# Counting Empirical Cluster Sizes Of Identical COVID-19 Genetic Sequences

by

Sjoerd van der Niet

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday July 2, 2024 at 1:00 PM.

| | | | |
|---|---|---|---|
| Student number: | 5182417 | | |
| Project duration: | January 1, 2024 – July 2, 2024 | | |
| Thesis committee: | dr. J. Komjáthy | TU Delft | chair, supervisor |
| | Prof. dr. G. Jongbloed | TU Delft | |

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

# Abstract

This thesis aims to enhance existing models that infer parameters describing the spread of a virus by analyzing the distribution of empirical cluster sizes of identical genetic sequences. An approach that has gained recent popularity assumes that each individual cluster can be modeled as a Bienaymé-Galton-Watson process, with the distribution of empirical cluster sizes being equal to the law of the final size $\widetilde{Y}_\infty$ of the branching process. By employing the theory of general branching processes counted by characteristics, we demonstrate that the empirical cluster size distribution $C^\alpha$ stochastically dominates $\widetilde{Y}_\infty$ due to the exponential growth of the branching process. Under the assumption that the underlying branching tree follows either a Bienaymé-Galton-Watson process or an age-dependent process, we show that the mean of the empirical cluster size distribution can be used for a (strongly) consistent estimator for the probability of mutation $\nu$. For both branching models, we compute $\mathbb{P}(C^\alpha = n)$ for $n = 1, 2$. We conjecture that $\mathbb{P}(C^\alpha = n)$ is independent of the underlying model and that it can be expressed as a function of the mean of the offspring distribution $X$, and the probability mass function of $\mathrm{BIN}(X, 1-\nu)$. An extension of the model is considered where the probability of mutation is sampled from a distribution $\nu$ for each cluster. We show that under this assumption the empirical mean of the cluster sizes estimates the quantity $\int \nu^{-1}(r)\,\mathrm{d}r$. We also show that the $\nu$ can still be estimated by the empirical mean of the cluster sizes, when the population is divided into a finite number of types with inhomogeneous offspring distributions.

# Contents

# Chapter 1

# Introduction

The importance of epidemic modelling dates back to at least the eighteenth century. Daniel Bernoulli studied the effects of variolation, the predecessor of modern vaccination against smallpox, on the mortality rates associated with the disease [3]. He used empirical data to support his model and advice on public health policies. Using a mathematical framework to address a public health concern marked a pivotal moment in the history of epidemiology.

The next benchmark in epidemic modelling was established by two British scientists in 1927 [23]. While they were not the first to divide the population into compartments based on infection status [30], their formulation of the SIR (Susceptible-Infectious-Recovered) compartment model, accompanied by a system of differential equations for simulating epidemics, became a foundational framework for future research. Moreover, the concept of the basic reproduction number ($R_0$) was introduced, which is defined as the average number of secondary infections by a single infected individual, in a fully susceptible population.

The basic reproduction number plays a crucial role in understanding and controlling infectious diseases, as it serves as an indicator for the epidemic potential and guides the need for public health interventions [10]. Although $R_0$ is not a biological constant, as it may differ for the same pathogen[1] in distinct populations and has to be interpreted with caution, it remains critically important [8].

In the 1950s it became evident that randomness plays a crucial role in epidemic spread. A key figure in this movement was Bartlett, who formulated the SIR compartment model in terms of transition rates instead of differential equations [2], resulting in a Markov chain. By understanding the stochastic nature of disease spread, the ability to predict the course of an epidemic could be improved.

At this point, it was not a significant leap to utilize the theory of branching processes to model epidemics [19]. This theory originated from the works of Bienaymé in 1845 [4], and Galton and Watson in 1874 [36], both of which aimed to investigate the survival chances of family names. While the latter study is more widely recognized, we refer to the model defined in Section 2.2, which is considered in both studies, as a Bienaymé-Galton-Watson process to acknowledge both contributions.

The Bienaymé-Galton-Watson process is a simple model compared to the vast generalized theory on branching processes. Nevertheless, it incorporates the stochastic elements of an epidemic better than earlier mentioned models. Unlike models that only consider the evolution of an epidemic due to a fixed rate, the Bienaymé-Galton-Watson process captures elements caused by random effects that occur in the early stages of an epidemic. In the Bienaymé-Galton-Watson

[1] A pathogen is defined as an organism causing disease to its host.

model, each infected individual produces an independent number of random infected individuals, according to some distribution. The exponential growth which arises in the earlier model still persists if the distribution is chosen appropriately. Moreover, properties that arise due to the stochastic nature of the model can be studied, such as the probability of extinction, i.e. the probability that a spreading virus does not become an epidemic.

