1 \leq n < A_i, \end{cases} \quad (2.1)$$

where the law of the $A_i - 1$ long sequence of type vectors $(\mathbf{D}_{i+1}(1), \dots, \mathbf{D}_{i+1}(A_i - 1))$ is distributed as the vector $\boldsymbol{\eta}_{A_i-1, \mathbf{D}_i(A_i)_{[2]}}$ of types of the left most surviving progeny (and their siblings) at generations $-1, \dots, -(A_i - 1)$ of a generation $-n = -(A_i - 1)$ individual whose type is $\ell = \mathbf{D}_i(A_i)_{[2]}$.

Before presenting the proof of Theorem 2.1, we first state a spine decomposition of a multi-type branching process conditioned on survival to a certain generation, which shows that, if we consider the infinite (back into the past) lineage of a current individual, at every generation back in the past the subtrees of siblings of the ancestor in that generation are independent of the infinite lineage and are distributed as trees of an unconditioned multi-type branching process. Moreover, *knowing the values of their own initial individuals*, these trees are independent from their sibling subtrees, and are independent of their rank in the planar ordering.

For single-type processes this result first appeared in [31] and [14] (Section 1.1.2). For multi-type processes a decomposition of a tree relative to a spine that is infinite *into the future* is stated in [26], and in [15] for branching in continuous time. We present a statement in the form of Lemma 1.1 ([14] see Section 1.1.2) for decomposition of trees

conditioned only to survive to a fixed generation, and give its proof.

Consider a multi-type branching process $\mathbf{Z} = (\mathbf{Z}_1^{(n)}, \mathbf{Z}_2^{(n)}, \dots)_{n \geq 0}$ which is still non-extinct at generation $n + 1$. Let $T^{(i)}, 1 \leq i \leq |\mathbf{Z}^{(1)}|$ denote the subtrees descending from the offspring in the first generation, under a uniform ordering $\mathbf{d}(\mathbf{Z}^{(1)})$ of all the offspring in the first generation, and let R_{n+1} be the rank of the first offspring whose descendants survive to generation $n + 1$.

Lemma 2.1. *The subtrees $T^{(i)}, 1 \leq i \leq |\mathbf{Z}^{(1)}|$, $d_i \in \{1, 2, \dots\}$, are conditionally independent given $\{\mathbf{Z}^{(0)} = \mathbf{e}_\ell, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j\}$, for $1 \leq j \leq |\mathbf{z}|$ and $\mathbf{z} = (z_1, z_2, \dots)$ with $\mathbf{d} = (d_1, \dots, d_{|\mathbf{z}|})$. Furthermore*

$$(T^{(i)} \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j) \stackrel{d}{=} \begin{cases} (\mathcal{T} \mid \mathbf{Z}^{(n)}(\mathcal{T}) = \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}), & 1 \leq i \leq j - 1, \\ (\mathcal{T} \mid \mathbf{Z}^{(n)}(\mathcal{T}) \neq \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}), & i = j, \\ (\mathcal{T} \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}), & j + 1 \leq i \leq |\mathbf{z}|, \end{cases}$$

where \mathcal{T} denotes the law of a tree of multi-type branching processes with the probability generating function of \mathbf{Z} . Further, the conditional joint distribution of $R_{n+1}, \mathbf{Z}^{(1)}, \mathbf{d}(\mathbf{Z}^{(1)})$ is given by

$$\begin{aligned} & \mathbb{P}(R_{n+1} = j, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}(\mathbf{z}) \mid \mathbf{Z}^{(n+1)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \\ &= \frac{\mathbb{P}(\boldsymbol{\xi}_\ell = \mathbf{z}) \mathbb{P}(\mathbf{d}(\mathbf{z}) = (d_1, \dots, d_{|\mathbf{z}|})) \mathbb{P}(\mathbf{Z}^{(n)} \neq \mathbf{0} \mid \mathbf{Z}^{(0)} = \mathbf{e}_{d_j}) \prod_{j'=1}^{j-1} \mathbb{P}(\mathbf{Z}^{(n)} = \mathbf{0} \mid \mathbf{Z}^{(0)} = \mathbf{e}_{d_{j'}})}{\mathbb{P}(\mathbf{Z}^{(n+1)} \neq \mathbf{0} \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell)}. \end{aligned} \tag{2.2}$$

Proof of Lemma 2.1. Let \mathcal{T} denote the tree of a branching process with the p.g.f of \mathbf{Z} .

Let \mathbf{z} be a sample value of $\mathbf{Z}^{(1)}$, and consider $j \in \{1, \dots, |\mathbf{z}|\}$ fixed. Let $(\mathcal{A}_i)_{1 \leq i \leq |\mathbf{z}|}$ be measurable subsets of the space of multi-type rooted planar trees with roots given by d_i , where $\mathbf{d}(\mathbf{z}) = (d_1, \dots, d_{|\mathbf{z}|})$. Assume that for $1 \leq i \leq j-1$, we have

$$\mathcal{A}_i \subseteq \{\mathcal{T} : \mathbf{Z}^{(n)}(\mathcal{T}) = \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}\}, \quad \mathcal{A}_j \subseteq \{\mathcal{T} : \mathbf{Z}^{(n)}(\mathcal{T}) \neq \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_j}\},$$

and assume no additional condition on \mathcal{A}_i for $i > j$. Since $\{T^{(i)} \in \mathcal{A}_i\}_{1 \leq i \leq j-1}$ implies that the first $j-1$ subtrees are extinct by generation n , and $T^{(j)} \in \mathcal{A}_j$, which implies that the j -th subtree $T^{(j)}$ has survived to generation n , together imply that $R_{n+1} = j$, we have

$$\left\{ \bigcap_{i=1}^{|\mathbf{z}|} \{T^{(i)} \in \mathcal{A}_i\}, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d} \right\} \subset \{R_{n+1} = j\}.$$

So from the independence of offspring trees without condition on their shape:

$$\begin{aligned} & \mathbb{P}(\{T^{(i)} \in \mathcal{A}_i\}_{1 \leq i \leq |\mathbf{z}|}, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \\ &= \mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z} \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \mathbb{P}(\mathbf{d}(\mathbf{z}) = \mathbf{d}) \prod_{i=1}^{|\mathbf{z}|} \mathbb{P}(T^{(i)} \in \mathcal{A}_i \mid \mathbf{Z}^{(0)}(T^{(i)}) = \mathbf{e}_{d_i}). \end{aligned} \quad (2.3)$$

From this equality we have that the subtrees $T^{(i)}$ are conditionally independent given $\{\mathbf{Z}^{(0)} = \mathbf{e}_\ell, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}(\mathbf{z}), R_{n+1} = j\}$, since the measurable sets \mathcal{A}_i are arbitrary under those conditions.

To prove the equality in distribution, we use the shorthand notation

$$E_i := \{\mathcal{T} : \mathbf{Z}^{(n)}(\mathcal{T}) = \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}\}, \text{ for } 1 \leq i \leq j-1,$$

$$E_j := \{\mathcal{T} : \mathbf{Z}^{(n)}(\mathcal{T}) \neq \mathbf{0}, \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_j}\},$$

$$E_i := \{\mathcal{T} : \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}\}, \text{ for } j < i \leq |\mathbf{z}|.$$

We need to show that

$$(T^{(i)} \mid \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell, R_{n+1} = j) \stackrel{d}{=} (\mathcal{T} \mid \mathcal{T} \in E_i).$$

Equivalently, that for every measurable subset $\hat{\mathcal{A}}_i \subset E_i$, we have

$$\begin{aligned} & \mathbb{P}(T^{(i)} \in \hat{\mathcal{A}}_i \mid \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell, R_{n+1} = j) \\ &= \mathbb{P}(\mathcal{T} \in \hat{\mathcal{A}}_i \mid \mathcal{T} \in E_i). \end{aligned} \quad (2.4)$$

The left hand side of the above equality can be rewritten as

$$\frac{\mathbb{P}(T^{(i)} \in \hat{\mathcal{A}}_i, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell)}{\mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell)}. \quad (2.5)$$

Using equation (2.3), the numerator of this expression becomes

$$\mathbb{P}(T^{(i)} \in \hat{\mathcal{A}}_i, \{T^{(r)} \in E_r\}_{r \neq i}, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) =$$

$$\mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z} \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \mathbb{P}(\mathbf{d}(\mathbf{z}) = \mathbf{d}) \mathbb{P}(\mathcal{T} \in \hat{\mathcal{A}}_i \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i}) \prod_{r \neq i} \mathbb{P}(\mathcal{T} \in E_r \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_r}),$$

while the denominator is equal to

$$\begin{aligned} & \mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \\ &= \mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, \{T^{(r)} \in E_r\}_{r=1, \dots, |\mathbf{z}|}, R_{n+1} = j \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \\ &= \mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z} \mid \mathbf{Z}^{(0)} = \mathbf{e}_\ell) \mathbb{P}(\mathbf{d}(\mathbf{z}) = \mathbf{d}) \prod_{r=1}^{|\mathbf{z}|} \mathbb{P}(\mathcal{T} \in E_r \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_r}). \end{aligned}$$

Together the last two equalities show that (2.5) is equal to

$$\mathbb{P}(T^{(i)} \in \hat{\mathcal{A}}_i \mid R_{n+1} = j, \mathbf{Z}^{(1)} = \mathbf{z}, \mathbf{d}(\mathbf{Z}^{(1)}) = \mathbf{d}, \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \frac{\mathbb{P}(\mathcal{T} \in \hat{\mathcal{A}}_i \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i})}{\mathbb{P}(\mathcal{T} \in E_i \mid \mathbf{Z}^{(0)}(\mathcal{T}) = \mathbf{e}_{d_i})},$$

which is equal to the right hand side of equation (2.4). This proves the statement about the law of the different subtrees. To prove the statement about the joint law of the index of the surviving subtree to generation n and the number and types of the offspring at generation 1, it is sufficient to condition on $\mathbf{Z}^{(1)} = \mathbf{z}$ and use independence. \star

Proof of Theorem 2.1. Observe that

$$\begin{aligned} A_i > n &\Leftrightarrow \forall n' \leq n, a_i(n') \neq a_{i+1}(n') \\ &\Leftrightarrow \forall n' \leq n, (-n', a_i(n')) \text{ has no surviving progeny in } \{(0, i+1), (0, i+2), \dots\} \\ &\Leftrightarrow \forall n' \leq n, \|\mathbf{D}_i(n')\| = 1, \end{aligned}$$

so that A_i is the level of the first term of the sequence \mathbf{D}_i such that $D_i(n) = \|\mathbf{D}_i(n)\| - 1 \neq 0$.

In addition, notice that $\mathbf{D}_{i+1}(n)$ records the types of the offsprings of the ancestor of $(0, i+1)$ at generation $-n$ which have surviving progeny embedded to the right of $(0, i+1)$, including the one that is on the lineage of $(0, i+1)$. Thus $\mathbf{D}_{i+1}(n)_{[1]}$ is the ancestor's type of $(0, i+1)$ at generation $-(n-1)$, which proves the second equality.

At level A_i we have the most recent common ancestor $a_i(A_i) = a_{i+1}(A_i)$ of individuals $(0, i)$ and $(0, i+1)$, whose offspring with surviving progeny embedded to the right of $(0, i+1)$ do not include the ancestor of $(0, i)$, which is recorded in $\mathbf{D}_i(A_i)_{[1]}$, but do include all the others. So,

$$\left(\mathbf{D}_{i+1}(A_i)_{[1]}, \mathbf{D}_{i+1}(A_i)_{[2]}, \mathbf{D}_{i+1}(A_i)_{[3]}, \dots\right) = \left(\mathbf{D}_i(A_i)_{[2]}, \mathbf{D}_i(A_i)_{[3]}, \dots\right).$$

At any level $n > A_i$ below the most recent common the ancestors of $(0, i)$ and $(0, i+1)$

are the same since $a_i(A_i) = a_{i+1}(A_i)$ implies $a_i(n) = a_{i+1}(n)$, so

$$\forall n > A_i, \mathbf{D}_i(n) = \mathbf{D}_{i+1}(n).$$

For levels $n < A_i$ above the most recent common ancestor, note that by Lemma 2.1 the subtrees descending from different surviving offspring of $(-A_i, a_i(A_i))$ are independent copies of multi-type branching processes whose initial individuals are of types $\mathbf{D}_i(A_i)_{[1]}, \mathbf{D}_i(A_i)_{[2]}, \dots$ and which are conditioned to survive for at least $n' := A_i - 1$ generations. In particular, the subtree containing the lineage of $(0, i + 1)$ above $(-A_i, a_i(A_i))$ is independent of the subtree whose lineage is recorded in $\{\mathbf{D}_i(n), n < A_i\}$ and is initiated by an individual of type $\ell := \mathbf{D}_i(A_i)_{[2]}$ (See Figure 2.4). By definition $(\mathbf{D}_{i+1}(n), 1 \leq n < A_i)$ records the survivor types (and their siblings) along the left most ancestral lineage of $(0, i + 1)$ above the level A_i . Furthermore, by Lemma 2.1 the distribution of this sequence of type vectors for a multi-type branching process with initial individual of type ℓ conditioned to survive at least n' generations is distributed as the sequence of type vectors $\boldsymbol{\eta}_{n', \ell}$. So,

$$(\mathbf{D}_{i+1}(1), \dots, \mathbf{D}_{i+1}(A_i - 1)) \stackrel{d}{=} \boldsymbol{\eta}_{A_i - 1, \mathbf{D}_i(A_i)_{[2]}} \Leftrightarrow \forall 1 \leq n < A_i, \mathbf{D}_{i+1}(n) \stackrel{d}{=} \boldsymbol{\eta}_{A_i - 1, \mathbf{D}_i(A_i)_{[2]}}(n).$$

As in the single type case, the sequence $\mathbf{D}_{i+1} = (\mathbf{D}_{i+1}(n), n \geq 1)$ depends only on \mathbf{D}_i and not on $\mathbf{D}_{i'}$ for $i' < i$; and its transition law is determined by values of $(\mathbf{D}_i(n), n \geq A_i)$ and an independent random variable $\boldsymbol{\eta}_{n', \ell}$ with $n' = A_i - 1$ and $\ell = \mathbf{D}_i(A_i)_{[2]}$. \star

The Markov chain $(\mathbf{D}_i)_{i \geq 1}$ allows us to calculate some statistical features of the ancestral tree. The most relevant are coalescence times $(A_i)_{i \geq 1}$ which indicate the shape of the ancestral tree, and form a non-Markovian process.

We have the following result on the law of the coalescent time A_1 .

Corollary 2.1. For a single type value $\mathbf{a}_{[n]} \in \{1, 2, \dots\}$ and branching process $\mathbf{Z} = (\mathbf{Z}^{(n)})_{n \geq 0}$ with infinitely many types

$$\mathbb{P}(A_1 > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \mathbb{P}\left(\sum_{\ell=1}^{\infty} Z_{\ell}^{(n)} = 1 \mid \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}}\right). \quad (2.6)$$

Proof. The fact that $\{A_1 \neq 1, \dots, A_1 \neq n\}$ iff the subtree of the ancestor $(-n, a_1(n))$ of $(0, 1)$ at generation $-n$ has exactly one offspring with surviving progeny, directly implies that

$$\mathbb{P}(A_1 > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \mathbb{P}\left(\sum_{\ell \geq 1} Z_{\ell}^{(n)} = 1 \mid \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}}\right).$$



We next give a more general formula for the *joint law* of a coalescent time A_1 of individuals $(0, 1)$ and $(0, 2)$ together with the values of types along the whole ancestral lineage $\mathbf{A}_0 = (t(0, a_1(0)), t(-1, a_1(1)), \dots)$ of individual $(0, 1)$. It illustrates the role of ancestral types when determining branching times in the ancestral tree of the standing population. For a sequence $\mathbf{a} \in \{1, 2, \dots\}^{\mathbb{N}_0}$ let $\mathbf{a}_{|n'}$ denote the vector of the first coordinates up to n' -th one in this sequence $\mathbf{a}_{|n'} := (\mathbf{a}_{[0]}, \mathbf{a}_{[1]}, \dots, \mathbf{a}_{[n']})$.

Proposition 2.1. For a sequence of types $\mathbf{a} = (\mathbf{a}_{[0]}, \mathbf{a}_{[1]}, \dots) \in \{1, 2, \dots\}^{\mathbb{N}_0}$

$$\mathbb{P}(A_1 > n, \mathbf{A}_{0|n-1} = \mathbf{a}_{|n-1} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \frac{1}{p_{n, \mathbf{a}_{[n]}}} \prod_{n'=1}^n \left(\frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}} \right), \quad (2.7)$$

where $\mathbf{1} - \mathbf{p}_{n'-1} := (1 - p_{n'-1,1}, 1 - p_{n'-1,2}, \dots) = \mathbf{f}^{(n'-1)}(\mathbf{0})$ is the vector of extinction probabilities by generation $n' - 1$ and \mathbf{f} is the probability generating function of the multi-type branching process.

Proof. Observe that $\{A_1 \neq 1, \dots, A_1 \neq n\}$ iff all the ancestors $(-n', a_1(n'))$ of $(0, 1)$ at generation $-n'$, $1 \leq n' \leq n$ have exactly one offspring with surviving progeny. Observe

also that when types of ancestral individuals are known, by Lemma 2.1 the events of having exactly one offspring surviving progeny are independent across different generations. If $\mathbf{a}_{[n']} = \mathbf{A}_{0[n']}$ denotes the type of the ancestor $(-n', a_1(n'))$ of $(0, 1)$ at generation $-n'$, furthermore by Lemma 2.1 then those events can be expressed in terms of the random variable $\boldsymbol{\eta}_{n, \mathbf{a}_{[n]}}$ and in terms of the random variables $\zeta'_{n', \mathbf{a}_{[n']}}, 1 \leq n' \leq n$ as

$$\begin{aligned}
& \mathbb{P}(A_1 > n, \mathbf{A}_{0[n-1]} = \mathbf{a}_{[n-1]} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) \\
&= \mathbb{P}(\boldsymbol{\eta}_{n, \mathbf{a}_{[n]}} = (\{\mathbf{a}_{[0]}\}, \{\mathbf{a}_{[1]}\}, \dots, \{\mathbf{a}_{[n-1]}\})) \\
&= \mathbb{P}(\forall 1 \leq n' \leq n: \zeta'_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta'_{n', \mathbf{a}_{[n']}, \ell'} = 0 \forall \ell' \neq \mathbf{a}_{[n'-1]}) \\
&= \prod_{n'=1}^n \mathbb{P}(\zeta'_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta'_{n', \mathbf{a}_{[n']}, \ell'} = 0 \forall \ell' \neq \mathbf{a}_{[n'-1]}),
\end{aligned}$$

where we can write the above as a product because the subtrees descending from different offspring are independent. For each product term we have

$$\begin{aligned}
& \mathbb{P}(\zeta'_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta'_{n', \mathbf{a}_{[n']}, \ell'} = 0 \forall \ell' \neq \mathbf{a}_{[n'-1]}) \\
&= \frac{\mathbb{P}(\zeta'_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta'_{n', \mathbf{a}_{[n']}, \ell'} = 0 \forall \ell' \neq \mathbf{a}_{[n'-1]})}{\mathbb{P}(\sum_{\ell' \geq 1} \zeta'_{n', \mathbf{a}_{[n']}, \ell'} \neq 0)}.
\end{aligned}$$

Conditioning on the value of variable $\boldsymbol{\xi}_{\mathbf{a}_{[n']}}$ which, when Bernoulli sampled by the vector $\mathbf{p}_{n'-1} := \mathbf{1} - \mathbf{f}^{(n'-1)}(0, 0, \dots)$ of survival probabilities of different types by generation $n' - 1$ (i.e., each entry i is kept with probability $p_{n'-1, i}$), gives the distribution of $\zeta'_{n', \mathbf{a}_{[n]}}$,

we get for the numerator

$$\begin{aligned}
& \mathbb{P}(\zeta_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta_{n', \mathbf{a}_{[n']}, \ell'} = 0 \ \forall \ell' \neq \mathbf{a}_{[n'-1]}) \\
&= \mathbb{E}\left(\mathbb{P}(\zeta_{n', \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \zeta_{n', \mathbf{a}_{[n']}, \ell'} = 0 \ \forall \ell' \neq \mathbf{a}_{[n'-1]} \mid \xi_{\mathbf{a}_{[n']}})\right) \\
&= \mathbb{E}\left(\xi_{\mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} \frac{p_{n'-1, \mathbf{a}_{[n'-1]}}}{(1 - p_{n'-1, \mathbf{a}_{[n'-1]}})} \prod_{\ell' \geq 1} (1 - p_{n'-1, \ell'})^{\xi_{\mathbf{a}_{[n']}, \ell'}}\right), \\
&= p_{n'-1, \mathbf{a}_{[n'-1]}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}},
\end{aligned}$$

and for the denominator

$$\mathbb{P}\left(\sum_{\ell' \geq 1} \zeta_{n', \mathbf{a}_{[n']}, \ell'} \geq 1\right) = 1 - f_{\mathbf{a}_{[n']}}^{(n')}(0, 0, \dots) = p_{n', \mathbf{a}_{[n']}}.$$

Since $p_{0, \mathbf{a}_{[0]}} = 1$, we have

$$\begin{aligned}
\mathbb{P}(A_1 > n, \mathbf{A}_{0|n-1} = \mathbf{a}_{|n-1} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) &= \prod_{n'=1}^n \left(\frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}} \frac{p_{n'-1, \mathbf{a}_{[n'-1]}}}{p_{n', \mathbf{a}_{[n']}}} \right) \\
&= \frac{1}{p_{n, \mathbf{a}_{[n]}}} \prod_{n'=1}^n \left(\frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}} \right).
\end{aligned}$$

Note that for $n' = 1$ the evaluation of the derivative is at $\mathbf{s} = \mathbf{1} - \mathbf{p}_0 = \mathbf{0}$. ⊛

An easy modification of the formula above gives $\mathbb{P}(A_1 = n, \mathbf{A}_{0|n-1} = \mathbf{a}_{|n-1} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]})$.

Remark 2.1. *In terms of applications the joint law of A_1 and $\mathbf{A}_{0|n-1}$ is more useful for reconstructing ancestral trees. Note that the choice of embedding the offspring of each parent uniformly at random in the tree is not reflected in these formulae at all. This is in particular evident in (2.6). Moreover, (2.6) can be obtained from (2.7) by summing over*

all the possible values of $\mathbf{A}_{0[n']}$ of types of the $0 \leq n' < n$ generations' ancestors of the individual $(0, 1)$ which start with $\mathbf{A}_{0[n]} = \mathbf{a}_{[n]}$. In the next section we consider a different choice of ordering the offspring for a specific offspring distribution, and show consistency of these two expressions.

A statistical feature which indicates the distribution of types in the ancestral tree are coalescence times between individuals in the standing population that are of the same type. Suppose the type of the first individual at generation 0 is $\mathbf{A}_{0[0]} = \ell$, and define the sequence $i_{\ell,0} := 0$, $i_{\ell,1} := \min\{i' > 0 : \mathbf{A}_{i'[0]} = \ell\}$, \dots , $i_{\ell,i} := \min\{i' > i_{\ell,i-1} : \mathbf{A}_{i'[0]} = \ell\}, \dots$ representing the indices of consecutive individuals of type ℓ from the standing population.

Definition 2.2. Define the sequence of same-type coalescence times for individuals of type ℓ by

$$B_{\ell,1} := \max\{A_{i_{\ell,1}}, \dots, A_{i_{\ell,2}-1}\}, \dots, B_{\ell,i} := \max\{A_{i_{\ell,i}}, \dots, A_{i_{\ell,i+1}-1}\}, \dots$$

As before, we also have the following result on the law of the same-type coalescent time $B_{\ell,1}$.

Corollary 2.2. For a single type value $\mathbf{a}_{[n]} \in \{1, 2, \dots\}$ and a multi-type branching process $\mathbf{Z} = (\mathbf{Z}^{(n)})_{n \geq 0}$, we have

$$\mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) = \mathbb{P}(Z_{\ell}^{(n)} = 1 \mid Z_{\ell}^{(n)} \neq 0, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}}). \quad (2.8)$$

Proof. The fact that $\{B_{\ell,1} > n\}$ iff the subtree of the ancestor $(-n, a_1(n))$ of $(0, 1)$ at generations $-n$ has exactly one descendant of type ℓ after n generations, directly implies

that

$$\mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) = \mathbb{P}(Z_\ell^{(n)} = 1 \mid Z_\ell^{(n)} \neq 0, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}}).$$



Using the values of types on the left most infinite spine \mathbf{A}_0 we can determine the joint distribution of $B_{\mathbf{A}_{0[0]},1}$ and $\mathbf{A}_{0[n-1]}$ as follows.

Proposition 2.2. *For a sequence of types $\mathbf{a} = (\mathbf{a}_{[0]}, \mathbf{a}_{[1]}, \dots) \in \{1, 2, \dots\}^{\mathbb{N}_0}$ with $\mathbf{a}_{[0]} = \ell$,*

$$\mathbb{P}(B_{\ell,1} > n, \mathbf{A}_{0[n-1]} = \mathbf{a}_{[n-1]} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) \quad (2.9)$$

$$= \frac{1}{p_{(n,\ell), \mathbf{a}_{[n-1]}} \prod_{n'=1}^n \left(\frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}^{(n'-1, \ell')}} \right)},$$

where $\mathbf{1} - \mathbf{p}^{(n'-1, \ell')} := (1 - p^{(n'-1, \ell'), 1}, 1 - p^{(n'-1, \ell'), 2}, \dots) = \mathbf{f}^{(n'-1)}(\hat{\mathbf{e}}_{\ell'})$ with $\hat{\mathbf{e}}_{\ell'} = \mathbf{1} - \mathbf{e}_{\ell'}$, is the vector of extinction probabilities for lineages with type ℓ' descendants after $n' - 1$ generations.

Proof. Observe that $\{B_{\ell,1} \neq 1, \dots, B_{\ell,1} \neq n\}$ iff all the ancestors $(-n', a_1(n'))$ of $(0, 1)$ at generations $-n', 1 \leq n' \leq n$ have exactly one descendant in the standing population that has type ℓ . As before, let $\mathbf{a}_{[n']} = \mathbf{A}_{0[n']}$ denote the type of the ancestor $(-n', a_1(n'))$ of $(0, 1)$ at generation $-n'$, and note that $\mathbf{a}_{[0]} = \ell$.

We need to introduce new random variables which will count the number of offspring with descendants of type ℓ in the standing population. If $\mathbf{f}^{(n'-1)} = (f_1^{(n'-1)}, f_2^{(n'-1)}, \dots)$ is the probability generating function of the $n' - 1$ generation in a multi-type branching process initiated by individuals of type $\{1, 2, \dots\}$, then the probability that a multi-type process after $n' - 1$ generations has no individuals of type ℓ is given by the vector $\mathbf{f}^{(n'-1)}(\hat{\mathbf{e}}_\ell)$, where $\hat{\mathbf{e}}_\ell := \mathbf{1} - \mathbf{e}_\ell$. Let $\mathbf{p}^{(n'-1, \ell)} := \mathbf{1} - \mathbf{f}^{(n'-1)}(\hat{\mathbf{e}}_\ell)$ denote the probability of having at least one descendant of type ℓ after $n' - 1$ generations, that is, for each

$j' \in \{1, 2, \dots\}$, we have $p_{(n'-1, \ell), j'} = 1 - f_{j'}^{(n'-1)}(\hat{\boldsymbol{e}}_\ell)$. For all different offspring types $j' \in \{1, 2, \dots\}$, let $\{\varepsilon_{(n', \ell), j'}^1, \varepsilon_{(n', \ell), j'}^2, \dots\}$ be independent sequences of Bernoulli variables with parameters $\mathbb{P}(\varepsilon_{(n', \ell), j'}^m = 1) = p_{(n'-1, \ell), j'}$ (for $m=1, 2, \dots$). For an independent offspring variable $\boldsymbol{\xi}_j$ with $\boldsymbol{\xi}_{j, j'}$ offspring of type j' define

$$\boldsymbol{\zeta}_{(n', \ell), j} := \left(\sum_{m=1}^{\boldsymbol{\xi}_{j,1}} \varepsilon_{(n', \ell), 1}^m, \sum_{m=1}^{\boldsymbol{\xi}_{j,2}} \varepsilon_{(n', \ell), 2}^m, \dots \right),$$

whose j' coordinate is denoted by $\boldsymbol{\zeta}_{(n', \ell), j, j'}$. Then, $\boldsymbol{\zeta}_{(n', \ell), j}$ records the number of offspring (of different types), in the first generation of a multi-type branching process initiated by an individual of type j , which have at least one descendant of type ℓ after $n' - 1$ generations.

Then the law of $\boldsymbol{\zeta}'_{(n', \ell), j}$ representing the number of offspring (of different types) at generation $-(n' - 1)$ of a type j ancestor from generation $-n'$ whose descendants contain an individual of type ℓ in the standing population, given that there is at least one, is given by:

$$\boldsymbol{\zeta}'_{(n', \ell), j} \stackrel{d}{=} \left(\boldsymbol{\zeta}_{(n', \ell), j} \mid \sum_{j' \geq 1} \boldsymbol{\zeta}_{(n', \ell), j, j'} \geq 1 \right).$$

The event $\{B_{\ell, 1} > n\} = \{B_{\ell, 1} \neq 1, \dots, B_{\ell, 1} \neq n\}$ can now be expressed in terms of the newly defined random variables $\boldsymbol{\zeta}'_{(n', \ell), \mathbf{a}_{[n']}}, 1 \leq n' \leq n$, as

$$\begin{aligned} & \mathbb{P}(B_{\ell, 1} > n, \mathbf{A}_{0[n-1]} = \mathbf{a}_{[n-1]} \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) \\ &= \mathbb{P}(\forall 1 \leq n' \leq n : \boldsymbol{\zeta}'_{(n', \ell), \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \boldsymbol{\zeta}'_{(n', \ell), \mathbf{a}_{[n']}, j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]}) \\ &= \prod_{n'=1}^n \mathbb{P}(\boldsymbol{\zeta}'_{(n', \ell), \mathbf{a}_{[n']}, \mathbf{a}_{[n'-1]}} = 1, \boldsymbol{\zeta}'_{(n', \ell), \mathbf{a}_{[n']}, j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]}), \end{aligned}$$

where the product form follows since, by Lemma 2.1, subtrees of different offspring are

independent. For each product term we have

$$\begin{aligned} & \mathbb{P}(\zeta'_{(n',\ell),\mathbf{a}_{[n']},\mathbf{a}_{[n'-1]}} = 1, \zeta'_{(n',\ell),\mathbf{a}_{[n']},j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]}) \\ &= \frac{\mathbb{P}(\zeta_{(n',\ell),\mathbf{a}_{[n']},\mathbf{a}_{[n'-1]}} = 1, \zeta_{(n',\ell),\mathbf{a}_{[n']},j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]})}{\mathbb{P}(\sum_{j' \geq 1} \zeta_{(n',\ell),j',j'} \geq 1)}. \end{aligned}$$

A similar calculation to the one in the proof of Proposition 2.1, conditioning on $\xi_{\mathbf{a}_{[n']}}$, gives the numerator to be

$$\begin{aligned} & \mathbb{P}(\zeta_{(n',\ell),\mathbf{a}_{[n']},\mathbf{a}_{[n'-1]}} = 1, \zeta_{(n',\ell),\mathbf{a}_{[n']},j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]}) \\ &= \mathbb{E}\left(\mathbb{P}(\zeta_{(n',\ell),\mathbf{a}_{[n']},\mathbf{a}_{[n'-1]}} = 1, \zeta_{(n',\ell),\mathbf{a}_{[n']},j'} = 0 \forall j' \neq \mathbf{a}_{[n'-1]} \mid \xi_{\mathbf{a}_{[n]}})\right) \\ &= \mathbb{E}\left(\xi_{\mathbf{a}_{[n']},\mathbf{a}_{[n'-1]}} \frac{p_{(n'-1,\ell),\mathbf{a}_{[n'-1]}}}{(1 - p_{(n'-1,\ell),\mathbf{a}_{[n'-1]}})} \prod_{j' \geq 1} (1 - p_{(n'-1,\ell),j'})^{\xi_{\mathbf{a}_{[n']},j'}}\right) \\ &= p_{(n'-1,\ell),\mathbf{a}_{[n'-1]}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial \mathbf{s}_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{(n'-1,\ell)}}, \end{aligned}$$

while the denominator is calculated in the same way and equals

$$\begin{aligned} \mathbb{P}\left(\sum_{\ell' \geq 1} \zeta_{(n',\ell),\mathbf{a}_{[n']},\ell'} \geq 1\right) &= 1 - \mathbb{E}\left(\prod_{\ell' \geq 1} (1 - p_{(n'-1,\ell),\ell'})^{\xi_{\mathbf{a}_{[n']},\ell'}}\right) \\ &= 1 - f_{\mathbf{a}_{[n']}}(\mathbf{1} - \mathbf{p}_{(n'-1,\ell)}) = 1 - f_{\mathbf{a}_{[n']}}^{(n')}(\hat{\mathbf{e}}_\ell) = p_{(n',\ell),\mathbf{a}_{[n']}}. \end{aligned}$$

The cross terms of probabilities of a lineage with descendants of type ℓ cancel, and

$$\mathbb{P}(B_{\ell,1} > n, \mathbf{A}_{0|n-1} = \mathbf{a}_{|n-1} \mid \mathbf{A}_{0|n} = \mathbf{a}_{[n]}, \mathbf{A}_{0|0} = \ell) = \frac{1}{p_{(n,\ell),\mathbf{a}_{[n]}}} \prod_{n'=1}^n \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial \mathbf{s}_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{(n'-1,\ell)}}.$$



It might be tempting to provide a formula for the distribution of A_2, A_3, \dots and $B_{\ell,2}, B_{\ell,3}, \dots$ in a similar vein using the values of the types on the ancestral lineage of the individuals $(0, 2), (0, 3), \dots$ respectively. The information, analogous to that of types of individuals on the left most (infinite into the past) spine \mathbf{A}_0 used in the formulae for A_1 and $B_{\ell,1}$, which one would need to use for A_{i+2} would be the types along the ancestral lineage of $(0, i + 1)$. That is, for $i \geq 0$ one could define the *infinite ancestral lineage* \mathbf{A}_i^∞ of individual $(0, i + 1)$ as the infinite sequence:

$$\mathbf{A}_i^\infty := (t(0, a_{i+1}(0)), t(-1, a_{i+1}(1)), t(-2, a_{i+1}(2)), \dots).$$

Note that the restriction of \mathbf{A}_i^∞ to its first A_i entries equals the sequence \mathbf{A}_i called the *ancestral lineage* of $(0, i + 1)$, and that $\mathbf{A}_0^\infty = \mathbf{A}_0$. It is easy to see, as a consequence of Theorem 2.1, that $(\mathbf{A}_i^\infty)_{i \geq 1}$ is also a functional of $(\mathbf{D}_i)_{i \geq 1}$ given by:

$$\mathbf{A}_i^\infty = (\mathbf{D}_{i+1}(1)_{[1]}, \mathbf{D}_{i+1}(2)_{[1]}, \mathbf{D}_{i+1}(3)_{[1]}, \dots).$$

This follows from the fact that the first $1 \leq n \leq A_i$ entries in this sequence are the same as in \mathbf{A}_i , while for the subsequent $n > A_i$ entries the ancestry of $(0, i + 1)$ is equal to the ancestry of $(0, i)$ (as their ancestors already coalesced) and we have $\mathbf{D}_{i+1}(n) = \mathbf{D}_i(n)$.

Providing a formula for $\mathbb{P}(A_{i+2} > n | \mathbf{A}_{i+1[n]} = \mathbf{a}_{[n]})$ for any $i \geq 0$ can be done only in case the branching mechanism is such that in the coalescent point-process all points are simple (have multiplicities equal to one). This is because, in case of multiple coalescence points, all ancestral lineages, except for that of $(0, 1)$, in addition to information about the lineage of individual $(0, i)$ also contains information about the ancestral lineages of $(0, i'), 1 \leq i' < i$. In other words, the calculation (used in the proofs of the Propositions 2.1 and 2.2) which exploits the equivalence $\{A_1 > n, \mathbf{A}_{0|n-1} = \mathbf{a}_{|n-1}\}$ iff $\{\text{individuals } \mathbf{a}_{[n]}, 1 \leq$

$n' \leq n$ on the ancestral lineage \mathbf{A}_0 have a single offspring with descendants surviving to generation 0} is no longer valid in general. This is a clear consequence of the fact that the process $(\mathbf{A}_i, i \geq 0)$ is itself not Markovian, unless we are in the special case for the offspring distribution which is of linear-fractional form. In the next section we explore this special case, and extend the above results for the coalescent times as well as for the same-type coalescence times.

In case the branching mechanism is such that in the coalescent point-process all points are simple, the results of Proposition 2.1 can be used to simulate ancestral trees with types along their lineages. In this case, the same joint law (2.7) applies to each A_i and $\mathbf{A}_{i-1|A_i-1}$ for $i \geq 1$. A modification of formula (2.7) gives $\mathbb{P}(A_i = n, \mathbf{A}_{i-1|n-1} = \mathbf{a}_{|n-1} | \mathbf{A}_{i-1|n} = \mathbf{a}_{[n]})$. In order to simulate the types along the ancestral tree, one would start by drawing from this distribution to obtain types along the ancestral lineage of individual $(0, 2)$ until its MRCA with individual $(0, 1)$. Then, one would draw independently from this distribution to obtain types along the ancestral lineage of individual $(0, 3)$ until its MRCA with individual $(0, 2)$, and so on. Notice that, if the value of MRCA for $(0, 2)$ and $(0, 1)$ drawn is smaller than the MRCA for $(0, 3)$ and $(0, 2)$, then the types on the ancestral lineage of individual $(0, 3)$ continue below its MRCA with individual $(0, 2)$ and are drawn along the left most ancestral spine \mathbf{A}_0 (see the example of $i = 3$ or $i = 5$ in Figure 2.2(right)). According to this algorithm one can simulate in a sequential manner the genealogy of an arbitrary number of individuals in the standing population.

Remark 2.2. *The construction of the ancestral tree in case of the most general branching process is based on the Markovian property of the auxiliary process $(\mathbf{D}_i)_{i \geq 1}$. A simulation algorithm for ancestral trees in general would require the simulation of the whole process $(\mathbf{D}_i)_{i \geq 1}$. To initiate this process requires either knowing \mathbf{D}_1 a priori, or drawing \mathbf{D}_1 from $\boldsymbol{\eta}_{\infty, \ell_\infty}$ for some (infinitely old) originating type ℓ_∞ . One way to draw from this distribution would be to assume non-extinction and time-reverse the quasi-stationary distribution from*

time 0 for the branching process (Q -process at end of Section 1.1). For the case of a single type branching process a construction like this was discussed in [14]. The multi-type extension is straightforward, with the infinitely old originating individual having, in the $\rho \leq 1$ case, the size-biased version of the offspring distribution (a branching distribution conditioned on never becoming extinct) given by

$$\widehat{\mathbb{P}}(\mathbf{Z}^{(1)} = \mathbf{z} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell) = \frac{\mathbb{P}(\mathbf{Z}^{(1)} = \mathbf{z} | \mathbf{Z}^{(0)} = \mathbf{e}_\ell)}{\rho} \frac{\mathbf{z} \cdot \mathbf{v}}{\mathbf{e}_\ell \cdot \mathbf{v}},$$

where \mathbf{u} is the right eigenvector of the mean matrix \mathbf{M} obtained under the assumptions at the end of Section 1.1. Refer to that section (or Seneta [42]) for the existence of quasi-stationary distributions in the countably many types branching process.

2.2 Special case: Multi-type linear-fractional branching processes

Many of the complications which arise in calculating the distribution of ancestral trees in multi-type branching processes simplify a great deal in the special case when the offspring distribution is of the linear-fractional (LF) type. This type of offspring distribution leads to a number of particularly nice features involving the memoryless property of the geometric distribution. The definition of the multi-type linear-fractional offspring distribution can be found in Section 1.1.1. We next give a series of specific results for the distribution of the ancestral tree of the standing population, which both illustrate and extend our general results from the previous Section.

2.2.1 Coalescent times in linear-fractional ancestral tree

We use the coalescent point-process construction to get simpler results for the distribution of coalescent times (and same-type coalescent times) for this special class of multi-type branching processes. For this purpose we make one change in our original construction pertaining to the embedding of the multi-type tree in the plane. For a general offspring distribution we made the assumption that the offspring of any parent are embedded in a left to right order chosen *uniformly* at random from all possible orderings. For the linear-fractional offspring distribution we make a particular assumption that the offspring with distribution given by the vector \mathbf{h} . is embedded as the left most individual, followed by the rest of the offspring according to an arbitrary order.

Proposition 2.3. *The coalescence times $(A_i)_{i \geq 0}$ in the ancestral tree of a LF($\mathbf{H}, \mathbf{g}, m$) branching process are independent identically distributed variables with*

$$\mathbb{P}(A_1 > n) = \prod_{n'=1}^n \frac{1}{1 + m - m \sum_{\ell' \geq 1} \mathbf{g}_{\ell'} h_{\ell'0}^{(n'-1)}} = \frac{1}{1 + m^{(n)}}, \quad (2.10)$$

where $h_{\ell'0}^{(n'-1)} = 1 - \mathbf{h}_{\ell'}^{(n'-1)} \mathbf{1}^\top$, $\mathbf{h}_{\ell'}^{(n'-1)}$ is the ℓ' -th row of the matrix $\mathbf{H}^{(n'-1)}$ from (1.1), $h_{\ell'0}^{(0)} = 0 \forall \ell'$. The law of the coalescent times also satisfies $\mathbb{P}(A_1 > n | \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \mathbb{P}(A_1 > n)$.

Proof. The most immediate approach to the proof uses Corollary 2.1 and follows from the fact that in a multi-type LF branching process, all offspring in the n -th generation other than the first one (which according to our current convention is the left-most one) are independent of the type of the parent, and have a multivariate-geometric distribution with mean $m^{(n)}$ and type distribution given by $g^{(n)}$ (whose formula is given in (1.1)). This fact was also used in [41] (see Sec 4.1) to establish the formula 1.1 using the jumping contour representation of the branching process and its nice Markovian structure. Since

by Corollary 2.1

$$\mathbb{P}(A_1 > n | \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \mathbb{P} \left(\sum_{\ell \geq 1} Z_\ell^{(n)} = 1 \mid \mathbf{Z}^{(n)} \neq \mathbf{0}, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}} \right),$$

by the result of Theorem 1.5 the right-hand side is simply the probability that the geometric variable with mean $m^{(n)}$ is 0, which is equal to $1/(1 + m^{(n)})$ regardless of the type of the initial individual $\mathbf{Z}^{(0)}$.

In addition to the above approach, we thought it would be instructive to show the agreement with the formula for the joint law of A_1 and $\mathbf{A}_{0|n-1}$ in Proposition 2.1 via a summation approach. We start from a formula based on (2.7)

$$\mathbb{P}(A_1 > n | \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = \sum_{\mathbf{a}_{[0], \dots, \mathbf{a}_{[n-1]}} \prod_{n'=1}^n \left(\frac{p_{n'-1, \mathbf{a}_{[n'-1]}}}{p_{n', \mathbf{a}_{[n']}}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}} \right),$$

in which we perform the summation in a ‘top-down’ order, from possible values for $\mathbf{a}_{[0]}$ down to $\mathbf{a}_{[n-1]}$, and observe that since the summations are nested we can write this as $\mathbb{P}(A_1 > n | \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}) = c_1 \cdots c_n$, where

$$c_{n'} \equiv c_{n'}(\mathbf{a}_{[n']}) = \frac{1}{p_{n', \mathbf{a}_{[n']}}} \sum_{\mathbf{a}_{[n'-1]} \geq 1} p_{n'-1, \mathbf{a}_{[n'-1]}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial s_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}_{n'-1}}. \quad (2.11)$$

It turns out that in the linear fractional case the $c_{n'}$ are constants which do not depend on $\mathbf{a}_{[n']}$. More specifically

$$c_{n'}(\mathbf{a}_{[n']}) = \frac{1}{U_{n'}},$$

where $U_{n'} = 1 + m - m \sum_{\ell' \geq 1} g_{\ell'} h_{\ell'0}^{(n'-1)}$, so

$$\mathbb{P}(A_1 > n) = \prod_{n'=1}^n \frac{1}{1 + m - m \sum_{\ell' \geq 1} \mathbf{g}_{\ell'} h_{\ell'0}^{(n'-1)}}.$$

We can rewrite this formula using the parameters defined in (1.1) from Theorem 1.5 according to which

$$\mathbf{H}^{(n)} = \mathbf{M}^n - \frac{m^{(n)}}{1 + m^{(n)}} \mathbf{M}^n \mathbf{1}^\top \mathbf{g}^{(n)}.$$

Multiplying by \mathbf{g} on the left and by $\mathbf{1}^\top$ on the right yields $\mathbf{g} \mathbf{H}^{(n)} \mathbf{1}^\top = \frac{\mathbf{g} \mathbf{M}^n \mathbf{1}^\top}{1 + m^{(n)}}$ or $\sum_{\ell \geq 1} g_\ell h_{\ell 0}^{(n)} = 1 - \frac{\mathbf{g} \mathbf{M}^n \mathbf{1}^\top}{1 + m^{(n)}}$. Using this equality in the formula for $\mathbb{P}(A_1 > n)$ we get

$$\begin{aligned} \mathbb{P}(A_1 > n) &= \prod_{n'=1}^n \frac{1}{1 + m - m \sum_{i \geq 1} g_i h_{i0}^{(n'-1)}} = \prod_{n'=1}^n \frac{1}{1 + m \frac{\mathbf{g} \mathbf{M}^{n'-1} \mathbf{1}^\top}{1 + m^{(n'-1)}}} \\ &= \prod_{n'=1}^n \frac{1 + m^{(n'-1)}}{1 + m^{(n'-1)} + m \mathbf{g} \mathbf{M}^{n'-1} \mathbf{1}^\top} = \prod_{n'=1}^n \frac{1 + m^{(n'-1)}}{1 + m^{(n')}} = \frac{1}{1 + m^{(n)}}, \end{aligned}$$

because $m^{(n'-1)} + m \mathbf{g} \mathbf{M}^{n'-1} \mathbf{1}^\top = m \mathbf{g} (\mathbf{I} + \mathbf{M} + \dots + \mathbf{M}^{n'-2}) \mathbf{1}^\top + m \mathbf{g} \mathbf{M}^{n'-1} \mathbf{1}^\top = m^{(n')}$, and $m^{(0)} = 0$.



Proposition 2.4. *For any type $\ell \in \{1, 2, \dots\}$, the same-type coalescence times $(B_{\ell,i})_{i \geq 0}$ are independent identically distributed variables with,*

$$\mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[0]} = \ell) = \prod_{n'=1}^n \frac{1}{1 + m - m \sum_{\ell' \geq 1} \mathbf{g}_{\ell'} \tilde{h}_{\ell'0}^{(n'-1)}} = \frac{1}{1 + m^{(n)} g_\ell^{(n)}}, \quad (2.12)$$

where, for $n' > 1$, $\tilde{h}_{\ell'0}^{(n'-1)}$ is given by

$$\tilde{h}_{\ell'0}^{(n'-1)} = h_{\ell'0}^{(n'-1)} + \frac{1 - h_{\ell'0}^{(n'-1)} - h_{\ell'\ell}^{(n'-1)}}{1 + m^{(n'-1)} g_\ell^{(n'-1)}},$$

and $\tilde{h}_{\ell'0}^{(0)} = 1 \forall \ell' \neq \ell$, while $\tilde{h}_{\ell 0}^{(0)} = 0$.