Moving forward in time, the inference of $R_0$ under various models continues to be a persisting goal in research [7, 9, 10, 15]. As a result, many studies assume a branching process driving the spread of an infectious disease [5, 12, 25, 28, 35]. Advances in the field of DNA sequencing enabled researchers to gather large amounts of genetic data sampled from infectious patients. As a result, models can incorporate the availability of this data.

When a pathogen possesses the property that its genetic code mutates over time, clusters of identical sequences appear in various sizes. This approach is used in [35] to infer model parameters, such as $R_0$ and a variance controlling parameter $k$. By considering a branching process and a reproduction number $pR_0$, where $p$ is the probability that an infected individual has an identical sequence to its infector, the distribution of cluster sizes is used to perform maximum likelihood inference. This model is in fact the result of a branching process equipped with the infinite alleles model [24], and the clusters can be modelled as branching processes embedded in the original tree.

This thesis is the result of an attempt for inferring age-contact matrices[2] from epidemiological data. The methods of [35] inspired us to adapt the techniques in such a way that contacts between groups can be inferred. However, it was suspected that the aforementioned methods were based on incorrect assumptions, which could lead to biased estimates. One of these assumptions is that each observed cluster has reached its final size and the sizes of the clusters can be modelled as stand-alone branching processes. As mentioned before, the clusters appear inside a larger tree. For this reason, if cluster sizes are observed at a certain time, inside a growing infection tree, then the sample is more likely to contain small cluster sizes. This reasoning is made visible in Figure 1.1.

Regarding the application of the Bienaymé-Galton-Watson model by [35], the machinery provided by the classic theory of branching processes appears to be too limited. Resolving these shortcomings is the main motivation for this thesis. Properties of the infinite alleles branching process have already been studied in various settings [11, 17, 26, 29, 37]. Motivated by [37], where general branching processes counted by characteristics are used to derive results on the frequency spectrum[3], we aim to adapt these techniques and apply them to derive an expression for the distribution of the clusters, as they appear in the larger infection tree.

The narrative of this thesis is divided in the following chapters. In Chapter 2 we formally introduce the Bienaymé-Galton-Watson process and equip it with the infinite alleles model. The derivations lead up to the probability mass function stated in Corollary 1 as it appears in [35], and the chapter concludes with a discussion that motivates the use of general branching processes. The theory given in Chapter 2 also serves as a steppingstone for the subsequent chapters.

In Chapter 3, the main theory is established. Basing our notation on that of [21], we introduce the Ulam-Harris family history space. This space forms the foundation for constructing the probability space. Each element from the sample space contains information on every conceivable individual. Using this model, we define measurable functions on the whole tree of conceivable individuals. The outcome of these functions, which we define as characteristics, depend on the information provided by the sample. As a result, we construct characteristics that take the

[2] An age-contact matrix $C = (c_{ij})$ is a representation for the intensity of contacts between different age groups within a population. Here, $c_{ij}$ denotes the average number of contacts for an individual in age group $i$ with individuals from age group $j$. The time interval for this average dependents on the specific model and context.

[3] Within the infinite alleles framework, the frequency spectrum if often referred to as the vector $(\alpha_j(t))_j$ where $\alpha_j(t)$ is the number of alleles represented by $j$ individuals at time $t$.