Proof. By Corollary 2.2

$$\mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) = \mathbb{P}(Z_\ell^{(n)} = 1 \mid Z_\ell^{(n)} \neq 0, \mathbf{Z}^{(0)} = \mathbf{e}_{\mathbf{a}_{[n]}}),$$

which is equal to $\partial_{s_\ell} f_{\mathbf{a}_{[n]}}^{(n)}(\hat{\mathbf{e}}_\ell)/(1 - f_{\mathbf{a}_{[n]}}^{(n)}(\hat{\mathbf{e}}_\ell))$, and as a result of Theorem 1.5 and some simple arithmetic can be shown to be equal to $1/(1 + m^{(n)}g_\ell^{(n)})$ regardless of the type of the initial individual $\mathbf{Z}^{(0)}$.

As in the proof of Proposition 2.3 we also show a different approach using (2.9) from Proposition 2.2, that is

$$\mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[n]} = \mathbf{a}_{[n]}, \mathbf{A}_{0[0]} = \ell) = \sum_{\mathbf{a}_{[0]}, \dots, \mathbf{a}_{[n-1]}} \prod_{n'=1}^n \left(\frac{p^{(n'-1, \ell), \mathbf{a}_{[n'-1]}}}{p^{(n', \ell), \mathbf{a}_{[n']}}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial \mathbf{s}_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}^{(n'-1, \ell)}} \right).$$

For each $1 \leq n' \leq n-1$ the sums

$$\tilde{c}_{n'} \equiv \tilde{c}_{n'}(\mathbf{a}_{[n']}) = \frac{1}{p^{(n', \ell), \mathbf{a}_{[n']}}} \sum_{\mathbf{a}_{[n'-1]} \geq \mathbf{1}} p^{(n'-1, \ell), \mathbf{a}_{[n'-1]}} \frac{\partial f_{\mathbf{a}_{[n']}}(\mathbf{s})}{\partial \mathbf{s}_{\mathbf{a}_{[n'-1]}}} \Big|_{\mathbf{s}=\mathbf{1}-\mathbf{p}^{(n'-1, \ell)}} \quad (2.13)$$

are independent of the value of $\mathbf{a}_{[n']}$ in the linear fractional case, with

$$\tilde{c}_{n'} = \frac{1}{\tilde{U}_{n'}} = \frac{1}{1 + m - m \sum_{\ell' \geq 1} g_{\ell'} \tilde{h}_{\ell'0}^{(n'-1)}}$$

and

$$\tilde{h}_{\ell'0}^{(n'-1)} = h_{\ell'0}^{(n'-1)} + \frac{1 - h_{\ell'0}^{(n'-1)} - h_{\ell'\ell}^{(n'-1)}}{1 + m^{(n'-1)}g_\ell^{(n'-1)}},$$

$$\text{so } \mathbb{P}(B_{\ell,1} > n \mid \mathbf{A}_{0[0]} = \ell) = \prod_{n'=1}^n \tilde{c}_{n'}.$$

We can rewrite this formula in a similar way as before by noting that the first product term in (2.12) is equal to $1/(1 + mg_\ell)$, and using many arithmetic steps established by the relationship of parameters in (1.1) the rest of the terms for $n' > 1$ can be shown to be equal to

$$\frac{(1 + m^{(n'-1)})(1 + m^{(n'-1)}g_\ell^{(n'-1)})}{(1 + m^{(n'-1)})(1 + mg\mathbf{H}^{(n'-1)}\mathbf{e}_\ell) + m^{(n'-1)}g_\ell^{(n'-1)}(1 + m^{(n')})} = \frac{1 + m^{(n'-1)}g_\ell^{(n'-1)}}{1 + m^{(n')}g_\ell^{(n')}},$$

hence

$$\mathbb{P}(B_{\ell,1} > n | \mathbf{A}_{0[0]} = \ell) = \frac{1}{1 + mg_\ell} \prod_{n'=2}^n \frac{1 + m^{(n'-1)} g_\ell^{(n'-1)}}{1 + m^{(n')} g_\ell^{(n')}} = \frac{1}{1 + m^{(n)} g_\ell^{(n)}}.$$



Although the multi-type LF offspring distribution has a seemingly small level of dependence between the offspring and parent type, it still affects the distribution of types in the ancestral tree. One can also consider a multi-type branching process where offspring distribution is completely independent of the parent type. In this case the shape of the tree and the types on the tree can be decoupled, and the distribution of types is only governed by the frequency of this type in the population.

Consider a special case of a LF distribution where $\mathbf{H} := \mathbf{1}^\top \mathbf{h}$, for $\mathbf{h}\mathbf{1}^\top = h_1 + h_2 + \dots \leq 1$. Then each parent has the same LF($\mathbf{h}, \mathbf{g}, m$) offspring distribution. Further, if we have no distinction between the first offspring and the rest, then we would have $\mathbf{h} := (1 - h_0)\mathbf{g}$, for $h_0 \in (0, 1)$. In this case all parents have the same offspring laws, where their number of offsprings has a single-type LF(h_0, m) distribution with probability generating function $h_0 + (1 - h_0)s/(1 + m - ms)$, and given the number of offsprings, the distribution of their types is multinomial with parameter \mathbf{g} . In this case we get the following formulae for the law of coalescence times and same-type coalescence times.

Corollary 2.3. *If the offspring distribution of each parent is independent of the parent's type with LF($(1 - h_0)\mathbf{1}^\top \mathbf{g}, \mathbf{g}, m$) distribution, then*

$$\mathbb{P}(A_1 > n) = \begin{cases} \frac{m - h_0(1 + m)}{m(1 + m)^n(1 - h_0)^n - h_0(1 + m)} & \text{if } (1 - h_0)(1 + m) \neq 1, \\ \frac{1 - h_0}{1 - h_0 + nh_0} & \text{if } (1 - h_0)(1 + m) = 1, \end{cases} \quad (2.14)$$

and

$$\mathbb{P}(B_{\ell,1} > n | \mathbf{A}_{0[0]} = \ell) = \frac{1 - \mathbb{P}(A_1 \leq n)}{1 - \mathbb{P}(A_1 \leq n)(1 - g_\ell)}. \quad (2.15)$$

Proof. One approach is to use equivalence of offspring laws for different parents. Since types do not affect the reproduction law, the ancestral tree shape can be decoupled from the individual types. In other words, we can first construct the branching process using the single-type LF offspring distribution with parameters (h_0, m) , and subsequently assign types to all individuals independently according to probabilities \mathbf{g} . As the only factor affecting the coalescent times of the standing population is the offspring number of each individual, their law is the same as in the case of the associated single-type LF coalescent times.

According to Proposition 5.1 from [29] for the single-type LF case, the coalescent times have distribution (in their notation $a \mapsto h_0, b \mapsto m/(1+m)$ and offspring mean $m \mapsto (1+m)(1-h_0)$)

$$\mathbb{P}(A_1 > n) = \frac{m - h_0(1+m)}{m(1+m)^n(1-h_0)^n - h_0(1+m)},$$

if $(1+m)(1-h_0) \neq 1$, while if $(1+m)(1-h_0) = 1$

$$\mathbb{P}(A_1 > n) = \frac{1 - h_0}{1 - h_0 + nh_0}.$$

To see that this agrees with result (2.10) note that $\mathbf{H} = (1 - h_0)\mathbf{1}^\top \mathbf{g}$, $\mathbf{M} = \mathbf{H} + m\mathbf{H}\mathbf{1}^\top \mathbf{g}$ implies

$$\mathbf{M} = (1-h_0)\mathbf{1}^\top \mathbf{g} + m(1-h_0)\mathbf{1}^\top \mathbf{g}\mathbf{1}^\top \mathbf{g} = (1-h_0)(1+m)\mathbf{1}^\top \mathbf{g}, \quad \mathbf{M}^{n'} = (1-h_0)^{n'}(1+m)^{n'}\mathbf{1}^\top \mathbf{g},$$

$$\begin{aligned}
m^{(n)} &= m\mathbf{g}(\mathbf{I} + (1 - h_0)(1 + m)\mathbf{1}^\top \mathbf{g} + \cdots + (1 - h_0)^{n-1}(1 + m)^{n-1}\mathbf{1}^\top \mathbf{g})\mathbf{1}^\top \\
&= m \sum_{n'=0}^{n-1} (1 - h_0)^{n'} (1 + m)^{n'} = m \frac{1 - (1 - h_0)^n (1 + m)^n}{1 - (1 - h_0)(1 + m)}, \quad \text{if } (1 - h_0)(1 + m) \neq 1.
\end{aligned}$$

On the other hand if $(1 - h_0)(1 + m) = 1$, then $\mathbf{M}^{n'} = \mathbf{M}$, $m^{(n)} = mn = nh_0/(1 - h_0)$. Using this in (2.10) the formula $\mathbb{P}(A_1 > n) = 1/(1 + m^{(n)})$ gives the same result as above.

From the result for A_1 we can obtain the law of $B_{\ell,1}$ using its original definition as the maximum of all coalescence times until the first next individual in the current population whose type is ℓ : $B_{\ell,1} := \max\{A_{i_{\ell,1}}, \dots, A_{i_{\ell,2}-1}\}$. The only reason why this calculation is simple is due to the decoupling of the branching tree and the individual types. Since, given the branching tree, all individuals are assigned types independently according to probabilities \mathbf{g} , the index $i_{\ell,2} := \min\{i' > i_{\ell,1} \equiv 1 : \mathbf{A}_{i'_{[0]}} = \ell\}$ is such that $i_{\ell,2} - 1$ is a random variable with a shifted geometric distribution with parameter g_ℓ . Conditioning on the value of $i_{\ell,2} - 1$, and using the fact that $(A_i)_{i \geq 1}$ is an i.i.d sequence, we get

$$\begin{aligned}
\mathbb{P}(B_{\ell,1} \leq n) &= \mathbb{E}\left(\mathbb{P}(\max\{A_1, \dots, A_{i_{\ell,2}-1}\} \leq n \mid i_{\ell,2} - 1)\right) \\
&= \sum_{i'-1=1}^{\infty} \mathbb{P}(A_1 \leq n)^{i'-1} (1 - g_\ell)^{i'-2} g_\ell = \frac{g_\ell \mathbb{P}(A_1 \leq n)}{1 - \mathbb{P}(A_1 \leq n)(1 - g_\ell)},
\end{aligned}$$

and

$$\mathbb{P}(B_{\ell,1} > n) = 1 - \mathbb{P}(B_{\ell,1} \leq n) = \frac{1 - \mathbb{P}(A_1 \leq n)}{1 - \mathbb{P}(A_1 \leq n)(1 - g_\ell)}.$$



As expected, the distribution of types \mathbf{g} has no effect on the law of $(A_i)_{i \geq 1}$ and the shape of the tree, but appears in the distribution of types in the tree as indicated by the law of $(B_{\ell,i})_{i \geq 1}$.

We can consider the process of coalescent times $(A_i)_{i \geq 0}$ as a simple point-process A on $\{1, 2, \dots\} \times \{-1, -2, \dots\}$ with intensity measure $\nu_A[\{i\} \times \{-(n+1), \dots\}] =$

$\mathbb{P}(A_1 > n), \forall i \geq 1$. Similarly for each $\ell \in \{1, 2, \dots\}$, $(B_{\ell,i})_{i \geq 0}$ can be regarded as a simple point-process B_ℓ with intensity $\nu_{B_\ell}[\{i\} \times \{-(n+1), \dots\}] = \mathbb{P}(B_{\ell,1} > n)$ on $\{1, 2, \dots\} \times \{-1, -2, \dots\}$. Note that for any $i \geq 1, n \geq 1$, (2.15) implies that

$$\nu_{B_\ell}[\{i\} \times \{-1, \dots, -n\}] = \frac{\nu_A[\{i\} \times \{-1, \dots, -n\}]g_\ell}{\nu_A[\{i\} \times \{-(n+1), \dots\}] + \nu_A[\{i\} \times \{-1, \dots, -n\}]g_\ell},$$

showing that only a fraction of all coalescent times are candidates for same-type coalescence times for type ℓ . Intuitively, when considering $B_{\ell,i}$ mark each coalescent time A_i, A_{i+1}, \dots with the probability that the next individual in the standing population is of type ℓ , which is g_ℓ . Then, from the filtered view of $B_{\ell,i}$, a coalescence time A_i, A_{i+1}, \dots either occurs outside the set $\{1, \dots, n\}$, or it occurs inside this set and it links to a standing individual of type ℓ . Note that the intensity measures $\{\nu_{B_\ell}\}_{\ell \in \{1, 2, \dots\}}$ do not partition in full the measure ν_A , since for any $n \geq 1$ such that $\mathbb{P}(A_i > n) > 0$ we have that

$$\sum_{\ell=1}^{\infty} \nu_{B_\ell}[\{i\} \times \{-1, \dots, -n\}]g_\ell < \sum_{\ell=1}^{\infty} \nu_A[\{i\} \times \{-1, \dots, -n\}]g_\ell = \nu_A[\{i\} \times \{-1, \dots, -n\}].$$

This is a consequence of the fact that not all coalescence times are in fact same-type coalescence times for some ℓ (for example, in Figure 2.2 the coalescence time $A_4 = 1$ of $(0, 4)$ and $(0, 5)$ is neither a same-type **1** nor a same-type **2** coalescence time).

2.2.2 Comparison of ancestral trees in two-type models

We next give an example of using the same-type coalescent times to investigate the effect of differences in offspring distribution on the distribution of types in the ancestral trees they produce. One question that motivated our work is the effect of different diversification rates for different types of individuals (phenotypes). We translate these questions into a discrete time defining an asymmetrical offspring distribution law.

Specifically, in a population with only two types of individuals, if the transition rates

of one type to the other are relatively high, while the other type never transitions into the first, this will be reflected in the distribution of types along the tree. In a discrete time process this is translated in the probability of a parent of the first type giving birth to individuals of the second type and vice versa. We consider this difference in the context of a two-type LF offspring distribution. In order to investigate only the effect on the distribution of types, we will make the distribution of the shape of the tree the same in both cases.

We consider the following two LF offspring distributions on $k = 2$ types of individuals. Let the parameters $\mathbf{g} = (g, 1 - g)$, $g \in [0, 1/2]$, $m > 0$ and $h_1 = 1 - h_0$, $h_0 \in [0, 1]$ be the same in both distributions, and for $p \in (0, 1)$ let

$$\mathbf{H}_s = h_1 \begin{pmatrix} p & 1 - p \\ 1 - p & p \end{pmatrix}, \quad \mathbf{H}_a = h_1 \begin{pmatrix} p & 1 - p \\ 0 & 1 \end{pmatrix},$$

be, respectively, associated with the symmetrical and the asymmetrical offspring distribution. In the symmetrical case parents of either type produce the first offspring of their own type and of the other type. In the asymmetrical case only a parent of type **1** will do that, while a parent of type **2** can only produce the first offspring of its own type. Since the number of offspring of each parent depends only on h_0 and m , the distribution of the ancestral tree with types erased will be the same in both cases. However, the distribution of the two types **1** and **2** are different, as can be seen in the following result.

Remark 2.3. *Note that we can assume without loss of generality that $g \in [0, 1/2]$, since in case $g \in [1/2, 1]$ we can simply reverse the notation of the two types. For $p = 1$ there is no asymmetry, nor are there offspring of different type than the parent - individuals in the whole tree are all of the same type. For $(g, p) = (1/2, 1/2)$ the symmetric case is special, and the offspring distribution is independent of the type of the parent, as discussed in Corollary 2.3.*

Proposition 2.5. *The distributions of coalescence times $(A_i)_{i \geq 1}$ are the same in both cases. The distribution of same-type coalescence times $(B_{1,i})_{i \geq 1}$ and $(B_{2,i})_{i \geq 1}$ satisfy the following stochastic dominance relations: $\forall p \in [0, 1]$,*

$$\mathbb{P}_{\mathbf{a}}(B_{1,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{1}) \geq \mathbb{P}_{\mathbf{s}}(B_{1,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{1}),$$

and

$$\mathbb{P}_{\mathbf{a}}(B_{2,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{2}) \leq \mathbb{P}_{\mathbf{s}}(B_{2,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{2}).$$

Also $\forall p \geq 1/2$ the two above inequalities are related by:

$$\mathbb{P}_{\mathbf{s}}(B_{1,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{1}) \geq \mathbb{P}_{\mathbf{s}}(B_{2,i} > n \mid \mathbf{A}_{0[0]} = \mathbf{2}).$$

Proof. Formulae (1.1) imply (after much arithmetic using Maple) that for $\mathbf{x} \in \{\mathbf{a}, \mathbf{s}\}$

$$m^{(n)} = m_{\mathbf{x}}^{(n)} = m \sum_{n'=0}^{n-1} h_1^{n'} (m+1)^{n'} = \frac{m(h_1^n (1+m)^n - 1)}{h_1(1+m) - 1},$$

$$\mathbf{g}_{\mathbf{s}}^{(n)} = \left((g - \frac{1}{2})G(2p-1) + \frac{1}{2}, -(g - \frac{1}{2})G(2p-1) + \frac{1}{2} \right), \quad \mathbf{g}_{\mathbf{a}}^{(n)} = (gG(p), -gG(p) + 1),$$

where G is a rather complicated polynomial

$$G(x) = \frac{h_1^{n-1}(h_1(1+m) - 1)}{h_1^n(1+m)^n - 1} \left[\sum_{n'=0}^{n-2} \left(\frac{h_1 m + 1}{h_1^{n-n'-1}} + \sum_{i=1}^{n-n'-2} (1+m)^i h_1^{i-n+n'+2} \right) x^{n'} + x^{n-1} \right],$$

which is increasing on $[0, 1]$ and satisfies $G(0) = 0, G(1) = 1$. From (2.10) we have the same formulae for the distribution of coalescent times in the two cases for $\mathbf{x} \in \{\mathbf{a}, \mathbf{s}\}$:

$$\mathbb{P}_{\mathbf{x}}(A_1 > n) = (1 + m^{(n)})^{-1},$$

and from (2.12) we get the following formulae for same-type coalescent times in the two cases:

$$\mathbb{P}_{\mathbf{x}}(B_{\ell,i} > n \mid \mathbf{A}_{0[0]} = \ell) = (1 + m^{(n)} g_{\mathbf{x}\ell}^{(n)})^{-1}, \quad \text{for } \ell \in \{\mathbf{1}, \mathbf{2}\},$$

where the two coordinates of vectors $\mathbf{g}_{\mathbf{x}}^{(n)} = (g_{\mathbf{x}\mathbf{1}}^{(n)}, g_{\mathbf{x}\mathbf{2}}^{(n)})$ for $\mathbf{x} \in \{\mathbf{a}, \mathbf{s}\}$ are given above.

We next prove that for $g \leq 1/2$ we have

$$1 + m^{(n)} g_{\mathbf{a}\mathbf{1}}^{(n)} \leq 1 + m^{(n)} g_{\mathbf{s}\mathbf{1}}^{(n)} \Leftrightarrow g_{\mathbf{a}\mathbf{1}}^{(n)} \leq g_{\mathbf{s}\mathbf{1}}^{(n)}, \quad 1 + m^{(n)} g_{\mathbf{s}\mathbf{2}}^{(n)} \leq 1 + m^{(n)} g_{\mathbf{a}\mathbf{2}}^{(n)} \Leftrightarrow g_{\mathbf{s}\mathbf{2}}^{(n)} \leq g_{\mathbf{a}\mathbf{2}}^{(n)}.$$

Both of these inequalities are equivalent to

$$gG(p) + \left(\frac{1}{2} - g\right)G(2p-1) - \frac{1}{2} \leq 0,$$

which holds since all multiplying coefficients of the polynomial $G(x)$ are nonnegative, so G is increasing and both $G(p), G(2p-1) \leq G(1) = 1$.

For the last comparison we need to show that for $g \leq 1/2 \leq p$

$$1 + m^{(n)} g_{\mathbf{s}\mathbf{1}}^{(n)} \leq 1 + m^{(n)} g_{\mathbf{s}\mathbf{2}}^{(n)} \Leftrightarrow g_{\mathbf{s}\mathbf{1}}^{(n)} \leq g_{\mathbf{s}\mathbf{2}}^{(n)},$$

which is equivalent to

$$2\left(\frac{1}{2} - g\right)G(2p-1) \geq 0,$$

and holds as long as $p \geq 1/2$ so that the polynomial $G(x)$ is evaluated on $x \geq 0$. 

We see that the consequence of asymmetry (irrespective of the value of p) is that the same-type coalescence times are typically shorter for type **2** than in the symmetrical case, while they are longer for type **1**. This intuitively make sense, since subtrees of a type **2** can only have a first offspring of type **2** and the probability of having subsequent

offsprings of type **1** is less than $1/2$ ($g \leq 1/2$).

We can also see the effect that the ‘strength’ p of not transitioning to a different type plays in the symmetric case. When $p \geq 1/2$ having the same type offspring as parent is more likely. In the symmetric case $g \leq 1/2$ further implies that type **1** is overall less frequent than type **2** in the tree. Hence, one would expect that the same-type coalescence times are typically going to be longer for type **1** than for type **2**.

Chapter 3

Distribution of k -type cherries

The contents of this chapter are part of ongoing research, and are likely to be expanded before the final submission of this thesis. We start the chapter by presenting an extension of some results by McKenzie & Steel [34]. We extend the concept of Yule trees (Section (1.2.3)) to the multi-type setting. We find exact and asymptotic expressions for the mean number of cherries in neutral two-type trees under a natural classification by types. To obtain asymptotic expressions we use an extended Pólya urn representation which implies that this distribution converges to a multi-dimensional normal with certain parameters. The main difficulty in using Pólya urns for more general settings is that it is often not possible to obtain a usable expression for the eigenvalues and eigenvectors of the model's matrix. Thus in order to obtain more general results for k -type Yule trees, we use a continuous model in which the discrete Yule model is embedded. This approach is presented at the end of this chapter (Section 3.2).

3.1 Distribution of two-type cherries in neutral two-type binary trees

We study the distribution of the number of cherries of different types in a *neutral two-type* Yule tree. This is a tree generated by the Yule process as described in Section 1.2.3, with nodes whose types (either 1 or 2) are chosen on every iteration with a distribution which depends on the parent type. More formally, the process starts with a single node of fixed or random type 1 or 2. This node is assumed to be a leaf. At each step, a leaf is chosen uniformly at random. If this leaf is of type i , then two leaves are appended to it. These two leaves are assigned types $\{j_1, j_2\} \subseteq \{1, 2\}$ with probability $q_i^{j_1 j_2}$ ($j_1 \leq j_2$).

In this and other sections we may use the word *time* or *generation* to refer to a given step in the Yule model. We may also refer to the leaves at time n as *individuals alive at time n* , and a node's children as its *offspring*. We assume that two-type Yule trees start at time 1 with one individual, so that the number of individuals at time n is equal to n . We may also disregard the concept of time and refer to the n -th iteration of the Yule model as a *Yule tree on n leaves*.

We are interested in finding means and variances of a few statistics for neutral two-type Yule trees. In particular the number $N_i(n)$ of type- i individuals at time n and the number of leaves of each type ij_1j_2 , where $j_1 \leq j_2$ are the types of the cherry's leaves and i is the type of their common parent. Figure 3.1 shows all 6 types of cherries in a two-type tree, as well as the 4 types of *pendant edges*, which we define as leaves which are not part of a cherry.

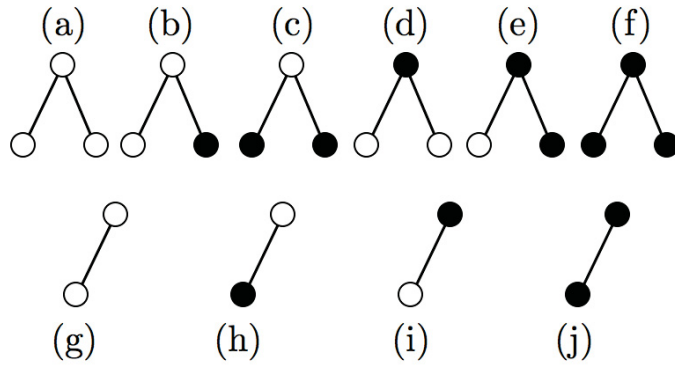


Figure 3.1: Two-type cherries and pendant edges. Individuals of type 1 are denoted by a blank circle, and individuals of type 2 are denoted by a full circle. There are 6 different types of cherry, namely (a) (type 111), (b) (type 112), (c) (type 122), (d) (type 211), (e) (type 212) and (f) (type 222), and there are 4 different types of pendant edges, namely (g) (type 11), (h) (type 12), (i) (type 21) and (j) (type 22).

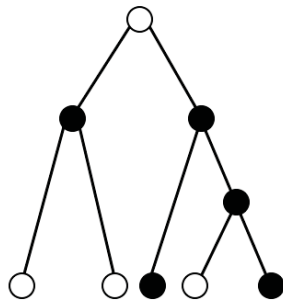


Figure 3.2: Example of a two-type tree. A two-type tree on five leaves with two cherries of types 211 and 212, and one pendant edge of type 22.

We obtain explicit formulae for the means and variances of these random variables under certain conditions, and then we provide some results on their asymptotic behaviour using Pólya urn models, under the condition that $q_i^{ii} \neq 1$ for all $i \in \{1, 2\}$.

3.1.1 Probability distribution of two-type cherries

To obtain most of the results that follow we have performed calculations involving generating functions. In some cases we have used the algebra system Maple to aid us in these calculations. Most of these results have also been verified for known particular cases, or matched against results obtained through different methods.

Proposition 3.1. *Let ν_n denote the mean number of leaves of type 1 on a random neutral two-type binary tree on n leaves. Assume the Yule distribution with parameters $q_i^{j_1 j_2}$ ($i, j_1, j_2 \in \{1, 2\}$) satisfying $c_1 - c_2 \notin \{-2, -1, 2\}$ for $c_1 := 2q_1^{11} + q_1^{12}$ and $c_2 := 2q_2^{11} + q_2^{12}$. Then for $n \geq 2$;*

$$\nu_n = \frac{c_2 n}{2 - c_1 + c_2} - \frac{(2c_2 + \nu_2 c_1 - \nu_2 c_2 - 2\nu_2)\Gamma(n - 1 + c_1 - c_2)}{(2 - c_1 + c_2)\Gamma(c_1 - c_2 + 1)\Gamma(n)},$$

where $\nu_2 = \begin{cases} 2q_1^{11} + q_1^{12} & \text{if root type is 1,} \\ 2q_2^{11} + q_2^{12} & \text{if root type is 2,} \end{cases}$ and $\Gamma(n)$ represents the classical gamma function of n .

Proof. The following recurrence is a direct result of the definition of a neutral two-type binary tree:

$$\begin{aligned} \mathbb{P}(N_1(n) = n_1) &= \left(\frac{n_1 q_1^{12}}{n-1} + \frac{(n-n_1-1)q_2^{22}}{n-1} \right) \mathbb{P}(N_1(n-1) = n_1) \\ &+ \left(\frac{(n_1-1)q_1^{11}}{n-1} + \frac{(n-n_1)q_2^{12}}{n-1} \right) \mathbb{P}(N_1(n-1) = n_1 - 1) \\ &+ \left(\frac{(n-n_1+1)q_2^{11}}{n-1} \right) \mathbb{P}(N_1(n-1) = n_1 - 2) \\ &+ \left(\frac{(n_1+1)q_1^{22}}{n-1} \right) \mathbb{P}(N_1(n-1) = n_1 + 1). \end{aligned}$$

This recurrence yields a differential equation on the generating functions:

$$G_n(x) = \sum_{i \geq 0} \mathbb{P}(N_1(n) = n_1) x^{n_1}.$$

From this differential equation we obtain a recurrence for the mean ν_n which yields the claimed result. ✪

Remark 3.1. *From Proposition 3.1 above we immediately obtain the mean number of leaves of type 2 simply by swapping types 1 and 2 on each index of the $q_i^{j_1 j_2}$. The conditions $c_1 - c_2 \neq 2$ and $c_1 - c_2 \neq -2$ rule out some trivial cases such as $q_1^{11} = q_2^{22} = 1$ and $q_2^{11} = q_1^{22} = 1$. Observe also that $c_1 - c_2 = 2(q_1^{11} + q_2^{22}) + (q_1^{12} + q_2^{12}) - 2$ remains invariant upon swapping types 1 and 2, and so the condition to obtain each mean is the same for both types.*

Let us denote by $K_i^{j_1 j_2}(n)$ ($j_1 \leq j_2$) the number of cherries of type $ij_1 j_2$ (with parent of type i and children of types j_1, j_2) in a neutral two-type binary tree at generation n . We study these quantities for $i = 1$ (as any result for $i = 2$ is analogous) using their bivariate probability generating function (pgf) with the number of leaves of type 1:

$$F_n^{1j_1 j_2}(x, y) = \sum_{n_1 \geq 0, l \geq 0} \mathbb{P}(N_1(n) = n_1, K_1^{j_1 j_2}(n) = l) x^{n_1} y^l.$$

Proposition 3.2. *Let $\mu_{1,n}$, $\mu_{2,n}$, $\mu_{3,n}$ be the mean number of cherries of type 111, 112, 122, respectively, for a neutral two-type binary tree on n leaves. Under the Yule distribution with the same conditions of Proposition 3.1, for $n \geq 3$:*

$$\begin{aligned} \mu_{1,n} &= \frac{3(2 - c_1 + c_2)(2\mu_{1,3} - q_1^{11}\nu_2) + n(n-1)(n-2)q_1^{11}c_2}{3(2 - c_1 + c_2)(n-1)(n-2)} - q_1^{11}C_n, \\ \mu_{2,n} &= \frac{3(2 - c_1 + c_2)(2\mu_{2,3} - q_1^{12}\nu_2) + n(n-1)(n-2)q_1^{12}c_2}{3(2 - c_1 + c_2)(n-1)(n-2)} - q_1^{12}C_n, \\ \mu_{3,n} &= \frac{3(2 - c_1 + c_2)(2\mu_{3,3} - q_1^{22}\nu_2) + n(n-1)(n-2)q_1^{22}c_2}{3(2 - c_1 + c_2)(n-1)(n-2)} - q_1^{22}C_n, \end{aligned}$$

where ν_2, c_1, c_2 are as in Proposition 3.1 and

$$C_n := \frac{(2c_2 + \nu_2 c_1 - \nu_2 c_2 - 2\nu_2)\Gamma(n-1+c_1-c_2)}{(2-c_1+c_2)\Gamma(c_1-c_2+2)\Gamma(n)},$$

$$\mu_{1,3} = \begin{cases} (q_1^{11})^2 + q_1^{11}q_1^{12}/2 & \text{if root type is 1} \\ q_1^{11}q_2^{11} + q_1^{11}q_2^{12}/2 & \text{if root type is 2} \end{cases}, \quad \mu_{2,3} = \begin{cases} (q_1^{12})^2/2 + q_1^{11}q_1^{12} & \text{if root type is 1} \\ q_2^{11}q_1^{12} + q_2^{12}q_1^{12}/2 & \text{if root type is 2} \end{cases},$$

$$\mu_{3,3} = \begin{cases} q_1^{11}q_1^{22} + q_1^{12}q_1^{22}/2 & \text{if root type is 1} \\ q_2^{12}q_1^{22}/2 + q_2^{11}q_1^{22} & \text{if root type is 2} \end{cases}.$$

Proof. For simplicity denote

$$f_n^{ij_1j_2}(n_1, l) := \mathbb{P}(N_1(n) = n_1, K_i^{j_1j_2}(n) = l).$$

The formulae claimed are obtained by using the recurrences below to derive recurrences for the generating functions $F_n^{ij_1j_2}(x, y)$. For cherries of type 111 we have:

$$\begin{aligned} f_n^{111}(n_1, l) &= \left(\frac{2lq_1^{11}}{n-1} + \frac{(n-n_1)q_2^{12}}{n-1} \right) f_{n-1}^{111}(n_1-1, l) \\ &+ \left(\frac{(n_1-2l)q_1^{12}}{n-1} + \frac{(n-n_1-1)q_2^{22}}{n-1} \right) f_{n-1}^{111}(n_1, l) \\ &+ \left(\frac{(n-n_1+1)q_2^{11}}{n-1} \right) f_{n-1}^{111}(n_1-2, l) \\ &+ \left(\frac{(n_1+1-2l)q_1^{22}}{n-1} \right) f_{n-1}^{111}(n_1+1, l) \\ &+ \left(\frac{(n_1-1-2(l-1))q_1^{11}}{n-1} \right) f_{n-1}^{111}(n_1-1, l-1) \\ &+ \left(\frac{2(l+1)q_1^{22}}{n-1} \right) f_{n-1}^{111}(n_1+1, l+1) \\ &+ \left(\frac{2(l+1)q_1^{12}}{n-1} \right) f_{n-1}^{111}(n_1, l+1). \end{aligned}$$

Similarly, for cherries of type 112:

$$\begin{aligned}
f_n^{112}(n_1, l) &= \left(\frac{lq_1^{12}}{n-1} + \frac{(n-n_1-l-1)q_2^{22}}{n-1} \right) f_{n-1}^{112}(n_1, l) \\
&+ \left(\frac{(n_1-l-1)q_1^{11}}{n-1} + \frac{(n-n_1-l)q_2^{12}}{n-1} \right) f_{n-1}^{112}(n_1-1, l) \\
&+ \left(\frac{(n-n_1-l+1)q_2^{11}}{n-1} \right) f_{n-1}^{112}(n_1-2, l) \\
&+ \left(\frac{(l+1)q_2^{11}}{n-1} \right) f_{n-1}^{112}(n_1-2, l+1) \\
&+ \left(\frac{(l+1)q_2^{12}}{n-1} + \frac{(l+1)q_1^{11}}{n-1} \right) f_{n-1}^{112}(n_1-1, l+1) \\
&+ \left(\frac{(l+1)q_2^{22}}{n-1} \right) f_{n-1}^{112}(n_1, l+1) \\
&+ \left(\frac{(n_1-l+1)q_1^{12}}{n-1} \right) f_{n-1}^{112}(n_1, l-1) \\
&+ \left(\frac{(n_1-l+1)q_1^{22}}{n-1} \right) f_{n-1}^{112}(n_1+1, l) \\
&+ \left(\frac{(l+1)q_1^{22}}{n-1} \right) f_{n-1}^{112}(n_1+1, l+1).
\end{aligned}$$

And for cherries of type 122:

$$\begin{aligned}
f_n^{122}(n_1, l) &= \left(\frac{(n-n_1-2l)q_2^{12}}{n-1} + \frac{(n_1-1)q_1^{11}}{n-1} \right) f_{n-1}^{122}(n_1-1, l) \\
&+ \left(\frac{(n-n_1-2l-1)q_2^{22}}{n-1} + \frac{n_1q_1^{12}}{n-1} \right) f_{n-1}^{122}(n_1, l) \\
&+ \left(\frac{2(l+1)q_2^{11}}{n-1} \right) f_{n-1}^{122}(n_1-2, l+1) \\
&+ \left(\frac{2(l+1)q_2^{12}}{n-1} \right) f_{n-1}^{122}(n_1-1, l+1) \\
&+ \left(\frac{2(l+1)q_2^{22}}{n-1} \right) f_{n-1}^{122}(n_1, l+1) \\
&+ \left(\frac{(n_1+1)q_1^{22}}{n-1} \right) f_{n-1}^{122}(n_1+1, l-1) \\
&+ \left(\frac{(n-n_1-2l+1)q_2^{11}}{n-1} \right) f_{n-1}^{122}(n_1-2, l).
\end{aligned}$$

All these recurrences can be written in terms of the corresponding bivariate pgfs by summing over n_1 and l . Differentiating these recurrences and evaluating at $x = y = 1$ yields recurrences for the means $\mu_{i,n}$, which may be solved to obtain the desired expressions. \star

Remark 3.2. Notice that, as expected, the sum of all the means from Proposition 3.2, together with the ones for parents of type 2 (which are obtained by swapping types 1 and 2 on each $q_i^{j_1 j_2}$) add up to $n/3$, the mean number of cherries in a single-type Yule model (see Theorem 1.7).

Particular cases for the probabilities $q_i^{j_1 j_2}$ are often useful in applications. The following corollary of Proposition 3.2 lists the means for some important particular cases.

Corollary 3.1. Consider a neutral two-type binary tree under the conditions of Proposition 3.2 above. Some particular values for the means $\mu_{i,n}$ ($i = 1, 2, 3$) are given by:

Cases	Mean
Extreme #1: $q_1^{11} = 1, q_1^{12} = q_1^{22} = 0$	$\mu_{1,n} = \frac{n}{3}, \mu_{2,n} = \mu_{3,n} = 0.$
Extreme #2: $q_1^{11} = q_2^{11} = 1$	$\mu_{1,n} = \frac{n}{3}, \mu_{2,n} = \mu_{3,n} = 0.$
Extreme #3: $q_1^{12} = q_2^{12} = 1$	$\mu_{2,n} = \frac{n}{6}, \mu_{1,n} = \mu_{3,n} = 0.$
Neutral: $q_1^{11} = q_2^{11}, q_1^{12} = q_2^{12}, q_1^{22} = q_2^{22}$	$\mu_{1,n} = \frac{nq_1^{11}c_1}{6}, \mu_{2,n} = \frac{nq_1^{12}c_1}{6}, \mu_{3,n} = \frac{nq_1^{22}c_1}{6}.$

3.1.2 Asymptotics on two-type cherries

More complicated calculations yield formulae for the variances of the number of cherries of each type. This is achieved by taking the second derivatives of the recurrences for the pgfs $F_n^{111}(x, y)$, $F_n^{112}(x, y)$ and $F_n^{122}(x, y)$, evaluating them at $x = y = 1$, and solving the

resulting system of recurrences. We use Maple to obtain these formulae, and while they turn out to be quite long and cluttered, a particularly interesting result is their order with respect to n .

Proposition 3.3. *Let $\sigma_{1,n}^2, \sigma_{2,n}^2, \sigma_{3,n}^2$ be the variance for the number of cherries of type 111, 112, 122, respectively, for a two-type rooted binary tree on n leaves. Assume that $c_1 - c_2 \notin \{3/2, 1, -1, 2, -2, 0\}$, where c_1, c_2 are as in Proposition 3.1. Under the Yule distribution with the conditions of Proposition 3.1, for $n \geq 5$, each of the variances $\sigma_{1,n}^2, \sigma_{2,n}^2, \sigma_{3,n}^2$ is the sum of three terms with orders, given by*

$$\sigma_{1,n}^2 \sim O(n) + O(n^{c_1 - c_2 - 1}) + O(n^{2(c_1 - c_2 - 1)}),$$

$$\sigma_{2,n}^2 \sim O(n) + O(n^{c_1 - c_2 - 1}) + O(n^{2(c_1 - c_2 - 1)}),$$

$$\sigma_{3,n}^2 \sim O(n) + O(n^{c_1 - c_2 - 1}) + O(n^{2(c_1 - c_2 - 1)}).$$

Remark 3.3. *Recall that the conditions $c_1 - c_2 \neq 2$ and $c_1 - c_2 \neq -2$ rule out some trivial cases such as $q_1^{11} = q_2^{22} = 1$ and $q_2^{11} = q_1^{22} = 1$.*

Proposition 3.4. *Some particular values for the variances of Proposition 3.3 are given by:*

<i>Cases</i>	<i>Variance</i>
<i>Extreme case 1:</i> $q_1^{11} = 1, q_1^{12} = q_1^{22} = 0$	$\sigma_{n,1}^2 = \frac{2n}{45}, \sigma_{n,2}^2 = \sigma_{n,3}^2 = 0.$
<i>Extreme case 2:</i> $q_1^{11} = q_2^{11} = 1$	$\sigma_{n,1}^2 = \frac{2n}{45}, \sigma_{n,2}^2 = \sigma_{n,3}^2 = 0.$
<i>Extreme case 3:</i> $q_1^{12} = q_2^{12} = 1$	$\sigma_{n,2}^2 = \frac{7n}{90}, \sigma_{n,1}^2 = \sigma_{n,3}^2 = 0.$
<i>Neutral case:</i> $q_1^{11} = q_2^{11}, q_1^{12} = q_2^{12}, q_1^{22} = q_2^{22}$	$\sigma_{n,1}^2 = \frac{nq_1^{11}(6(q_1^{11})^2 + 15c_1 - 8q_1^{11}c_1^2)}{90},$ $\sigma_{n,2}^2 = \frac{nq_1^{12}(6q_1^{11}q_1^{12} + 15c_1 - 8q_1^{12}c_1^2)}{90},$ $\sigma_{n,3}^2 = \frac{nq_1^{22}(6q_1^{11}q_1^{22} + 15c_1 - 8q_1^{22}c_1^2)}{90},$

Notice that, as one would expect, the mean and variance in the first two cases coincide with the mean and variance of (single-type) Yule model in McKenzie & Steel [34]. Analogous results are obtained for cherries of types 211, 212 and 222, simply by exchanging indices 1 and 2 in the corresponding formulae.

3.1.3 Pólya urn representation

The Yule process on neutral two-type binary trees can be viewed as a Pólya model (Section 1.3). In this model the balls are all the cherries and pendant edges of the graph at time n . There are thus 10 types of balls, as shown in Figure 3.1. The types corresponding to cherries have weight $a_i = 2$, and those corresponding to pendant edges have weight $a_i = 1$. This is because it is twice as likely to choose a cherry than a pendant edge when a leaf is picked uniformly at random. The generating 10×10 matrix \mathbf{A} of this Pólya urn

process (see Section 1.3) is given by

$$\begin{bmatrix} -2(q_1^{12} + q_1^{22}) & q_1^{11} & 0 & 0 & q_1^{11} & 2q_1^{11} & q_1^{11} & 0 & 0 & q_1^{11} \\ 2q_1^{12} & -(2 - q_1^{12}) & 0 & 0 & q_1^{12} & 2q_1^{12} & q_1^{12} & 0 & 0 & q_1^{12} \\ 2q_1^{22} & q_1^{22} & -2 & 0 & q_1^{22} & 2q_1^{22} & q_1^{22} & 0 & 0 & q_1^{22} \\ 0 & q_2^{22} & 2q_2^{22} & -2(q_2^{11} + q_2^{12}) & q_2^{22} & 0 & 0 & q_2^{22} & q_2^{22} & 0 \\ 0 & q_2^{12} & 2q_2^{12} & 2q_2^{12} & -(2 - q_2^{12}) & 0 & 0 & q_2^{12} & q_2^{12} & 0 \\ 0 & q_2^{11} & 2q_2^{11} & 2q_2^{11} & q_2^{11} & -2 & 0 & q_2^{11} & q_2^{11} & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Here the 10 ball types have been ordered as follows: 111, 112, 122, 222, 212, 211, 11, 12, 22, 21. In this process, the entries of the vector \mathbf{X}_n count the number of cherries and pendant edges of each of these types in a neutral two-type binary tree on n leaves.

Using Maple we have obtained the eigenvalues of \mathbf{A} , namely $\lambda_1 = 1$, $\lambda_2 = c_1 - c_2 - 1 = 2q_1^{11} + q_1^{12} - 2q_2^{11} - q_2^{12} - 1 \leq \lambda_1$, $\lambda_3 = \lambda_4 = -1$ and $\lambda_5 = \dots = \lambda_{10} = -2$. The maximum

real eigenvalue is $\lambda_1 = 1$, whose normalized right and left eigenvectors are:

$$\mathbf{v}_1 = \frac{1}{3(2 - c_1 + c_2)} \begin{bmatrix} q_1^{11} c_2 \\ q_1^{12} c_2 \\ q_1^{22} c_2 \\ q_2^{22} (2 - c_1) \\ q_2^{12} (2 - c_1) \\ q_2^{11} (2 - c_1) \\ c_1 c_2 / 2 \\ (2 - c_1) c_2 / 2 \\ (2 - c_1)(2 - c_2) / 2 \\ (2 - c_1) c_2 / 2 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Also, $\lambda_2 < 1$ is a simple eigenvalue. Its normalized right and left eigenvectors are given by:

$$\mathbf{v}_2 = \frac{c_2}{(2 - c_1 + c_2)(c_2 - c_1 - 1)} \begin{bmatrix} q_1^{11} \\ q_1^{12} \\ q_1^{22} \\ q_2^{22} \\ q_2^{12} \\ q_2^{11} \\ c_1 / (c_1 - c_2) \\ (2 - c_1) / (c_1 - c_2) \\ -(2 - c_1) / (c_1 - c_2) \\ c_2 / (c_1 - c_2) \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{c_2} \begin{bmatrix} -2(2 - c_1) \\ c_1 + c_2 - 2 \\ 2 \\ 2 \\ c_1 + c_2 - 2 \\ -2(2 - c_1) \\ c_1 - 2 \\ 1 \\ 1 \\ c_1 - 2 \end{bmatrix}.$$

If we assume that $q_i^{ii} \neq 1$ for $i, j_1, j_2 \in \{1, 2\}$, then it is possible to eventually have

any type ij_1j_2 of cherry with $q_i^{j_1j_2} \neq 0$ or pendant edge ij with $q_i^j \neq 0$, starting from any other such type. Hence under this assumption, conditions $(\alpha_1) - (\alpha_5)$ of Section 1.3 are satisfied, which allows us to apply Theorems 1.8 and 1.9 to obtain the following results.

Corollary 3.2. *Let \mathbf{X}_n be the Pólya urn model corresponding to a neutral two-type binary tree with generating probabilities $q_i^{j_1j_2}$ ($i, j_1, j_2 \in \{1, 2\}$) such that $q_i^{ii} \neq 1$. That is, \mathbf{X}_n is a 10-dimensional random vector whose entries count the number of cherries and pendant edges of types 111, 112, 122, 222, 212, 211, 11, 12, 22, 21 in a neutral two-type binary tree on n leaves. Then*

$$\frac{\mathbf{X}_n}{n} \rightarrow \mathbf{v}_1 = \frac{1}{3(2 - c_1 + c_2)} \begin{bmatrix} q_1^{11}c_2 \\ q_1^{12}c_2 \\ q_1^{22}c_2 \\ q_2^{11}(2 - c_1) \\ q_2^{12}(2 - c_1) \\ q_2^{22}(2 - c_1) \\ (c_1c_2)/2 \\ (2 - c_1)c_2/2 \\ (2 - c_1)(2 - c_2)/2 \\ (2 - c_1)c_2/2 \end{bmatrix} \quad a.s. \text{ as } n \rightarrow \infty.$$

In particular

$$\lim_{n \rightarrow \infty} \frac{X_{n,i}}{n} = \lim_{n \rightarrow \infty} \frac{\mu_{i,n}}{n} \quad \text{for } i = 1, 2, 3,$$

where the $\mu_{n,i}$ are as in Proposition 3.2.