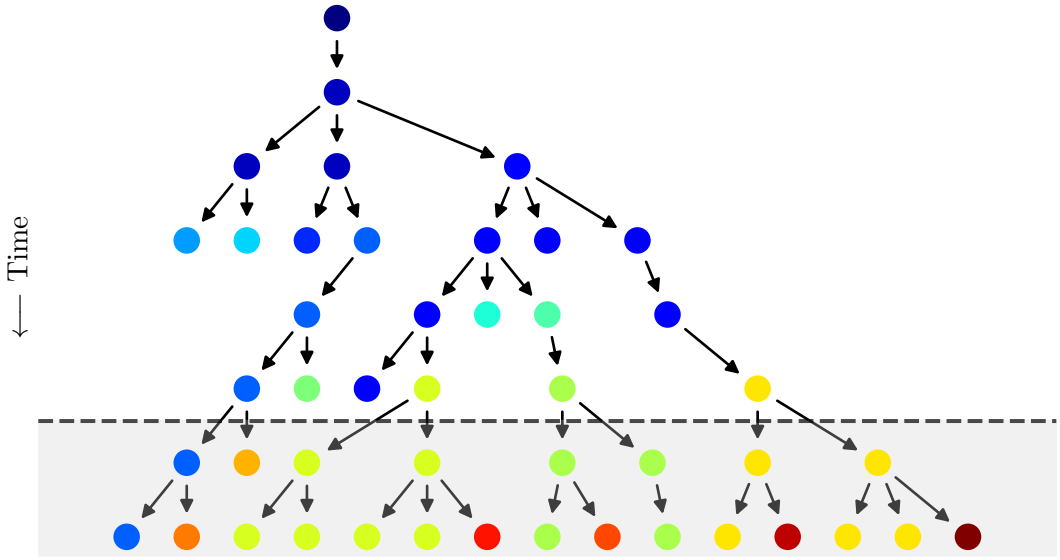


Figure 1.1: An example of a Bienaymé-Galton-Watson process realization equipped with the infinite alleles model. The dashed line denotes the time when the process is observed. If the time of observation is before the time of extinction, it is possible that some clusters have yet to reach their final size. This illustration demonstrates that, assuming all clusters have reached their final size, larger clusters are disproportionately overrepresented in the sample, while smaller clusters are underrepresented.

growth of the infection tree into account. Provided that the branching process is supercritical, i.e. it grows exponentially in size, we may apply results from [21, 27] to obtain analogues for the results derived in Chapter 2. The analogues are for the Bienaymé-Galton-Watson process and the age-dependent branching process. In Theorem 5 we state that under the Bienaymé-Galton-Watson model, the newly found random variable $C^\alpha$ stochastically dominates the random variable counting the observed cluster sized found in Chapter 2.

The mean of the observed cluster sizes $C^\alpha$ for both models turns out to be independent of the offspring distribution, but solely depends on the mutation parameter $\nu$. The mean coincides for the two models. In Theorem 6 we give the main result of this chapter. It states that the sample mean of the observed cluster sizes $\overline{C_n^\alpha}$ gives a strongly consistent estimator for $\nu$. Moreover, we explicitly compute the probability mass function $\mathbb{P}(C^\alpha = n)$ for $n = 1, 2$, where the expressions again coincide for the two models. We conclude the chapter with Conjecture 1, which states that the expression for $\mathbb{P}(C^\alpha = n)$ is independent of the two models, and only depends on the offspring distribution.

In Chapter 4 we extend the single-type model we define in Chapter 3, where the individuals are assumed to be indistinguishable, to a multi-type model. The statements from Chapter 3 are reformulated under the assumption that individuals carry types from a type space $S$. We utilize this model to analyse the mean of the observed cluster sizes, under the assumption that

each cluster has a different probability of mutation. The mean of the observed cluster sizes is also studied in the context where individuals carry a type, which characterizes a type-dependent offspring distribution, i.e. the offspring distribution is not homogeneous among different types.

In Chapter 5 we apply the main results of Chapter 3. By means of a simulation study we show that in the supercritical regime, the statement in Theorem 5 is supported by the biased estimates of the model derived in Chapter 2. We also show that Theorem 6 holds in a similar simulation study.

# Chapter 2

# Classical branching processes

## 2.1 Model motivation

An infectious individual infects a random number of other individuals. We assume that every newly infected individual also infects other individuals according to the same probability distribution, which we refer to as the *offspring distribution*. This happens independent of all other individuals. Continuing this process gives a random infection tree, and these trees can be modelled as a branching process. In this section we model the tree as a classical Bienaymé-Galton-Watson process.

When an epidemic is the result of a spreading virus, the virus is characterised by its genetic sequence. Because the sequences mutate over time [31], infected individuals generally do not carry identical sequences. Assuming there exists a probability $\nu$ such that for each infection the newly infected individual is infected with a different genetic code than its infector, we can model the mutation process according to the infinite alleles model [24]. When the root produces offspring, each child assumes the same label as the parent, or the child carries a newly introduced unique label. We reserve the term *type* when the model is extend to a multi-type branching process. We make use of the property that the individuals carrying the root's label can be modelled by a branching process itself. The details are further explained in this section, which results in the derivation of the probability mass function for the number of individuals carrying the same label. We refer to the individuals with an identical label as a *cluster*.