Remark 3.4. *Notice that if the entries $v_{1,1}, v_{1,2}, v_{1,3}, \dots, v_{1,10}$ of \mathbf{v}_1 are known, then we*

can recover all probabilities $q_i^{j^l}$ ($i, j, l \in \{1, 2\}$, $j \leq l$). Indeed

$$q_1^{11} = \frac{v_{1,1}}{v_{1,1} + v_{1,2} + v_{1,3}}, q_1^{12} = \frac{v_{1,2}}{v_{1,1} + v_{1,2} + v_{1,3}}, q_1^{22} = \frac{v_{1,3}}{v_{1,1} + v_{1,2} + v_{1,3}},$$

$$q_2^{22} = \frac{v_{1,4}}{v_{1,4} + v_{1,5} + v_{1,6}}, q_1^{12} = \frac{v_{1,5}}{v_{1,4} + v_{1,5} + v_{1,6}}, q_1^{22} = \frac{v_{1,6}}{v_{1,4} + v_{1,5} + v_{1,6}}.$$

The denominators above are non-zero, since $c_2 = 0$ would imply $q_2^{22} = 1$, and $c_1 = 2$ would imply $q_1^{11} = 1$.

Corollary 3.3. *Let \mathbf{X}_n be the Pólya urn process from Corollary 3.3 above. Suppose that $c_1 - c_2 = 3/2$. Then as $n \rightarrow \infty$*

$$\frac{\mathbf{X}_n - n\mathbf{v}_1}{n \ln(n)} \rightarrow N(0, \Sigma) \text{ in distribution,}$$

where the covariance matrix is given by

$$\Sigma = C \begin{bmatrix} (q_1^{11})^2 & q_1^{11}q_1^{12} & q_1^{11}q_1^{22} & -q_1^{11}q_2^{22} & -q_1^{11}q_2^{12} & -q_1^{11}q_2^{11} & * & * & * & * \\ \checkmark & (q_1^{12})^2 & q_1^{12}q_1^{22} & -q_1^{12}q_2^{22} & -q_1^{12}q_2^{12} & -q_1^{12}q_2^{11} & * & * & * & * \\ \checkmark & \checkmark & (q_1^{22})^2 & -q_1^{22}q_2^{22} & -q_1^{22}q_2^{12} & -q_1^{22}q_2^{11} & * & * & * & * \\ \checkmark & \checkmark & \checkmark & (q_2^{22})^2 & q_2^{22}q_2^{12} & q_2^{22}q_2^{11} & * & * & * & * \\ \checkmark & \checkmark & \checkmark & \checkmark & (q_2^{12})^2 & q_2^{12}q_2^{11} & * & * & * & * \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & (q_2^{11})^2 & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * \end{bmatrix}$$

where

$$C := -\frac{8}{25}(9 + 12(q_1^{11})^2 + 2q_2^{11}q_1^{12} + 4q_1^{11}q_2^{11} + 4(q_1^{12})^2 + 14q_1^{11}q_1^{12} - 4q_2^{11} - 12q_1^{12} - 21q_1^{11})$$

The entries * represent the covariances between the number of cherries and the number of pendant edges, which we have purposefully omitted, and the entries marked with a ✓ are omitted because of the symmetry of the covariance matrix.

Example 3.1. In Corollary 3.3, we can take for example $q_1^{12} + q_2^{12} = 1/2$ and $q_2^{11} = q_1^{22} = 0$ and observe that indeed $c_1 - c_2 = 3/2$.

Corollary 3.4. Let \mathbf{X}_n be the Pólya urn process from Corollary 3.3 above. Suppose that $c_1 - c_2 < 3/2$. We have as $n \rightarrow \infty$

$$\frac{\mathbf{X}_n - n\mathbf{v}_1}{\sqrt{n}} \rightarrow N(0, \Sigma) \text{ in distribution,}$$

for a covariance matrix Σ with constant entries.

The entries of the covariance matrix of corollary 3.4 above have been calculated explicitly using Maple.

Remark 3.5. Observe that this result is consistent with Proposition 3.3, since when $(c_1 - c_2) < 3/2$, we have $(c_1 - c_2 - 1) < 1/2$ and $2(c_1 - c_2 - 1) < 1$. So the orders of the variances are equal to n .

Remark 3.6. Notice that of all the particular cases considered in Corollary 3.1 and Proposition 3.4, only the neutral case could satisfy $q_i^{j_1 j_2} > 0$ for all i, j_1, j_2 . In the neutral case $c_1 = c_2$, and so Corollary 3.4 holds.

3.1.4 Robust solvability of the reconstruction problem for neutral multi-type Yule trees

In this section we use the results of Janson & Mossel [21] to study the *solvability* of the neutral multi-type Yule tree under the condition that types of siblings are independent amongst each other. The concept of solvability means being able to obtain some information about the root type of the tree, from the types of its tips (leaves).

Recall that in the neutral k -type Yule model we start with a single node (the root) of type in $\{1, \dots, k\}$, and on each step, a leaf which is chosen uniformly at random, has two children of types $j_1 \leq j_2$ with probability $q_i^{j_1 j_2}$. This process continues until the tree has n leaves. Notice that this construction is equivalent to building a single type Yule tree on n leaves, and then *colouring* (or *propagating* the types of) the nodes successively, where $q_i^{j_1 j_2}$ is the probability that the children of a node of type i are j_1 and j_2 for $i, j_1, j_2 \in \{1, 2, \dots, k\}$. A particular model of propagation appears when the type of each child is chosen independently (from its siblings) according to a $k \times k$ stochastic transition matrix $\mathbf{S} = (s_{ij})_{1 \leq i, j \leq k}$ corresponding to an ergodic (irreducible and positive-recurrent) Markov chain. More precisely, s_{ij} is the conditional probability that a node is of type j , given that its parent is of type i . The relations below follow;

$$q_i^{j_1 j_2} = \begin{cases} 2s_{ij_1}s_{ij_2}, & j_1 < j_2. \\ (s_{ij_1})^2, & j_1 = j_2. \end{cases}$$

Janson & Mossel [21] study the solvability of general (not necessarily binary) trees under this propagation rule. More specifically they study robust solvability, which considers the leaves of the tree under certain random noise as we explain next. Let \mathcal{T} be a finite rooted binary tree with n leaves. Let $\alpha_\ell^{\mathcal{T}}$ denote the probability distribution of the types of the nodes of \mathcal{T} under the propagation rule above, conditioned on the root being

of type $\ell \in \{1, \dots, k\}$. Let $\mathbf{N} = (N_{ij})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k'}}$ be a $k \times k'$ stochastic matrix for some k' . Suppose that every tip of \mathcal{T} is “perturbed” using \mathbf{N} , i.e. with probability N_{ij} , a node of type $i \in \{1, \dots, k\}$ becomes of type $j \in \{1, \dots, k'\}$. Let $\alpha_\ell^\mathcal{T}[\mathbf{N}]$ denote the probability distribution of the node types in the tree \mathcal{T} under this noise. We are interested in the following cases for \mathbf{N} :

- For $n_1 \geq 0$, set $\mathbf{N} = \mathbf{S}^{n_1}$, and write $\alpha_\ell^\mathcal{T}[n_1] = \alpha_\ell^\mathcal{T}[\mathbf{S}^{n_1}]$.
- Given a distribution π on $\{1, 2, \dots, k\}$, and a real number $\varepsilon \in [0, 1]$, consider $N_{ij} = (1 - \varepsilon)\delta_{i=j} + \varepsilon\pi_j$. In this case we write $\alpha_\ell^\mathcal{T}[\pi, \varepsilon] = \alpha_\ell^\mathcal{T}[\mathbf{N}]$.
- Given $0 \leq \varepsilon \leq 1$, let \mathbf{N} be the $k \times (k + 1)$ matrix with

$$N_{ij} = \begin{cases} 1 - \varepsilon, & i = j. \\ \varepsilon, & j = k + 1. \\ 0, & \text{else.} \end{cases}$$

In this case we write $\alpha_\ell^\mathcal{T}[\varepsilon] = \alpha_\ell^\mathcal{T}[\mathbf{N}]$.

Denote by D_V the total variation distance between two distributions α and α' defined on the same space Ω ,

$$D_V(\alpha, \alpha') = \frac{1}{2} \sum_{\omega \in \Omega} |\alpha(\omega) - \alpha'(\omega)|.$$

The original formulation of the following theorem involves graph theoretical notions and general infinite trees. Here we write in an equivalent specialized form in terms of binary trees.

Theorem 3.1. (*Janson & Mossel [21]*) *Let \mathbf{S} be a stochastic matrix of an ergodic Markov chain such that $2|\lambda_2|^2 > 1$, where λ_2 is the eigenvalue of \mathbf{S} having the second largest real part. Recall that this matrix defines distributions $\alpha_i^\mathcal{T}$ on the tips of each binary tree \mathcal{T} conditioned on the root being of type $i = 1, \dots, k$.*

- For each $n_1 \geq 0$, there exist $i, j \in \{1, 2, \dots, k\}$ so that

$$\inf_{\mathcal{T}} D_V(\alpha_i^{\mathcal{T}}[n_1], \alpha_j^{\mathcal{T}}[n_1]) > 0,$$

where the inf is over all finite rooted binary trees.

- For each nondegenerate distribution π , and each $\varepsilon \in [0, 1)$, there exist types $i, j \in \{1, 2, \dots, k\}$ so that

$$\inf_{\mathcal{T}} D_V(\alpha_i^{\mathcal{T}}[\pi, \varepsilon], \alpha_j^{\mathcal{T}}[\pi, \varepsilon]) > 0.$$

- For each $\varepsilon \in [0, 1)$, there exist $i, j \in \{1, 2, \dots, k\}$ so that

$$\inf_{\mathcal{T}} D_V(\alpha_i^{\mathcal{T}}[\varepsilon], \alpha_j^{\mathcal{T}}[\varepsilon]) > 0.$$

This theorem means that, depending on the value of the second eigenvalue of \mathbf{S} , the distribution of leave types on a neutral multi-type Yule tree, as it grows to infinity, with a propagation rule as above, may provide some information on the type of the root.

Remark 3.7. *In the two type case, the condition $2|\lambda_2|^2 > 1$ simply becomes $2(s_{11} + s_{22} - 1)^2 > 1$. Furthermore, if the rates s_{11}, s_{22} are unknown, but the asymptotic proportion of cherries and pendant edges are known, this condition may be written in terms of these asymptotic statistics. Using the formulae and notation of Corollary 3.2 and Remark 3.4, we obtain the following result for the two-type case.*

Corollary 3.5. *If $2 \left(\sqrt{v_{1,1}/(v_{1,1} + v_{1,2} + v_{1,3})} + \sqrt{v_{1,4}/(v_{1,4} + v_{1,5} + v_{1,6})} - 1 \right)^2 > 1$ in a particular two-type Yule model, then the results of Theorem 3.1 hold.*

The previous Corollary suggests that in the two-type scenario the asymptotic proportion of the number of cherries and pendant edges helps not only in getting information

about the transition probabilities of the process, but also in the reconstruction of the root type.

3.2 Distribution of k-type cherries in multi-type Yule trees

We would like to explore Yule trees where besides the splitting probabilities $q_i^{j_1 j_2}$ for $i, j_1, j_2 \in \{1, 2, \dots, k\}$, it is possible for an individual of type i to mutate to a different type j with probability q_i^j for $i, j \in \{1, 2, \dots, k\}$, $i \neq j$. In this setting we have for all $i \in \{1, \dots, k\}$

$$\sum_{j_1 \leq j_2} q_i^{j_1 j_2} + \sum_{j \neq i} q_i^j = 1.$$

These mutations appear naturally when considering the ancestral tree of a multi-type birth-death process, that is, the tree resulting from removing all non-surviving progeny in a birth-death process started at time 0 conditioned on having survived to the present time T (See Figure 3.3, far right). If an individual dies off after having produced an offspring of a different type, and this offspring has surviving progeny at time T , then this may be viewed as the individual having mutated to a different type in the ancestral tree. The first part of this section is devoted to deriving some properties of these ancestral trees, as a way of highlighting the motivation for studying processes involving mutations.

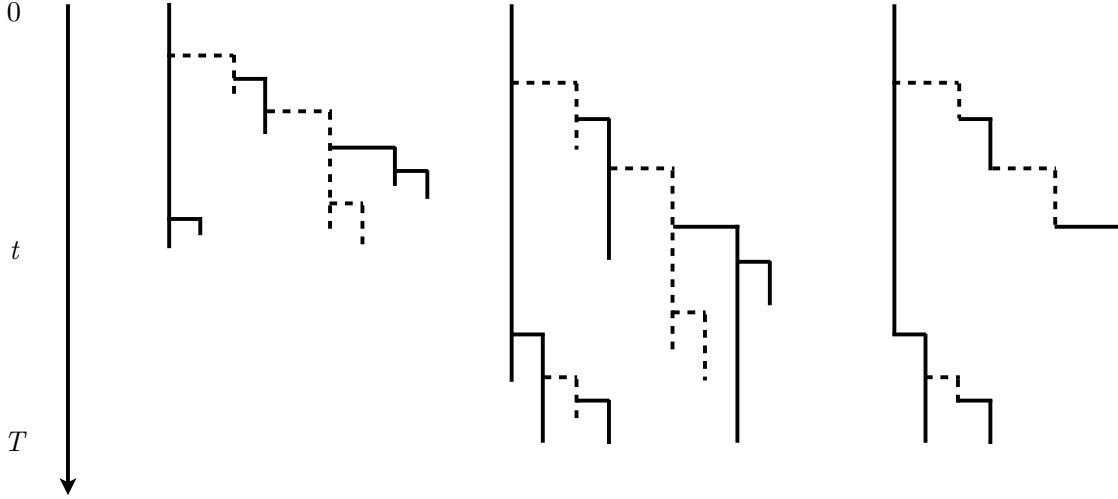


Figure 3.3: Sample trees of birth-death process. From left to right: A sample tree of a two-type birth-death process. A sample tree of a birth-death process conditioned on having survived to the present time T . The corresponding ancestral tree.

Consider a constant-rate multi-type birth-death process that is surviving at the present time $T > 0$, where \tilde{b}_i^{ij} is the constant rate at which an individual of type i has an offspring of type j . Also, let

$$\tilde{b}_i = \sum_{j \geq i} \tilde{b}_i^{ij},$$

denote the birth rate of a type- i individual and \tilde{d}_i its death rate. Thus the probability that a particular offspring of a type- i individual is of type j is given by;

$$p_i^{ij} = \frac{\tilde{b}_i^{ij}}{\tilde{b}_i}.$$

It turns out, as we show next, that the ancestral tree of this birth-death process is a pure birth process with mutations, whose splitting rates $\tilde{q}_{i,t}^{ij}$ and mutation rates $\tilde{q}_{i,t}^j$ are time-dependent. These rates may be calculated in terms of the rates \tilde{b}_i^{ij} and of the extinction probabilities $p_{e_i}^0(t, T)$ at time T of a birth-death process started at time t by a single particle of type i .

We start by proving the following lemma;

Lemma 3.1. *Let $\mathbf{Y}_t = (Y_1(t), \dots, Y_k(t))$ denote a continuous-time multi-type birth-death process starting at time 0 and surviving at the present time T . Its ancestral process $\mathbf{W}_t = (W_1(t), \dots, W_k(t))$ is also a Markov process.*

Proof. For $0 \leq t_0 \leq t_1 \leq \dots \leq t_n \leq T$ ($n \geq 1$), and a process \mathbf{Y}_t started at time t_0 with $\mathbf{Y}_{t_0} = \mathbf{y}_0$, we use the notation

$$P_{t_0; t_1, \dots, t_n}(\mathbf{y}_0; \mathbf{w}_1, \dots, \mathbf{w}_n) = \mathbb{P}(\mathbf{W}_{t_j} = \mathbf{w}_j, j = 1, \dots, n \mid \mathbf{Y}_{t_0} = \mathbf{y}_0).$$

We claim that for all $n \geq 2$

$$P_{t_0; t_1, \dots, t_n}(\mathbf{y}_0; \mathbf{w}_1, \dots, \mathbf{w}_n) = P_{t_0; t_1, \dots, t_{n-1}}(\mathbf{y}_0; \mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \frac{P_{t_0; t_{n-1}, t_n}(\mathbf{y}_0; \mathbf{w}_{n-1}, \mathbf{w}_n)}{P_{t_0; t_{n-1}}(\mathbf{y}_0; \mathbf{w}_{n-1})}. \quad (3.1)$$

This is evident for $n = 2$. Assume now that $n \geq 3$ and that equation (3.1) is true for all smaller values of n . Notice that

$$P_{t_0; t_1, \dots, t_n}(\mathbf{y}_0; \mathbf{w}_1, \dots, \mathbf{w}_n) = \sum_{\mathbf{y}_1} \mathbb{P}(\mathbf{Y}_{t_1} = \mathbf{y}_1 \mid \mathbf{Y}_{t_0} = \mathbf{y}_0) P_{t_1; t_1, \dots, t_n}(\mathbf{y}_1; \mathbf{w}_1, \dots, \mathbf{w}_n). \quad (3.2)$$

Let us focus on the probability

$$P_{t_1; t_1, \dots, t_n}(\mathbf{y}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) = \mathbb{P}(\mathbf{W}_{t_j} = \mathbf{w}_j, j = 1, \dots, n \mid \mathbf{Y}_{t_1} = \mathbf{y}_1).$$

Since the process spanned by a subset of the individuals at time t_1 is independent of the one spanned by its complement, and since all individuals surviving at time T must be

descendants of the process $\mathbf{W} = (\mathbf{W}_t)_{t \geq 0}$, then this probability is equal to:

$$\begin{aligned} P_{t_1; t_1, \dots, t_n}(\mathbf{y}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) &= C_{\mathbf{y}_1, \mathbf{w}_1} \mathbb{P}(\mathbf{W}_{t_j} = \mathbf{w}_j, j = 1, \dots, n \mid \mathbf{Y}'_{t_1} = \mathbf{w}_1) p_{\mathbf{y}_1 - \mathbf{w}_1}^0(t_1, T), \\ &= C_{\mathbf{y}_1, \mathbf{w}_1} P_{t_1; t_1, \dots, t_n}(\mathbf{w}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) p_{\mathbf{y}_1 - \mathbf{w}_1}^0(t_1, T), \end{aligned} \quad (3.3)$$

where \mathbf{Y}'_{t_1} is another birth-death process starting with \mathbf{w}_1 individuals, $C_{\mathbf{y}_1, \mathbf{w}_1}$ is the number of ways of choosing \mathbf{w}_1 individuals out of \mathbf{y}_1 , and $p_{\mathbf{y}}^0(t, T)$ is the extinction probability before time T of a process started at time t with $\mathbf{Y}'_t = \mathbf{y}$.

Let us now focus on the probability

$$P_{t_1; t_1, \dots, t_n}(\mathbf{w}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) = \mathbb{P}(\mathbf{W}_{t_j} = \mathbf{w}_j, j = 1, \dots, n \mid \mathbf{Y}'_{t_1} = \mathbf{w}_1).$$

Observe that under the condition $\mathbf{Y}'_{t_1} = \mathbf{w}_1$, the process $(\mathbf{Y}'_t)_{t \geq t_1}$ is the sum of the processes defined by each of the subtrees $\mathcal{T}^{(i)}$ ($i = 1, \dots, |\mathbf{w}_1|$) spanned by each of the $|\mathbf{w}_1|$ individuals at time t_1 . We may assume that each $\mathcal{T}^{(i)}$ spans from an individual of type $\tau^{(i)}$, where $\tau^{(1)}, \dots, \tau^{(|\mathbf{w}_1|)}$ is the sequence of types counted by \mathbf{w}_1 in increasing order, since the ordering of the trees $\mathcal{T}^{(i)}$ does not affect their distribution. The conditions in the probability above may then be written as

$$\begin{aligned} \text{(a)} \quad \sum_{i=1}^{|\mathbf{w}_1|} \mathbf{W}_{t_j}(\mathcal{T}^{(i)}) &= \mathbf{w}_j \quad (j = 2, \dots, n), \\ \text{(b)} \quad \mathbf{W}_{t_j}(\mathcal{T}^{(i)}) &\neq \mathbf{0} \quad (j = 2, \dots, n, i = 1, \dots, |\mathbf{w}_1|), \end{aligned}$$

where $\mathbf{W}_t(\mathcal{T}^{(i)})$ counts the individuals of $\mathcal{T}^{(i)}$ at time t which survive to time T . Hence, because these trees are independent,

$$P_{t_1; t_1, \dots, t_n}(\mathbf{w}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) = \sum_{\substack{\mathbf{w}_j^{(i)} \\ (j \in [2, n], \\ i \in [1, |\mathbf{w}_1|])}} \prod_{i=1}^{|\mathbf{w}_1|} P_{t_1; t_2, \dots, t_n}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_n^{(i)}), \quad (3.4)$$

where $\mathbf{e}_{\tau^{(i)}}$ denotes the unit vector with 1 in the $\tau^{(i)}$ -th coordinate. This sum is over all nonzero values of the vectors $\mathbf{w}_j^{(i)}$ ($j = 2, \dots, n, i = 1, \dots, |\mathbf{w}_1|$) such that

$$\sum_{i=1}^{|\mathbf{w}_1|} \mathbf{w}_j^{(i)} = \mathbf{w}_j \quad (j = 2, \dots, n).$$

From the inductive hypothesis (3.1) for $n - 1$, the argument of the product in equation (3.4) is equal to

$$\begin{aligned} & P_{t_1; t_2, \dots, t_n}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_n^{(i)}) \\ &= P_{t_1; t_2, \dots, t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_{n-1}^{(i)}) \frac{P_{t_1; t_{n-1}, t_n}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_{n-1}^{(i)}, \mathbf{w}_n^{(i)})}{P_{t_1; t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_{n-1}^{(i)})} \\ &= P_{t_1; t_2, \dots, t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_{n-1}^{(i)}) \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n^{(i)} | \mathbf{W}_{t_{n-1}} = \mathbf{Y}'_{t_{n-1}} = \mathbf{w}_{n-1}^{(i)}), \end{aligned}$$

where the last equality results from equations (3.2) and (3.3). The first factor above does not depend on $\mathbf{w}_n^{(i)}$ ($i = 1, \dots, |\mathbf{w}_1|$). Thus the sum of equation (3.4) may be split into two sums: an outer sum indexed by $\mathbf{w}_j^{(i)}$ ($j = 2, \dots, n - 1, i = 1, \dots, |\mathbf{w}_1|$), and an inner sum, as follows:

$$\sum_{\substack{\mathbf{w}_j^{(i)} \\ (j \in [2, n-1], \\ i \in [1, |\mathbf{w}_1|])}} \left(\prod_{i=1}^{|\mathbf{w}_1|} P_{t_1; t_2, \dots, t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_{n-1}^{(i)}) \right) \sum_{\substack{\mathbf{w}_n^{(i)} \\ (i \in [1, |\mathbf{w}_1|])}} \prod_{i=1}^{|\mathbf{w}_1|} \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n^{(i)} | \mathbf{W}_{t_{n-1}} = \mathbf{Y}'_{t_{n-1}} = \mathbf{w}_{n-1}^{(i)}).$$

By the same argument of the subtrees above, but this time splitting the individuals at time t_{n-1} into subsets of sizes $\mathbf{w}_{n-1}^{(i)}$ ($i = 1, \dots, |\mathbf{w}_1|$), the inner sum is equal to

$$\sum_{\substack{\mathbf{w}_n^{(i)} \\ (i \in [1, |\mathbf{w}_1|])}} \prod_{i=1}^{|\mathbf{w}_1|} \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n^{(i)} | \mathbf{W}_{t_{n-1}} = \mathbf{Y}_{t_{n-1}} = \mathbf{w}_{n-1}^{(i)}) = \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n | \mathbf{W}_{t_{n-1}} = \mathbf{Y}'_{t_{n-1}} = \mathbf{w}_{n-1}),$$

which can be extracted from the outer sum to obtain

$$\mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n | \mathbf{W}_{t_{n-1}} = \mathbf{Y}'_{t_{n-1}} = \mathbf{w}_{n-1}) \sum_{\substack{\mathbf{w}_j^{(i)} \\ (j \in [2, n-1], \\ i \in [1, |\mathbf{w}_1|])}} \prod_{i=1}^{|\mathbf{w}_1|} P_{t_1; t_2, \dots, t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_{n-1}^{(i)}).$$

Furthermore, again from equations (3.2) and (3.3), this is equal to:

$$\frac{P_{t_0; t_{n-1}, t_n}(\mathbf{y}_0; \mathbf{w}_{n-1}, \mathbf{w}_n)}{P_{t_0; t_{n-1}}(\mathbf{y}_0; \mathbf{w}_{n-1})} \sum_{\substack{\mathbf{w}_j^{(i)} \\ (j \in [2, n-1], \\ i \in [1, |\mathbf{w}_1|])}} \prod_{i=1}^{|\mathbf{w}_1|} P_{t_1; t_2, \dots, t_{n-1}}(\mathbf{e}_{\tau^{(i)}}; \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_{n-1}^{(i)}).$$

Hence by reusing equation (3.4), we obtain

$$P_{t_1; t_1, \dots, t_n}(\mathbf{w}_1; \mathbf{w}_1, \dots, \mathbf{w}_n) = P_{t_1; t_1, \dots, t_{n-1}}(\mathbf{w}_1; \mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \frac{P_{t_0; t_{n-1}, t_n}(\mathbf{y}_0; \mathbf{w}_{n-1}, \mathbf{w}_n)}{P_{t_0; t_{n-1}}(\mathbf{y}_0; \mathbf{w}_{n-1})}.$$

By using once again equations (3.2) and (3.3), this becomes equation (3.1). Equation (3.1) may be written in terms of conditional probabilities as follows:

$$\begin{aligned} & \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n \mid \mathbf{W}_{t_j} = \mathbf{w}_j, j = 1, \dots, n-1, \mathbf{Y}_{t_0} = \mathbf{y}_0) \\ &= \mathbb{P}(\mathbf{W}_{t_n} = \mathbf{w}_n \mid \mathbf{W}_{t_{n-1}} = \mathbf{w}_{n-1}, \mathbf{Y}_{t_0} = \mathbf{y}_0), \end{aligned}$$

which is the Markovian property for $(\mathbf{W}_t)_{t \geq 0}$. ✪

Next we find the ancestral rates of a constant-rate multi-type birth-death process.

Theorem 3.2. *The ancestral process of a continuous multi-type birth-death process with birth rates \tilde{b}_i^{ij} and death rates \tilde{d}_i , which is surviving at time T , is a pure birth process with time-dependent mutation and splitting rates given by*

$$\tilde{q}_{i,t}^{ij} = \tilde{b}_i^{ij}(1 - p_{\mathbf{e}_j}^{\mathbf{0}}(t, T)), \quad \tilde{q}_{i,t}^j = \tilde{b}_i^{ij} \frac{(1 - p_{\mathbf{e}_j}^{\mathbf{0}}(t, T))p_{\mathbf{e}_i}^{\mathbf{0}}(t, T)}{1 - p_{\mathbf{e}_i}^{\mathbf{0}}(t, T)}, \quad (3.5)$$

where $p_{\mathbf{e}_i}^{\mathbf{0}}(t, T)$ is the probability that a birth-death process started at time t with a single

type- i individual, goes extinct before time T .

Proof. By Lemma 3.1, we know that the ancestral process $(\mathbf{W}_t)_{t \geq 0}$ is a Markov process. It is thus sufficient to show that its transition rates are those of a pure birth process allowing mutations with the claimed rates. Observe first that $|\mathbf{W}_t|$ is increasing and so $\mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{y} \mid \mathbf{W}_t = \mathbf{w})$ is equal to 0 or to $O((\Delta t)^2)$ for all $\mathbf{y} \notin \{\mathbf{e}_i, i = 1, \dots, k\} \cup \{\mathbf{e}_i - \mathbf{e}_j, i, j = 1, \dots, k\}$. Furthermore

$$\begin{aligned} \mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_i \mid \mathbf{W}_t = \mathbf{w}) &= \frac{\mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_i, \mathbf{W}_t = \mathbf{w})}{\mathbb{P}(\mathbf{W}_t = \mathbf{w})} \\ &= \frac{\sum_{\mathbf{y}} \mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_i, \mathbf{W}_t = \mathbf{w}, \mathbf{Y}_t = \mathbf{y})}{\sum_{\mathbf{y}} \mathbb{P}(\mathbf{W}_t = \mathbf{w}, \mathbf{Y}_t = \mathbf{y})}. \end{aligned}$$

If only one birth or death event occurs in the time Δt , then the conditions $\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_i, \mathbf{W}_t = \mathbf{w}, \mathbf{Y}_t = \mathbf{y}$ imply that $\mathbf{Y}_{t+\Delta t} = \mathbf{y} + \mathbf{e}_i$. Thus

$$\begin{aligned} &\mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_i \mid \mathbf{W}_t = \mathbf{w}), \\ &= \frac{\sum_{\mathbf{y} \geq \mathbf{w}} \mathbb{P}(\mathbf{Y}_t = \mathbf{y}) C_{\mathbf{y}, \mathbf{w}} \sum_{\ell=1}^k w_\ell b_\ell^{\ell j} \Delta t (\mathbf{1} - \mathbf{p}_{t+\Delta t, T})^{\mathbf{w} + \mathbf{e}_i} \mathbf{p}_{t+\Delta t, T}^{\mathbf{y} - \mathbf{w}} + O(\Delta t)}{\sum_{\mathbf{y}} \mathbb{P}(\mathbf{Y}_t = \mathbf{y}) C_{\mathbf{y}, \mathbf{w}} (\mathbf{1} - \mathbf{p}_{t, T})^{\mathbf{w}} \mathbf{p}_{t, T}^{\mathbf{y} - \mathbf{w}}} \quad (3.6) \\ &= \sum_{\ell=1}^k w_\ell \tilde{b}_\ell^{\ell i} (1 - p_{\mathbf{e}_i}^0(t, T)) \Delta t + O(\Delta t), \end{aligned}$$

where $\mathbf{p}_{t, T}^{\mathbf{w}} := \prod_{i=1}^k p_{\mathbf{e}_i}^0(t, T)^{w_i}$ and $(\mathbf{1} - \mathbf{p}_{t, T})^{\mathbf{w}} := \prod_{i=1}^k (1 - p_{\mathbf{e}_i}^0(t, T))^{w_i}$. Similarly for $i \neq j$,

we have

$$\begin{aligned}
& \mathbb{P}(\mathbf{W}_{t+\Delta t} = \mathbf{w} + \mathbf{e}_j - \mathbf{e}_i \mid \mathbf{W}_t = \mathbf{w}), \\
&= \frac{\sum_{\mathbf{y} \geq \mathbf{w}} \mathbb{P}(\mathbf{Y}_t = \mathbf{y}) C_{\mathbf{y}, \mathbf{w}} \sum_{\ell=1}^k w_\ell b_\ell^{\ell j} \Delta t (1 - \mathbf{p}_{t+\Delta t, T})^{\mathbf{w} + \mathbf{e}_j - \mathbf{e}_i} \mathbf{p}_{t+\Delta t, T}^{\mathbf{y} - \mathbf{w} + \mathbf{e}_i} + O(\Delta t)}{\sum_{\mathbf{y}} \mathbb{P}(\mathbf{Y}_t = \mathbf{y}) C_{\mathbf{y}, \mathbf{w}} (1 - \mathbf{p}_{t, T})^{\mathbf{w}} \mathbf{p}_{t, T}^{\mathbf{y}}} \quad (3.7) \\
&= w_j \tilde{b}_i^{ij} \frac{(1 - p_{\mathbf{e}_j}^0(t, T)) p_{\mathbf{e}_i}^{\mathbf{y}}(t, T)}{1 - p_{\mathbf{e}_i}^0(t, T)} \Delta t + O(\Delta t).
\end{aligned}$$

As claimed, the transition rates (3.6) and (3.7) correspond to the transition rates of a pure birth process allowing mutations (Section 1.2.2) with the splitting and mutation rates (3.5).



Remark 3.8. *Even though the death rates do not appear explicitly in the rates of the ancestral tree. They are implicitly involved with the probability of extinction, because the probability of extinction $p_{\mathbf{e}_i}^0(t, T)$ satisfies the following differential equation (which can be obtained from the formulae in [37] and [24] by a straightforward calculation)*

$$\frac{dp_{\mathbf{e}_i}^0(t, T)}{dt} = \tilde{d}_i - (\tilde{b}_i + \tilde{d}_i) p_{\mathbf{e}_i}^0(t, T) + \sum_{j=1}^k \tilde{b}_i^{ij} p_{\mathbf{e}_i}^0(t, T) p_{\mathbf{e}_j}^0(t, T), \text{ for } i \in \{1, 2, \dots, k\}.$$

From the rates of Theorem 3.2 we can also obtain the probability that a given split or mutation of a type- i individual is of a particular type ij or ij ;

Corollary 3.6. *The mutation and splitting probabilities of the ancestral tree of a continuous birth-death process are as follows:*

$$q_{i,t}^{ij} = \frac{p_i^{ij} (1 - p_{\mathbf{e}_i}^0(t, T)) (1 - p_{\mathbf{e}_j}^0(t, T))}{1 - \sum_{j_1=1}^k p_i^{ij_1} p_{\mathbf{e}_{j_1}}^0(t, T) - p_i^{ii} p_{\mathbf{e}_i}^0(t, T) (1 - p_{\mathbf{e}_i}^0(t, T))},$$

$$q_{i,t}^j = \frac{p_i^{ij}(1 - p_{e_j}^0(t, T))p_{e_i}^0(t, T)}{1 - \sum_{j_1=1}^k p_i^{ij_1} p_{e_{j_1}}^0(t, T) - p_i^{ii} p_{e_i}^0(t, T)(1 - p_{e_i}^0(t, T))}$$

where the notation is as in Theorem 3.2.

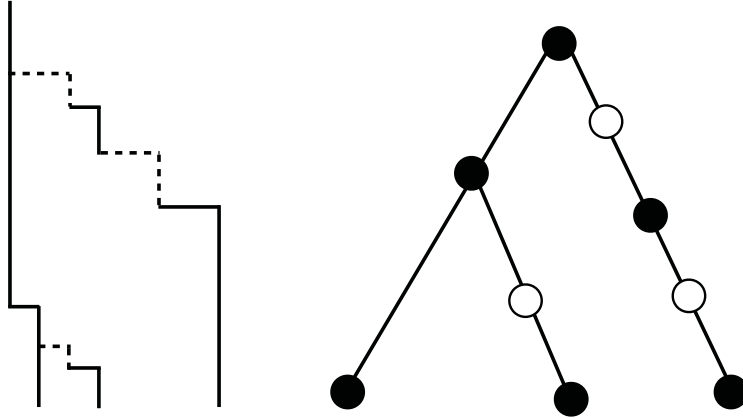


Figure 3.4: The ancestral tree of Figure 3.3 and the corresponding two-type Yule tree with mutations. The Yule tree has one cherry of type 222 and a pendant edge of type 22.

3.2.1 Multi-type Yule trees with mutations

The ancestral model above explains the motivation to consider multi-type Yule trees with mutations. A binary tree with mutations representing the topology of an ancestral multi-type birth-death tree must be so that one of the offsprings of each type- i individual is of type i as well (see Figure 3.4 for an example of a multi-type Yule tree with mutations resulting from the ancestral tree of Figure 3.3). However, in the Yule model we study next we ignore this restriction and assume that a type- i individual may split into offsprings of types $j_1 \leq j_2$ for $j_1, j_2 \neq i$.

Our general multi-type Yule model is as follows. We start at time 1 with one individual of fixed type. At each step of the multi-type Yule process, an individual is randomly chosen with weights \hat{w}_i for each type $i \in \{1, \dots, k\}$. With probability $q_i^{j_1 j_2}$ ($j_1 \leq j_2$), the

chosen individual has two children of types j_1, j_2 , and with probability q_i^j , it mutates and becomes of type $j \neq i$. As we mentioned above, for all $i \in \{1, \dots, k\}$

$$\sum_{j_1 \leq j_2} q_i^{j_1 j_2} + \sum_{j \neq i} q_i^j = 1.$$

In this k -type scenario we have a total of

$$k \binom{k+1}{2}$$

different k -type cherries. A cherry of type ij_1j_2 consists of two individuals of types j_1 and j_2 (possibly after any number of mutations) with a common parent of type i . A pendant edge of type ij consists of an individual of type j (possibly after mutations) whose parent is of type i (see Figure 3.4). Having different weights \hat{w}_i for different types makes a generating function approach inaccessible, since not only derivatives but also integrals would appear in the resulting differential equations, making them apparently impossible to solve with traditional methods. Also an extended Pólya urn model approach on all cherries and pendant edges leads to equations which algebra systems like Maple are not able to solve, even for particular cases like $k = 2$. We aim to resolve this issue by working with a continuous-time multi-type pure birth process where the distribution of cherries is the same as in the multi-type Yule tree just described. In this continuous time process, each individual of type i splits into two children of types $j_1 \leq j_2$ at a rate $\tilde{q}_i^{j_1 j_2}$ or becomes of type $j \neq i$ at a rate \tilde{q}_i^j . Unlike regular birth-death processes, it is possible for j_1, j_2 to be both different from i . Assume all of the summation indices in the next formulae range over the set $[k] := \{1, 2, \dots, k\}$. Denote

$$\tilde{q}_i = \sum_{j \leq l} \tilde{q}_i^{j l} + \sum_{j \neq i} \tilde{q}_i^j.$$

The discrete multi-type Yule process with weights and probabilities satisfying

$$\hat{w}_i = \frac{\tilde{q}_i}{\sum_i \tilde{q}_i}, \quad q_i^{j_1 j_2} = \frac{\tilde{q}_i^{j_1 j_2}}{\tilde{q}_i}, \quad q_i^j = \frac{\tilde{q}_i^j}{\tilde{q}_i},$$

is embedded in the continuous-time process and has the same distribution for the proportion of cherries of each type at a time when the process first reaches a set number of leaves. This multi-type birth-death model can be generalized further by assuming that the rates are continuous functions on the time variable t . The notation for these rates then becomes $\tilde{q}_{i,t}^{j_1 j_2}$ for $j_1 \leq j_2$, and $\tilde{q}_{i,t}^j$ for $j \neq i$.

We would like to calculate the proportion of the mean number of cherries of each type. We start by obtaining results on the proportion of leaves of each type then continue to study the proportion of the mean number of cherries.

Theorem 3.3. *Let $\tilde{\mathbf{v}}_t = (\tilde{v}_t^1, \tilde{v}_t^2, \dots, \tilde{v}_t^k)$ be the vector containing all mean number of leaves of type $1, 2, \dots, k$ at time t in the continuous-time pure birth process with mutations described above, with time-dependent rates, starting at time 0 with a single individual of type $a \in [k]$. Then*

$$\partial_t \tilde{\mathbf{v}}_t = \mathbf{B}_t \tilde{\mathbf{v}}_t, \quad \tilde{\mathbf{v}}_0 = \mathbf{e}_a,$$

where \mathbf{B}_t is the $k \times k$ matrix with entries;

$$B_{\ell_1 \ell_2, t} = \begin{cases} \tilde{q}_{\ell_1, t}^{\ell_1 \ell_1} - \sum_{\substack{i \leq j \\ i, j \neq \ell_1}} \tilde{q}_{\ell_1, t}^{ij} - \sum_{i \neq \ell_1} \tilde{q}_{\ell_1, t}^i, & \text{when } \ell_1 = \ell_2, \\ 2\tilde{q}_{\ell_2, t}^{\ell_1 \ell_1} + \tilde{q}_{\ell_2, t}^{\ell_1} + \sum_{j < \ell_1} \tilde{q}_{\ell_2, t}^{j \ell_1} + \sum_{j > \ell_1} \tilde{q}_{\ell_2, t}^{\ell_1 j}, & \text{when } \ell_1 \neq \ell_2. \end{cases}$$

Proof. This statement is equivalent to the claim that each \tilde{v}_t^ℓ satisfies the following dif-

ferential equation:

$$\partial_t \tilde{\nu}_t^\ell = \sum_{i \neq \ell} (2\tilde{q}_{i,t}^{\ell\ell} + \tilde{q}_{i,t}^\ell + \sum_{j < \ell} \tilde{q}_{i,t}^{j\ell} + \sum_{j > \ell} \tilde{q}_{i,t}^{\ell j}) \tilde{\nu}_t^i + (\tilde{q}_{\ell,t}^{\ell\ell} - \sum_{\substack{i \leq j \\ i, j \neq \ell}} \tilde{q}_{\ell,t}^{ij} - \sum_{i \neq \ell} \tilde{q}_{\ell,t}^i) \tilde{\nu}_t^\ell, \text{ for } 1 \leq \ell \leq k.$$

Indeed, in a small interval of time, for each $\ell \in [k]$, the mean number $\tilde{\nu}_t^\ell$ of leaves of type ℓ , increases by 2 whenever we add a cherry of type $i\ell\ell$ for some $i \neq \ell$. It increases by 1 if we add a cherry of type $ij\ell$ for some $i, j \neq \ell$ or for $i = j = \ell$, or if we apply a change of type $i\ell$ for some $i \neq \ell$. It decreases by one upon adding cherries of types lij or changes of type li for $i, j \neq \ell$. \star

Corollary 3.7. *Let $\tilde{\nu}_t$ and \mathbf{B}_t be as in Theorem 3.3 above. If \mathbf{B}_t commutes with $\int_0^t \mathbf{B}_\tau d\tau$ for all $t \geq 0$, then:*

$$\tilde{\nu}_t = \exp\left(\int_0^t \mathbf{B}_\tau d\tau\right) \mathbf{e}_a.$$

Corollary 3.8. *Let $\tilde{\nu}_t$ and \mathbf{B}_t be as in Theorem 3.3 above. If $\mathbf{B}_t = \mathbf{B}$ is a constant (time-independent) matrix, then:*

$$\tilde{\nu}_t = \exp(\mathbf{B}t) \mathbf{e}_a.$$

By adding up all leave counts we obtain the following.

Corollary 3.9. *Let ρ_t be the mean number of leaves at time t of the continuous-time pure birth process described above. If \mathbf{B}_t commutes with $\int_0^t \mathbf{B}_\tau d\tau$ for all $t \geq 0$, then*

$$\rho_t = \mathbf{1}^\top \exp\left(\int_0^t \mathbf{B}_\tau d\tau\right) \mathbf{e}_a.$$

In particular, if $\mathbf{B}_t = \mathbf{B}$ is time-independent, then

$$\rho_t = \mathbf{1}^\top \exp(\mathbf{B}t) \mathbf{e}_a.$$

We now present a theorem on the differential equations satisfied by the mean number of k -type cherries for fixed parent type ℓ .

Theorem 3.4. *Let $\boldsymbol{\mu}_t^\ell = (\mu_t^{\ell 11}, \mu_t^{\ell 12}, \dots, \mu_t^{\ell kk})$ be the vector whose entries are the mean number of k -type cherries of types ℓij for $1 \leq i \leq j \leq k$, in a continuous-time pure birth process as described before with time-dependent rates. Then,*

$$\partial_t \boldsymbol{\mu}_t^\ell = \mathbf{A}_t \boldsymbol{\mu}_t^\ell + \tilde{\mathbf{q}}_{(\ell),t} \tilde{\boldsymbol{\nu}}_t^\ell,$$

where

$$\tilde{\mathbf{q}}_{(\ell),t} := [\tilde{q}_{\ell,t}^{11}, \tilde{q}_{\ell,t}^{12}, \tilde{q}_{\ell,t}^{13}, \dots, \tilde{q}_{\ell,t}^{kk}]^\top,$$

and \mathbf{A}_t is a $\binom{k+1}{2} \times \binom{k+1}{2}$ matrix with entries given by

$$A_{\ell ij, \ell mn, t} = \begin{cases} -(\tilde{q}_{i,t} + \tilde{q}_{j,t}), & \text{for } (m, n) = (i, j), \\ \delta_{m,i} \tilde{q}_{n,t}^i + \delta_{n,i} \tilde{q}_{m,t}^i, & \text{for } (m, n) \neq (i, j), i = j, \\ \delta_{m,i} \tilde{q}_{n,t}^j + \delta_{m,j} \tilde{q}_{n,t}^i + \delta_{n,i} \tilde{q}_{m,t}^j + \delta_{n,j} \tilde{q}_{m,t}^i, & \text{for } (m, n) \neq (i, j), i \neq j, \end{cases}$$

assuming that the vectors $\boldsymbol{\mu}_t^\ell$, $\tilde{\mathbf{q}}_{(\ell),t}$ and the entries of the matrix \mathbf{A}_t have a consistent order.

Proof. We just need to show that each $\mu_t^{\ell ij}$ satisfies the following differential equation:

$$\partial_t \mu_t^{\ell ij} = \begin{cases} \sum_{\substack{m \leq n \\ (m,n) \neq (i,i)}} (\delta_{m,i} \tilde{q}_{n,t}^i + \delta_{n,i} \tilde{q}_{m,t}^i) \mu_t^{\ell mn} - 2\tilde{q}_{i,t} \mu_t^{\ell ii} + \tilde{q}_{\ell,t}^{ii} \nu_t^\ell, & \text{for } i = j, \\ \sum_{\substack{m \leq n \\ (m,n) \neq (i,j)}} (\delta_{m,i} \tilde{q}_{n,t}^j + \delta_{m,j} \tilde{q}_{n,t}^i + \delta_{n,i} \tilde{q}_{m,t}^j + \delta_{n,j} \tilde{q}_{m,t}^i) \mu_t^{\ell mn} - (\tilde{q}_{i,t} + \tilde{q}_{j,t}) \mu_t^{\ell ij} + \tilde{q}_{\ell,t}^{ij} \nu_t^\ell, & \text{for } i \neq j. \end{cases}$$

Indeed, the mean number of cherries of type lij will increase upon adding a cherry of type lij , or by a change of type which produces a cherry of type lij . Also it will decrease by mutating or splitting the leaves of cherries of type lij . \star

Remark 3.9. *In the previous theorem, the matrix \mathbf{A}_t is diagonally dominant by columns. This is because in the column lmn , every rate of the form $\tilde{q}_{m,t}^i$ and every rate of the form $\tilde{q}_{n,t}^i$ appears exactly once (when $n = m$, each one appears twice), and the sum of these rates is less than or equal to $\tilde{q}_{n,t} + \tilde{q}_{m,t}$. We use this fact in future proofs.*

We now give a result regarding the mean number of pendant edges. This statistic is useful in subsequent theorems where we reconstruct birth and mutation rates from observed asymptotic means.

Theorem 3.5. *Let $\boldsymbol{\gamma}_t = (\gamma_t^{11}, \gamma_t^{12}, \dots, \gamma_t^{kk})$ be the vector whose entries are the mean number of pendant edges of different types in a continuous pure birth process with time-dependent rates. Then,*

$$\partial_t \boldsymbol{\gamma}_t = \mathbf{C}_t \boldsymbol{\gamma}_t + \mathbf{U}_t \boldsymbol{\mu}_t,$$

where

$$C_{lm,ij,t} = \begin{cases} -\tilde{q}_{m,t} & \text{for } (\ell, m) = (i, j), \\ \tilde{q}_{j,t}^m & \text{for } \ell = i, m \neq j, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$U_{\ell m, \ell' ij, t} = \begin{cases} 2 \sum_{j_1 \leq j_2} \tilde{q}_{m, t}^{j_1 j_2} & \text{for } \ell = \ell', m = i = j, \\ \sum_{j_1 \leq j_2} \tilde{q}_{i, t}^{j_1 j_2} & \text{for } \ell = \ell', m = j > i, \\ \sum_{j_1 \leq j_2} \tilde{q}_{j, t}^{j_1 j_2} & \text{for } \ell = \ell', m = i < j, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We claim that each $\gamma_t^{\ell m}$ satisfies

$$\partial_t \gamma_t^{\ell m} = \sum_{j \neq m} \tilde{q}_{j, t}^m \gamma_t^{\ell j} - \tilde{q}_{m, t} \gamma_t^{\ell m} + \sum_{i < m} \left(\sum_{j_1 \leq j_2} \tilde{q}_{i, t}^{j_1 j_2} \right) \mu_t^{\ell i m} + \sum_{i > m} \left(\sum_{j_1 \leq j_2} \tilde{q}_{i, t}^{j_1 j_2} \right) \mu_t^{\ell m i} + 2 \sum_{j_1 \leq j_2} \tilde{q}_{m, t}^{j_1 j_2} \mu_t^{\ell m m}.$$

Indeed, the number of pendant edges of type ℓm will increase by one if a mutation of type $j m$ is added on a pedant edge of type ℓj . Also, it will increase if any cherry of type $i j_1 j_2$ is added to any cherry of type $\ell i m$ ($i \neq m$), or if a cherry of type $m j_1 j_2$ is added to a cherry of type $\ell m m$. It will decrease by one if we add any cherry or mutation on a pendant edge of type ℓm . \odot

Notice that the matrix \mathbf{B}_t of Theorem 3.3 has nonnegative entries, except possibly those in the diagonal. Thus by the Perron-Frobenius theorem, if it is irreducible, there exists a largest simple eigenvalue λ_t of \mathbf{B}_t with right and left eigenvectors $\mathbf{u}_t, \mathbf{v}_t$, respectively. We can assume that $\mathbf{1} \cdot \mathbf{u}_t = 1$. In the case that $\mathbf{B}_t = \mathbf{B}$ is time independent, we have the following result.

Theorem 3.6. *In the time-independent case, if $\boldsymbol{\eta}_t^\ell = \frac{\boldsymbol{\mu}_t^\ell}{\rho_t}$, the $\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ exists and \mathbf{B} is irreducible, then*

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell (\mathbf{A} - \lambda \mathbf{I})^{-1} \tilde{\mathbf{q}}_{(\ell)},$$

where λ is the largest real eigenvalue of \mathbf{B} and $\mathbf{u} = (u_1, u_2, \dots, u_k)$ is its corresponding right eigenvector.

Proof. Let us take $\boldsymbol{\mu}_t^\ell = \rho_t \boldsymbol{\eta}_t^\ell$ in the differential equation $\partial_t \boldsymbol{\mu}_t^\ell = \mathbf{A} \boldsymbol{\mu}_t^\ell + \tilde{\mathbf{q}}^{(\ell)} \tilde{\nu}_t^\ell$ from Theorem 3.4. We obtain

$$\partial_t \boldsymbol{\eta}_t^\ell = \left(\mathbf{A} - \frac{\partial_t \rho_t}{\rho_t} \mathbf{I} \right) \boldsymbol{\eta}_t^\ell + \frac{\tilde{\mathbf{q}}^{(\ell)} \tilde{\nu}_t^\ell}{\rho_t}.$$

We know that $\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ exists. Taking limit as $t \rightarrow \infty$ on both sides and since $\boldsymbol{\eta}_t^\ell$ is continuous, we get

$$0 = \left(\mathbf{A} - \lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t} \mathbf{I} \right) \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell + \tilde{\mathbf{q}}^{(\ell)} \lim_{t \rightarrow \infty} \frac{\tilde{\nu}_t^\ell}{\rho_t}. \quad (3.8)$$

Let us compute $\lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t}$ and $\lim_{t \rightarrow \infty} \frac{\tilde{\mathbf{q}}^{(\ell)} \tilde{\nu}_t^\ell}{\rho_t}$. Let \mathbf{J} denote the Jordan representation form of \mathbf{B} , so that $\mathbf{B} = \mathbf{P} \mathbf{J} \mathbf{P}^{-1}$ and $\exp(\mathbf{B}t) = \mathbf{P} \exp(\mathbf{J}t) \mathbf{P}^{-1}$. By Corollary 3.9,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t} &= \lim_{t \rightarrow \infty} \frac{\mathbf{1}^\top \mathbf{B} \exp(\mathbf{B}t) \mathbf{e}_a}{\mathbf{1}^\top \exp(\mathbf{B}t) \mathbf{e}_a} \\ &= \frac{\mathbf{1}^\top \lambda \mathbf{u} \mathbf{v}^\top \mathbf{e}_a}{\mathbf{1}^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a} \quad (\text{since the dominating terms are only those involving } e^{\lambda t}) \\ &= \lambda. \end{aligned}$$

Similarly, using Corollary 3.8,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\tilde{\nu}_t^\ell}{\rho_t} &= \frac{\mathbf{e}_\ell^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a}{\mathbf{1}^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a} \\ &= u_\ell. \end{aligned}$$

We now claim that $(\mathbf{A} - \lambda \mathbf{I})$ is invertible. This is true because \mathbf{A} is diagonally dominant by columns (see Remark 3.9), and $\lambda \geq 0$ (since ρ_t is positive and increasing), which means that $(\mathbf{A} - \lambda \mathbf{I})$ is diagonally dominant by columns as well. Hence, from

Equation (3.8), we obtain

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell (\mathbf{A} - \lambda \mathbf{I})^{-1} \tilde{\mathbf{q}}_{(\ell)}.$$



Using similar arguments, and the fact that the matrix $\mathbf{C} - \lambda \mathbf{I}$ is diagonally dominant, we obtain the following Theorem.

Theorem 3.7. *In the time-independent case, if $\boldsymbol{\eta}_t^* = \frac{\boldsymbol{\gamma}_t}{\rho_t}$, the $\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^*$ exists and \mathbf{B} is irreducible, then*

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^* = -(\mathbf{C} - \lambda \mathbf{I})^{-1} \mathbf{U} \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_t^1, \boldsymbol{\eta}_t^2, \dots, \boldsymbol{\eta}_t^k)$.

To obtain time-dependent versions of Theorems 3.6 and 3.7 we need some assumptions on the matrices involved.

Theorem 3.8. *Suppose all of the following assumptions are true.*

- a) *The limits $\lim_{t \rightarrow \infty} \tilde{q}_{i,t}^{j_1 j_2}$ and $\lim_{t \rightarrow \infty} \tilde{q}_{i,t}^j$ exist for all $i, j, j_1, j_2 \in [k]$ with $i \neq j$ and $j_1 \leq j_2$.*
- b) *The limit $\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ exists for all $\ell \in [k]$.*
- c) *The matrix $\lim_{t \rightarrow \infty} \mathbf{B}_t$ is irreducible and has right and left eigenvectors \mathbf{u}, \mathbf{v} , respectively, with associated maximum eigenvalue λ .*

Then,

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell \lim_{t \rightarrow \infty} (\mathbf{A}_t - \lambda \mathbf{I})^{-1} \lim_{t \rightarrow \infty} \tilde{\mathbf{q}}_{(\ell),t}.$$

Proof. Let us replace $\boldsymbol{\mu}_t^\ell = \rho_t \boldsymbol{\eta}_t^\ell$ in the differential equation $\partial_t \boldsymbol{\mu}_t^\ell = \mathbf{A}_t \boldsymbol{\mu}_t^\ell + \tilde{\mathbf{q}}_{(\ell),t} \tilde{\boldsymbol{\nu}}_t^\ell$ from Theorem 3.4. We obtain,

$$\partial_t \boldsymbol{\eta}_t^\ell = \left(\mathbf{A}_t - \frac{\partial_t \rho_t}{\rho_t} \mathbf{I} \right) \boldsymbol{\eta}_t^\ell + \frac{\tilde{\mathbf{q}}_{(\ell),t} \tilde{\boldsymbol{\nu}}_t^\ell}{\rho_t}.$$

We know that $\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ exists. Taking limit as $t \rightarrow \infty$ on both sides and since $\boldsymbol{\eta}_t^\ell$ is continuous, we get

$$0 = \left(\lim_{t \rightarrow \infty} \mathbf{A}_t - \lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t} \mathbf{I} \right) \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell + \lim_{t \rightarrow \infty} \tilde{\mathbf{q}}_{(\ell),t} \lim_{t \rightarrow \infty} \frac{\tilde{\boldsymbol{\nu}}_t^\ell}{\rho_t}. \quad (3.9)$$

We thus need to find expressions for $\lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t}$ and $\lim_{t \rightarrow \infty} \frac{\tilde{\boldsymbol{\nu}}_t^\ell}{\rho_t}$. From Theorem 3.3, we know that $\partial_t \tilde{\boldsymbol{\nu}}_t = \mathbf{B}_t \tilde{\boldsymbol{\nu}}_t$, and by letting $\tilde{\boldsymbol{\beta}}_t = \frac{\tilde{\boldsymbol{\nu}}_t}{\rho_t}$, we have

$$\partial_t \tilde{\boldsymbol{\beta}}_t = \mathbf{B}_t \tilde{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_t \frac{\partial_t \rho_t}{\rho_t}.$$

As before by taking $\lim_{t \rightarrow \infty}$ on both sides, we obtain

$$\left(\lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t} \right) (\lim_{t \rightarrow \infty} \tilde{\boldsymbol{\beta}}_t) = (\lim_{t \rightarrow \infty} \mathbf{B}_t) (\lim_{t \rightarrow \infty} \tilde{\boldsymbol{\beta}}_t).$$

Notice that the matrix $\lim_{t \rightarrow \infty} \mathbf{B}_t$ has all finite entries (by (a)) and from (c) it is irreducible. Since the vector $\lim_{t \rightarrow \infty} \tilde{\boldsymbol{\beta}}_t$ only has nonnegative entries, the Perron-Frobenius Theorem implies that this vector is the eigenvector \mathbf{u} and that $\lambda = \lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t}$.

We now claim that $(\mathbf{A}_t - \lambda \mathbf{I})$ is invertible. This is true because \mathbf{A}_t is diagonally dominant by columns (see Remark 3.9), and $\lambda \geq 0$ (since ρ_t is positive and increasing), which means that $(\mathbf{A}_t - \lambda \mathbf{I})$ is diagonally dominant by columns as well. Hence, from Equation (3.9), we obtain,

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell \lim_{t \rightarrow \infty} (\mathbf{A}_t - \lambda \mathbf{I})^{-1} \lim_{t \rightarrow \infty} \tilde{\mathbf{q}}_{(\ell),t}.$$



Remark 3.10. *In the particular case where for all t, t' the matrices $\mathbf{B}_t, \mathbf{B}_{t'}$ are irreducible, mutually diagonalizable, and have the same right and left eigenvectors \mathbf{u}, \mathbf{v} for their corresponding maximum eigenvalues $\lambda_t, \lambda_{t'}$, we can drop the assumption that \mathbf{B} is irreducible. In that case we have*

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell \lim_{t \rightarrow \infty} (\mathbf{A}_t - \lambda_t \mathbf{I})^{-1} \lim_{t \rightarrow \infty} \tilde{\mathbf{q}}_{(\ell), t}.$$

Indeed, from Corollary 3.9

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\partial_t \rho_t}{\rho_t} &= \lim_{t \rightarrow \infty} \frac{\mathbf{1}^\top \mathbf{B}_t \exp(\int_0^t \mathbf{B}_\tau d\tau) \mathbf{e}_a}{\mathbf{1}^\top \exp(\int_0^t \mathbf{B}_\tau d\tau) \mathbf{e}_a} \\ &= \lim_{t \rightarrow \infty} \frac{\mathbf{1}^\top \mathbf{P} \mathbf{D}_t \mathbf{P}^{-1} \mathbf{P} \exp(\int_0^t \mathbf{D}_\tau d\tau) \mathbf{P}^{-1} \mathbf{e}_a}{\mathbf{1}^\top \mathbf{P} \exp(\int_0^t \mathbf{D}_\tau d\tau) \mathbf{P}^{-1} \mathbf{e}_a} \\ &= \lim_{t \rightarrow \infty} \frac{\mathbf{1}^\top \mathbf{P} \mathbf{D}_t \exp(\int_0^t \mathbf{D}_\tau d\tau) \mathbf{P}^{-1} \mathbf{e}_a}{\mathbf{1}^\top \mathbf{P} \exp(\int_0^t \mathbf{D}_\tau d\tau) \mathbf{P}^{-1} \mathbf{e}_a} \\ &= \lim_{t \rightarrow \infty} \lambda_t \frac{\mathbf{1}^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a}{\mathbf{1}^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a} \text{ (since the dominating terms are only those involving } e^{\int_0^t \lambda_1, \tau d\tau} \\ &\quad \text{with } \mathbf{u}, \mathbf{v} \text{ as right and left eigenvectors of } \mathbf{B}_t, \text{ respectively.)} \\ &= \lim_{t \rightarrow \infty} \lambda_t, \end{aligned}$$

and as before from Corollary 3.7

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\tilde{\nu}_t^\ell}{\rho_t} &= \frac{\mathbf{e}_\ell \mathbf{u} \mathbf{v}^\top \mathbf{e}_a}{\mathbf{1}^\top \mathbf{u} \mathbf{v}^\top \mathbf{e}_a} \\ &= u_\ell. \end{aligned}$$

Using similar arguments, we obtain the following result.

Theorem 3.9. *If all of the assumptions of Theorem 3.8 are satisfied, and the limit*

$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^*$ exists, where $\boldsymbol{\eta}_t^* = \frac{\boldsymbol{\gamma}_t}{\rho_t}$, then

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^* = - \lim_{t \rightarrow \infty} (\mathbf{C}_t - \lambda \mathbf{I})^{-1} \mathbf{U}_t \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t,$$

with $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_t^1, \boldsymbol{\eta}_t^2, \dots, \boldsymbol{\eta}_t^k)$.

Remark 3.11. *In the time-independent case, this pure-birth process depends on a total of*

$$k \binom{k+1}{2} + k(k-1)$$

parameters (rates). In the previous result we have obtained the limits of

$$k \binom{k+1}{2} + k^2$$

ratios which satisfy the relation

$$2 \sum_{\substack{\ell \in [k] \\ i \leq j}} \eta_t^{\ell ij} + \sum_{i,j} \eta_t^{*ij} = 1.$$

This plays an important role when we wish to obtain information about the birth and transition rates of a process, having only some information about the distribution of types at the present time.

The following results illustrate some particular cases for which it is possible to reconstruct the birth and transition rates of a multi-type Yule process knowing the behaviour in the long run of the mean number of cherries and pendant edges.

Proposition 3.5. *If $\sum_{j_1 \leq j_2} \tilde{q}_{i,t}^{j_1, j_2} = r_{i,t}$, $\forall i \in [k]$, for some functions $r_{i,t}$ whose limit when $t \rightarrow \infty$ is known, and the maximum real eigenvalue λ of $\lim_{t \rightarrow \infty} \mathbf{B}_t$ is also known,*

then it is possible to compute all the limits of the mutation and birth rates in terms of the limit vectors $\mathbf{w}^* = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^*$ and $\mathbf{w}^\ell = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ (for all $\ell \in [k]$).

Proof. Notice that:

$$C_{\ell m, ij} = \begin{cases} -r_{m,t} - \sum_{i \neq m} \tilde{q}_{m,t}^i & \text{for } (\ell, m) = (i, j), \\ \tilde{q}_{j,t}^m & \text{for } \ell = i, m \neq j, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$U_{\ell m, \ell' ij} = \begin{cases} 2r_{m,t} & \text{for } \ell = \ell', m = i = j, \\ r_{i,t} & \text{for } \ell = \ell', m = j > i, \\ r_{j,t} & \text{for } \ell = \ell', m = i < j, \\ 0 & \text{otherwise.} \end{cases}$$

Since λ is known, from Theorem 3.9, we obtain the following system

$$\lim_{t \rightarrow \infty} (\mathbf{C}_t - \lambda \mathbf{I}) \mathbf{w}^* + \lim_{t \rightarrow \infty} \mathbf{U}_t \mathbf{w} = \mathbf{0},$$


where $\mathbf{w} = (\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^k)$. Notice that this system is linear and depends only on the limits of the mutation rates $\tilde{q}_{i,t}^j$ for $i, j \in [k]$ and $i \neq j$. Hence, there exists a set of solutions for these limits. For each solution, these limits can be replaced into each of the

systems, and we have

$$\lim_{t \rightarrow \infty} (\mathbf{A}_t - \lambda \mathbf{I}) \mathbf{w}^\ell + u_\ell \tilde{\mathbf{q}}_{(\ell),t} = \mathbf{0}, \text{ for all } \ell \in [k].$$

We claim that this is a linear system on the branching rates. This is true because

$$u_\ell = 2 \sum_i w_i^{\ell\ell} + \sum_{\substack{i \in [k] \\ j < \ell}} w_i^{j\ell} + \sum_{\substack{i \in [k] \\ j > \ell}} w_i^{\ell j} + \sum_{i \in [k]} w_i^\ell.$$

Therefore it is possible to obtain a set of solutions for each limit $\lim_{t \rightarrow \infty} \tilde{q}_{i,t}^{j_1 j_2}$ in terms of the limit vectors \mathbf{w}, \mathbf{w}^* , as claimed. 

Corollary 3.10. *In the time-independent case, if $\sum_{j_1 \leq j_2} \tilde{q}_i^{j_1, j_2} = r_i, \forall i \in [k]$, for some known constants r_i and the maximum real eigenvalue λ of \mathbf{B} is also known, then it is possible to compute all mutation and birth rates in terms of the limit vectors $\mathbf{w}^* = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^*$ and $\mathbf{w}^\ell = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ (for all $\ell \in [k]$).*

Corollary 3.11. *Assuming that $\sum_{j_1 \leq j_2} \tilde{q}_i^{j_1, j_2} = r$ for some constant r which does not depend on i , and given the limit vectors $\mathbf{w}^* = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^*$ and $\mathbf{w}^\ell = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell$ (for all $\ell \in [k]$), it is possible to compute all mutation and birth rates in terms of those vectors.*

Proof. This is a direct result from the previous Proposition by noticing that the maximum eigenvalue of \mathbf{B} is $\lambda = r$. Indeed, since r is the birth rate of the total population, we have:

$$\begin{aligned} r \rho_t &= \partial_t \rho_r \\ &= \partial_t (\mathbf{1} \cdot \tilde{\boldsymbol{\nu}}_t) \\ &= \mathbf{1} \cdot (\partial_t \tilde{\boldsymbol{\nu}}_t) \\ &= \mathbf{1} \cdot (\mathbf{B} \tilde{\boldsymbol{\nu}}_t). \end{aligned}$$

Thus

$$r = \mathbf{1} \cdot \begin{pmatrix} \tilde{\mathbf{B}} \tilde{\boldsymbol{\nu}}_t \\ \rho_t \end{pmatrix}.$$

Taking the limit as t goes to infinity, we have

$$r = \mathbf{1} \cdot (\lambda \mathbf{u}).$$

Hence, $r = \lambda$ and the result follows. ◉

3.2.2 Predictive accuracy of root type estimation in continuous neutral two-type Yule models

Consider a time-independent, neutral continuous pure-birth process with mutations, where offsprings are always of the parent type. In this model the overall splitting rates are given by $r = \tilde{q}_1^{11} = \tilde{q}_2^{22}$ and $\tilde{q}_1^{12} = \tilde{q}_2^{12} = \tilde{q}_1^{22} = \tilde{q}_2^{11} = 0$, and the mutation rates are $\tilde{q}_1^2 = \tilde{q}_2^1$.

We further assume that the root type has certain non-trivial distribution. Gascuel & Steel [13] study the reconstruction of the types of the root and internal nodes of these trees given the leaves' types.

For a method \mathcal{M} for finding the root type of a random continuous tree given the leaves' types at a certain time $t \geq 0$, its *predictive accuracy* is defined as the expected value over all sample trees of the probability that the predicted root type is correct. The same concept can be defined for methods to predict the type of an internal node selected uniformly at random. A well known method is the *maximum parsimony* method, which estimates the type at a node to be the one that minimizes the number of substitutions (type changes from parent to child) needed to explain the evolution of the types observed at the leaves. This method only takes into account the tree topology, but not its transition rates or branch lengths. Another common method used is the *majority rule* method that

assigns to the root type the most frequent type at the leaves.

In the two-type case, a trivial lower bound that can be set for the predictive accuracy of any method is $1/2$ which results from choosing types uniformly at random.

This particular two-type case is *symmetric* because $\tilde{q}_i^{j_1 j_2} = \tilde{q}_i^{j'_1 j'_2}$, $\forall i, j_1, j_2$ where $j' = 1$ if $j = 2$ and viceversa. Assume further that it is *stationary*, meaning that the root type (and thus every other node's type) follows the stationary distribution of the corresponding transition matrix.

Using the reconstruction results of Corollary 3.11, we can rewrite Proposition 5 of [13] as follows,

Corollary 3.12. *Consider the time-independent continuous neutral model defined above starting at time t (from a single lineage) assumed to be symmetric and stationary, then as t grows;*

- *The predictive accuracy of the maximum parsimony method for estimating the root state of the tree converges to the trivial bound $1/2$ if and only if the asymptotic proportion of the mean numbers of cherries and pendant edges satisfy*

$$\frac{2w_1^{11} + w_1^{12} - w_1^2}{w_1^1 - w_1^2} \geq \frac{1}{6}.$$

- *The predictive accuracy of any method for estimating the root type converges to the trivial bound if the asymptotic proportion of the mean numbers of cherries and pendant edges satisfy*

$$\frac{2w_1^{11} + w_1^{12} - w_1^2}{w_1^1 - w_1^2} \geq \frac{1}{4}.$$

Proof. By Proposition 5 from [13], the first claim is true if and only if the speciation rate is less than 6 times the substitution rate ($\tilde{q}_1^2 = \tilde{q}_2^1$), while the second claim is true if the speciation rate is less than 4 times the substitution rate. So we just need to show that

these rates can be written in terms of the asymptotic proportion of cherries and pendant edges.

The speciation rate is given by $\tilde{q}_1^{11} = \tilde{q}_2^{22} = r$, while the substitution rate is $\tilde{q}_1^2 = \tilde{q}_2^1$.

We would like to write \tilde{q}_1^2 in terms of the asymptotic proportion of cherries and pendant edges. With the notation of Corollary 3.11, since $\tilde{q}_1^2 = \tilde{q}_2^1$, we have that;

$$\mathbf{C} - r\mathbf{I} = \begin{bmatrix} -2r - \tilde{q}_1^2 & \tilde{q}_1^2 & 0 & 0 \\ \tilde{q}_1^2 & -2r - \tilde{q}_1^2 & 0 & 0 \\ 0 & 0 & -2r - \tilde{q}_1^2 & \tilde{q}_1^2 \\ 0 & 0 & \tilde{q}_1^2 & -2r - \tilde{q}_1^2 \end{bmatrix}.$$

$$\mathbf{U} = \begin{bmatrix} 2r & r & 0 & 0 & 0 & 0 \\ 0 & r & 2r & 0 & 0 & 0 \\ 0 & 0 & 0 & 2r & r & 0 \\ 0 & 0 & 0 & 0 & r & 2r \end{bmatrix}.$$

Assuming we know the vector $\mathbf{w}^\ell = [w_\ell^{11}, w_\ell^{12}, w_\ell^{22}]$ for $\ell \in \{1, 2\}$ and $\mathbf{w}^\star = [w_1^1, w_1^2, w_2^1, w_2^2]$, we can solve the following system for the rate \tilde{q}_1^2 ,

$$(\mathbf{C} - r\mathbf{I})\mathbf{w}^\star + \mathbf{U}\mathbf{w} = \mathbf{0},$$

Hence,

$$\tilde{q}_1^2 = \frac{r(2w_1^{11} + w_1^{12} - w_1^2)}{w_1^1 - w_1^2}.$$

Therefore the inequalities $r < 6\tilde{q}_1^2$ and $r < 4\tilde{q}_1^2$ become

$$\frac{2w_1^{11} + w_1^{12} - w_1^2}{w_1^1 - w_1^2} > \frac{1}{6} \text{ and } \frac{2w_1^{11} + w_1^{12} - w_1^2}{w_1^1 - w_1^2} > \frac{1}{4}, \text{ respectively.}$$



Mossel & Steel [38] proved the converse implication for the majority rule method, concluding that its predictive accuracy converges to the trivial bound if and only if $r \leq \tilde{q}_1^2$. From that result we immediately deduce the following corollary;

Corollary 3.13.

The predictive accuracy of the majority rule method for estimating the root type converges to the trivial bound if and only if the asymptotic proportion of the mean numbers of cherries and pendant edges satisfy

$$\frac{2w_1^{11} + w_1^{12} - w_1^2}{w_1^1 - w_1^2} \geq \frac{1}{4}.$$

This means that the majority rule method performs better than the maximum parsimony method for the symmetric stationary model, as long as the fraction above is between 1/6 and 1/4.

3.2.3 Monotonicity of the asymptotic proportion of cherries in certain two-type models

In this section we show monotonicity results for the asymptotic proportion of cherries of each type in two particular two-type models. In the first model the monotonicity is obtained in terms of the weight \hat{w}_1 , and in the second one this weight is fixed $\hat{w}_1 = 1/2$ and the monotonicity is in terms of the splitting probability q_1^{11} .

Consider the two-type continuous birth process with splitting rates $\tilde{q}_i^{j_1, j_2}$ ($i, j_1, j_2 \in \{1, 2\}, j_1 \leq j_2$), and no mutations. This process has the same mean asymptotic proportion of cherries of each type as the discrete Yule process with splitting probabilities,

$$q_1^{11} = \frac{\tilde{q}_1^{11}}{\tilde{q}_1}, \quad q_1^{12} = \frac{\tilde{q}_1^{12}}{\tilde{q}_1}, \quad q_1^{22} = \frac{\tilde{q}_1^{22}}{\tilde{q}_1},$$

$$q_2^{11} = \frac{\tilde{q}_2^{11}}{\tilde{q}_2}, \quad q_2^{12} = \frac{\tilde{q}_2^{12}}{\tilde{q}_2}, \quad q_2^{22} = \frac{\tilde{q}_2^{22}}{\tilde{q}_2},$$

and where the probability of choosing a leaf of type i at each step is weighted by the ratio $\hat{w}_i = \frac{\tilde{q}_i}{\tilde{q}_1 + \tilde{q}_2}$. Observe that $\hat{w}_1 + \hat{w}_2 = 1$. This process is neutral when $\hat{w}_1 = 1/2$ ($\tilde{q}_1 = \tilde{q}_2$). We do not restrict ourselves to the neutral case. In fact, we seek to find a relationship between the weight \hat{w}_1 and the asymptotic proportions $w_\ell^{j_1 j_2}$ of the mean number of cherries of type $\ell j_1 j_2$ ($j_1 \leq j_2$) when the splitting probabilities satisfy the following relation

$$q_1^{11} + q_2^{22} = 1 + q_1^{22} + q_2^{11}. \quad (3.10)$$

This relationship is not far fetched. Indeed, in the process where siblings are independent amongst each other and are decided from their parent's type according to a Markov process with transition probabilities $(s_{ij})_{i,j \in \{1,2\}}$, the splitting probabilities are given by;

$$q_1^{11} = (1 - s_{12})^2, \quad q_1^{12} = 2(1 - s_{12})s_{12}, \quad q_1^{22} = (s_{12})^2,$$

$$q_2^{22} = (1 - s_{21})^2, \quad q_2^{12} = 2(1 - s_{21})s_{21}, \quad q_2^{11} = (s_{21})^2,$$

and thus condition (3.10) becomes the equality $s_{12} + s_{21} = 1/2$. The neutral case $\hat{w}_1 = 1/2$ of this two-type Yule process is the one studied in Section 3.1.4, and by Janson & Mossel [21].

From Theorem 3.6, we know that the vector $\mathbf{w}^\ell = [w_\ell^{11}, w_\ell^{12}, w_\ell^{22}]$ can be written as follows,

$$\mathbf{w}^\ell = \lim_{t \rightarrow \infty} \boldsymbol{\eta}_t^\ell = -u_\ell (\mathbf{A} - \lambda \mathbf{I})^{-1} \tilde{\mathbf{q}}^{(\ell)},$$

By replacing $\tilde{q}_i^{j_1 j_2} = q_i^{j_1 j_2} \hat{w}_i (\tilde{q}_1 + \tilde{q}_2)$, and recalling that $\tilde{q}_i^j = 0$, this expression may be written entirely in terms of the splitting probabilities $q_i^{j_1 j_2}$ and weights \hat{w}_i ($i, j \in \{1, 2\}$), and by replacing $\hat{w}_2 = 1 - \hat{w}_1$, $q_\ell^{12} = 1 - q_\ell^{11} - q_\ell^{22}$ for $\ell \in \{1, 2\}$ and condition (3.10), we attain expressions for the mean asymptotic proportion of cherries which depend only on

\hat{w}_1 and $g_\ell^{j_1 j_2}$ for $\ell, j_1 = j_2 \in \{1, 2\}$.

These calculations are straightforward but rather chaotic, and were obtained using the algebra system Maple. However, the formulas simplify greatly when we assume condition (3.10). We use the following notation to avoid confusion between weights and the proportion of cherries;

$$w_\ell^{j_1 j_2} = g_\ell^{j_1 j_2}(\hat{w}_1), \text{ for } \ell, j_1, j_2 \in \{1, 2\}, j_1 \leq j_2.$$

Notice that even though $g_\ell^{j_1 j_2}$ depends also on the splitting probabilities we do not write it explicitly to simplify the notation. For this specific case we obtain the following monotonicity result,

Proposition 3.6. *For the time independent continuous-time two-type birth-death model with no mutations such that its splitting probabilities satisfy equation (3.10), the following monotonicity for the asymptotic proportion of cherries is satisfied in terms of the weight \hat{w}_1 for $1 \leq j_1 \leq j_2 \leq 2$;*

$$\hat{w}_1 < \hat{w}'_1 \quad \Rightarrow \quad g_1^{j_1 j_2}(\hat{w}_1) < g_1^{j_1 j_2}(\hat{w}'_1), \quad g_2^{j_1 j_2}(\hat{w}_1) > g_2^{j_1 j_2}(\hat{w}'_1).$$

Proof. From Theorem 3.6, as outlined above, we obtain the following expressions from straightforward calculations using the algebra system Maple:

$$\begin{aligned} g_1^{11}(\hat{w}_1) &= \frac{\hat{w}_1 q_1^{11} (q_1^{11} - q_1^{22})}{2q_1^{11} \hat{w}_1 + \hat{w}_1 - 2q_1^{22} \hat{w}_1 - q_1^{11} + q_1^{22} + 1}, \\ g_1^{12}(\hat{w}_1) &= \frac{\hat{w}_1 q_1^{12} (q_1^{11} - q_1^{22})}{2q_1^{11} \hat{w}_1 - \hat{w}_1 - 2q_1^{22} \hat{w}_1 - q_1^{11} + q_1^{22} + 2}, \\ g_1^{22}(\hat{w}_1) &= \frac{\hat{w}_1 q_1^{22} (q_1^{11} - q_1^{22})}{2q_1^{11} \hat{w}_1 - 3\hat{w}_1 - 2q_1^{22} \hat{w}_1 - q_1^{11} + q_1^{22} + 3}, \end{aligned}$$

$$\begin{aligned}
g_2^{11}(\hat{w}_1) &= \frac{q_2^{11}(1 - q_1^{11} + q_1^{22})(1 - \hat{w}_1)}{2q_1^{11}\hat{w}_1 + \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 1}, \\
g_2^{12}(\hat{w}_1) &= \frac{q_2^{12}(1 - q_1^{11} + q_1^{22})(1 - \hat{w}_1)}{2q_1^{11}\hat{w}_1 - \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 2}, \\
g_2^{11}(\hat{w}_1) &= \frac{q_2^{22}(1 - q_1^{11} + q_1^{22})(1 - \hat{w}_1)}{2q_1^{11}\hat{w}_1 - 3\hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 3},
\end{aligned}$$

In order to prove the monotonicity of $g_\ell^{j_1 j_2}(\hat{w}_1)$ for $\ell, j_1, j_2 \in \{1, 2\}$, it suffices to prove that their first derivate with respect to \hat{w}_1 is positive for $\ell = 1$ and negative for $\ell = 2$.

Indeed,

$$\begin{aligned}
\frac{\partial g_1^{11}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{q_1^{11}(q_1^{11} - q_1^{22})(1 - q_1^{11} + q_1^{22})}{(2q_1^{11}\hat{w}_1 + \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 1)^2} > 0, \\
\frac{\partial g_1^{12}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{q_1^{12}(q_1^{11} - q_1^{22})(2 - q_1^{11} + q_1^{22})}{(2q_1^{11}\hat{w}_1 - \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 2)^2} > 0, \\
\frac{\partial g_1^{22}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{q_1^{22}(q_1^{11} - q_1^{22})(3 - q_1^{11} + q_1^{22})}{(2q_1^{11}\hat{w}_1 - 3\hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 3)^2} > 0, \\
\frac{\partial g_2^{11}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{-q_2^{11}(2 - (q_1^{11} - q_1^{22})(1 + q_1^{11} - q_1^{22}))}{(2q_1^{11}\hat{w}_1 + \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 1)^2} < 0, \\
\frac{\partial g_2^{12}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{-q_2^{12}(1 - (q_1^{11} - q_1^{22})^2)}{(2q_1^{11}\hat{w}_1 - \hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 2)^2} < 0, \\
\frac{\partial g_2^{22}(\hat{w}_1)}{\partial \hat{w}_1} &= \frac{-q_2^{22}(q_1^{11} - q_1^{22})(1 - q_1^{11} + q_1^{22})}{(2q_1^{11}\hat{w}_1 - 3\hat{w}_1 - 2q_1^{22}\hat{w}_1 - q_1^{11} + q_1^{22} + 3)^2} < 0,
\end{aligned}$$

and the result follows. ◉

Notice that when $\hat{w}_1 = 1/2$, we obtain the same proportions as in Corollary 3.2, as expected. Also, we have the following Corollary concerning the comparison between neutral and non-neutral models.

Corollary 3.14. *The asymptotic proportion of the mean number of cherries of type 1 (respectively type 2) parent for the neutral ($\hat{w}_1 = 1/2$) two-type Yule model above is greater than (respectively less than) the same asymptotic proportion for the non-neutral model with $\hat{w}_1 < 1/2$, and less than (respectively greater than) the proportion for the*

non-neutral model with $\hat{w}_1 > 1/2$.

Let us now consider the time-independent, neutral continuous pure-birth process with mutations from Section 3.2.2 and studied by Gascuel & Steel [13], where offsprings are always of the parent type, but may mutate at certain rate. In this model the overall splitting rates are given by $r = \tilde{q}_1^{11} = \tilde{q}_2^{22}$ and $\tilde{q}_1^{12} = \tilde{q}_2^{12} = \tilde{q}_1^{22} = \tilde{q}_2^{11} = 0$, and the mutation (substitution) rates are $\tilde{q}_1^2 = \tilde{q}_2^1$. Thus the event probabilities are given as follows,

$$q_1^{11} = q_2^{22} = \frac{r}{r + \tilde{q}_1^2}, \quad q_1^2 = q_2^1 = \frac{\tilde{q}_1^2}{r + \tilde{q}_1^2} = 1 - q_1^{11}.$$

Notice that in this case the weight \hat{w}_1 is equal to $1/2$, and so we can write the asymptotic proportion of cherries in terms of $q_1^{11} = q$ by using Theorem 3.6. As before we use the notation

$$w_\ell^{j_1 j_2} = g_\ell^{j_1 j_2}(q), \text{ for } \ell, j_1, j_2 \in \{1, 2\}, j_1 \leq j_2.$$

We obtain the following result,

Proposition 3.7. *For the time-independent, neutral continuous two-type pure-birth process with splitting probability $q = q_1^{11} = q_2^{22}$ and mutation probability $1 - q = q_1^2 = q_2^1$, the following monotonicity for the asymptotic proportion of cherries is satisfied in terms of q ;*

$$q < q' \quad \Rightarrow \quad g_1^{11}(q) < g_1^{11}(q'), \quad g_1^{j_1 j_2}(q) > g_1^{j_1 j_2}(q'), \text{ for } j_1 \in \{1, 2\}.$$

The same thing holds for $g_2^{j_1 j_2}(q)$ by the symmetry of this case.

Proof. From Theorem 3.6, we obtain the following expressions from straightforward calculations using the algebra system Maple:

$$g_1^{11}(q) = \frac{2 + 8q - q^2}{6(8 + 2q - q^2)}, \quad g_1^{12}(q) = \frac{1 - q}{3(4 - q)}, \quad g_1^{22}(q) = -\frac{(1 - q)^2}{3(8 + 2q - q^2)}.$$

with corresponding derivatives

$$\frac{\partial g_1^{11}(q)}{\partial q} = \frac{q^2 - 2q + 10}{(8 + 2q - q^2)^2} > 0, \quad \frac{\partial g_1^{12}(q)}{\partial q} = -\frac{1}{(4 - q)^2} < 0, \quad \frac{\partial g_1^{22}(q)}{\partial q} = \frac{6q - 6}{(8 + 2q - q^2)^2} < 0.$$

which implies the monotonicity results. 

This result may be explained intuitively; as the probability $q_1^{11} = q_2^{22} = q$ increases, one would expect to have more cherries of types 111 and 222. And since this also means that the mutation probability $1 - q$ becomes smaller, the expected number of cherries of types 112, 212, 122 and 211 should decrease.

Chapter 4

Conclusion and outlook

Our work on the multi-type coalescent point process provided a method for reconstructing multi-type ancestral trees under certain assumptions by extending previously known results on single-type processes. This introduced a wider and more accurate framework for the investigation of ancestral evolutionary trees, since different traits are likely to yield different reproductive patterns in real-world biological applications.

Lemma 2.1 provides an independence property for subtrees of a multi-type branching process under conditions related to their descendants at the present time. This Lemma helped in the proof of Theorem 2.1, which provides a new backward algorithm for the construction of the multi-type coalescent point processes. Corollaries 2.1, 2.2, and Propositions 2.1, 2.2, give expressions for the distribution of the coalescent times of two consecutive individuals and the coalescent times of same type individuals in the multi-type coalescent point process. Propositions 2.3, 2.4 specialize these results to the case of the multi-type linear fractional coalescent point process. Corollary 2.3 and Proposition 2.5 give more specific formulae for particular linear fractional cases. Future research on the multi-type coalescent point process could focus on extending this backward construction to avoid the need for prior assumptions on the types of the ancestors of the standing population's leftmost individual (the spine of the process), possibly in the context of

particular cases for which the quasi-stationary distribution of the spine can be derived explicitly.

Our formulae for the asymptotic proportion of cherries and pendant edges of different types are useful for the reconstruction of ancestral features of multi-type birth models, since in many interesting cases it is possible to revert these formulae to obtain approximations for the speciation and mutations rates of the model. This is particularly important because cherries are formed by the standing population and their immediate ancestors, and because the proportion of types in the standing population alone does not provide enough information for such reconstruction. As corollaries to these formulae, it is possible to determine the solvability and predictive accuracy of certain models in terms of their asymptotic proportion of cherries of each type.

In the context of random rooted binary trees under the neutral two-type Yule model, this work provided explicit formulae and asymptotic properties for means and variances of the numbers of leaves, cherries and pendant edges of different types (Propositions 3.1, 3.2, 3.4, 3.3 and Corollaries 3.1, 3.2, 3.3).

As a way to present a motivation for the study of multi-type Yule trees with mutations, we have pointed out that type mutations appear naturally in the ancestral trees of multi-type birth-death process, and we have proved that these ancestral trees are themselves multi-type pure-birth processes, deriving their time-dependent rates and their defining probabilities explicitly (Lemma 3.1, Theorem 3.2 and Corollary 3.6). Subsequently we have provided some differential equations and asymptotic formulae for the mean number of leaves, cherries and pendant edges of each type, as well as formulae for reconstructing rates from these means in a continuous-time constant-rate multi-type pure-birth process with mutations (Theorems 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, Proposition 3.5, and their corollaries). This continuous-time process embeds and has the same asymptotic properties as a multi-type Yule tree with mutations under certain relations between the parameters of

each process. From this reconstruction we have obtained Corollaries 3.5, 3.12, 3.13 and 3.14 and Propositions 3.6 and 3.7, providing certain reversibility properties of particular two-type models by exploiting the relationship between the mean asymptotic proportion of cherries of each type and the defining parameters of those models. As part of future research on multi-type pure-birth processes with mutations we could explore particular time-dependent rates for which the statistics on cherries and pendant edges can be calculated explicitly, and test the reconstructive properties of these processes by recovering their defining parameters from simulated phylogenetic trees.

Bibliography

- [1] David Aldous. Probability distributions on cladograms. *Random Discrete Structures*, 76:1–18, 1995.
- [2] David Aldous and Lea Popovic. A critical branching process model for biodiversity. *Advances in Applied Probability*, 37(4):1094–1115, 2005.
- [3] Krishna B. Athreya. Some results on multitype continuous time Markov branching processes. *Annals of Mathematical Statistics*, 39:347–357, 1968.
- [4] Krishna B. Athreya and Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Annals of Mathematical Statistics*, 39:1801–1817, 1968.
- [5] Krishna B. Athreya and Peter E. Ney. *Branching Processes*. Dover Publications Inc., Mineola, NY, 2004.
- [6] Sandrine Dallaporta and Anatole Joffe. The Q -process in a multitype branching process. *International Journal of Pure and Applied Mathematics*, 42(2):235–240, 2008.
- [7] Jean-Stéphane Dhersin and Laurent Serlet. A stochastic calculus approach for the Brownian snake. *Canadian Journal of Mathematics. Journal Canadien de Mathématiques*, 52(1):92–118, 2000.

- [8] Filippo Disanto and Thomas Wiehe. Exact enumeration of cherries and pitchforks in ranked trees under the coalescent model. *Mathematical Biosciences*, 242(2):195–200, 2013.
- [9] Jay D. Evans and Dawn Gundersen-Rindal. Beenomes to bombyx: future directions in applied insect genomics. *Genome Biology*, 4:107–110, 2003.
- [10] Richard G. Fitzjohn, Wayne P. Maddison, and Sarah P. Otto. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, 57:595–611, 2009.
- [11] Simon D. W. Frost and Erik M. Volz. Modelling tree shape and structure in viral phylodynamics. *Philosophical transactions of the Royal Society B*, 368(1614), 2013.
- [12] Francis Galton and Henry W. Watson. On the probability of the extinction of families. *Journal of the Anthropological Institute of Great Britain*, 4:138–144, 1875.
- [13] Olivier Gascuel and Mike Steel. Predicting the ancestral character changes in a tree is typically easier than predicting the root state. *Systematic Biology*, 63(3):421–435, 2014.
- [14] Jochen Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *Journal of Applied Probability*, 36(2):301–309, 1999.
- [15] Hans-Otto Georgii and Ellen Baake. Supercritical multitype branching processes: the ancestral types of typical individuals. *Advances in Applied Probability*, 35(4):1090–1110, 2003.
- [16] Tanja Gernhard. The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4):769–778, 2008.

- [17] Emma Goldberg and Boris Igc. On phylogenetic tests of irreversible evolution. *Evolution*, 62(11):2727–2741, 2008.
- [18] Emma Goldberg and Boris Igc. Tempo and mode in plant breeding system evolution. *Evolution*, 66(12):3701–3709, 2012.
- [19] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, third edition, 2001.
- [20] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110(2):177–245, 2004.
- [21] Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *Annals of Probability*, 32:2630–2649, 2004.
- [22] Anatole Joffe and Gérard Letac. Multitype linear fractional branching processes. *Journal of Applied Probability*, 43(4):1091–1106, 2006.
- [23] Anatole Joffe and Frank Spitzer. On multitype branching processes with $\rho \leq 1$. *Journal of Mathematical Analysis and Applications*, 19:409–430, 1967.
- [24] Graham Jones. Calculations for multi-type age-dependent binary branching processes. *Journal of Mathematical Biology*, 63(1):33–56, 2011.
- [25] Marek Kimmel and David E. Axelrod. *Branching Processes in Biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, 2002.
- [26] Thomas Kurtz, Russell Lyons, Robin Pemantle, and Yuval Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In *Classical and Modern Branching Processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 181–185. Springer, New York, 1997.

- [27] Amaury Lambert. Population dynamics and random genealogies. *Stochastic Models*, 24(suppl. 1):45–163, 2008.
- [28] Amaury Lambert. The contour of splitting trees is a Lévy process. *The Annals of Probability*, 38(1):348–395, 2010.
- [29] Amaury Lambert and Lea Popovic. The coalescent point process of branching trees. *The Annals of Applied Probability*, 23(1):99–144, 2013.
- [30] Jean-François Le Gall. A class of path-valued Markov processes and its applications to superprocesses. *Probability Theory and Related Fields*, 95(1):25–46, 1993.
- [31] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3):1125–1138, 1995.
- [32] Wayne P. Maddison, Peter E. Midford, and Sarah P. Otto. Estimating a binary character’s effect on speciation and extinction. *Systematic Biology*, 56(5):701–710, 2007.
- [33] Karen Magnuson-Ford and Sarah P. Otto. Linking the investigations of character evolution and species diversification. *The American Naturalist*, 180(2):225–245, 2012.
- [34] Andy McKenzie and Mike Steel. Distributions of cherries for two models of trees. *Mathematical Biosciences*, 164(1):81–92, 2000.
- [35] Sylvie Méléard and Denis Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9:340–410, 2012.
- [36] Arnau Mir, Francesc Rosselló, and Lucia Rotger. A new balance index for phylogenetic trees. *Mathematical Biosciences*, 241(1):125–136, 2013.

- [37] Charles J. Mode. Some multi-dimensional birth and death processes and their applications in population genetics. *International Biometric Society*, 18(4):543–567, 1962.
- [38] Elchanan Mossel and Mike Steel. Majority rule has transition ratio 4 on yule trees under a 2-state symmetric model. *Preprint, arXiv:1404.2681v1*, pages 1–6, 2014.
- [39] Sean Nee, Robert H. May, and Paul H. Harvey. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B*, 344(1309):305–311, 1994.
- [40] Lea Popovic. Asymptotic genealogy of a critical branching process. *The Annals of Applied Probability*, 14(4):2120–2148, 2004.
- [41] Serik Sagitov. Linear-fractional branching processes with countably many types. *Stochastic Processes and their Applications*, 123(8):2940–2956, 2013.
- [42] E. Seneta and D. Vere-Jones. On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *Journal of Applied Probability*, 3:403–434, 1966.
- [43] Eugene Seneta. *Non-Negative Matrices and Markov Chains*. Springer Series in Statistics. Springer, New York, 2006.
- [44] Robert T. Smythe. Central limit theorems for urn models. *Stochastic Processes and their Applications*, 65(1):115–137, 1996.
- [45] Tree of life Web project. Genetic connections between organisms. <http://tolweb.org/tree/learn/concepts/geneticconnections.html>, April 2006.