As a result, one can perform statistical inference using the obtained probability mass function. By counting the sizes of clusters of identical genetic sequences, a likelihood function can be constructed to infer model parameters using a maximum likelihood estimation, which is done in [35]. In this section, the most simplistic model for counting cluster sizes is considered. In later sections we make the model more elaborate with the help of theory on general branching processes.

## 2.2 The Bienaymé-Galton-Watson process

The Bienaymé-Galton-Watson process is a nonnegative integer valued stochastic process $(Z_n)_{n\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, starting with one initial individual alive, $Z_0 = 1$, which we call the root. When we require that the branching process starts with $i$ individuals, i.e. $Z_0 = i$, we denote the process by $(Z_n^{(i)})_{n\in\mathbb{N}}$.

The root lives for one unit of time, and gives birth upon death to a random number of children according to the offspring distribution given by a random variable $X$, $Z_1 \sim X$. Now

every individual in the first generation, the children of the root, randomly produces offspring again upon death, after one unit of time. The offspring distribution of every individual in generation 1 is again given by $X$ and occurs independent of every other individual alive. Thus if $X_0 \sim X$ denotes the root's offspring, then $Z_2 = \sum_{i=1}^{X_0} X_{1,i}$, where $(X_{1,i})_{i\in\mathbb{N}}$ is a sequence of i.i.d. copies of $X$. The same logic applies to

$$Z_3 = \sum_{i=1}^{X_0} \sum_{j=1}^{X_{1,j}} X_{2,j} = \sum_{i=1}^{X_0} Z_{2,i},$$

where $(Z_{2,i})_{i\in\mathbb{N}}$ are now i.i.d. copies of $Z_2$. Thus the number of individuals in the third generation, $Z_3$, follows the same distribution as the sum of $X_0$ independent copies of $Z_2$. It is easily verified that this holds for any $n \geq 0$, which gives the characteristic property

$$Z_{n+1} = \sum_{i=1}^{X_0} Z_{n,i},$$

where $(Z_{n,i})_{i\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} Z_n$ and $X_0 \sim X$. This follows from the additive property [1].

**Proposition 1** (Additive property)**.** *The process $(Z_n^{(i)})_{n\in\mathbb{N}}$ with $i$ initial individuals is the sum of $i$ independent copies of the branching process $(Z_n)_{n\in\mathbb{N}}$.*

As mentioned at the beginning of this section, every newly introduced individual inherits its parent's label, or it introduces a new label with probability $\nu \in (0, 1)$. We refer to $\nu$ as the *probability of mutation*. Given an offspring distribution $X$, the number of individuals that inherits the label of the parent then follows a $\widetilde{X} \sim \text{BIN}(X, 1-\nu)$ distribution. Now every descendent carrying the same label gives birth to a random number of children with this label according to $\widetilde{X}$ too. As a result, the number of individuals carrying the same label, follows a branching process embedded in the initial tree of the branching process $(Z_n)_{n\in\mathbb{N}}$. We refer to this embedded branching process as $(\widetilde{Z}_n)_{n\in\mathbb{N}}$, where $\widetilde{Z}_0 = 1$ corresponds to the first individual to receive the label after a mutation. Every such embedding is an i.i.d. copy of the embedding of individuals carrying the *ancestral label*, which is the label assigned to the root. Because of this property, from this point forward, the branching process $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ with offspring distribution $\widetilde{X}$ is considered as a standalone process independent of $(Z_n)_{n\in\mathbb{N}}$.

We can use the process $(\widetilde{Z}_n)_{n\in\mathbb{N}}$ to obtain a distribution for the cluster sizes. When a new label is introduced with a newborn individual, we want to count all the descendent of this individual which carry this exact same label. From the following lemma we conclude that the descendants of an individual can be modelled as a branching process as well. The proof follows from applying the additive property.

**Lemma 1.** *Suppose that $\mathcal{T}$ is the random rooted tree generated by a branching process $(Z_n)_{n\in\mathbb{N}}$. For an individual $v$ in $\mathcal{T}$, let $\mathcal{T}(v)$ be the subtree rooted at $v$, pointing away from the root of $\mathcal{T}$. If $Z_n(v)$ denotes the size of $n$-th generation of $\mathcal{T}(v)$, then $(Z_n(v))_{n\in\mathbb{N}}$ is an i.i.d. copy of $(Z_n)_{n\in\mathbb{N}}$.*

Combining this with the fact that the descendants carrying the ancestral label are distributed as the branching process $(\widetilde{Z}_n)_{n\in\mathbb{N}}$, we need to count all the descendants of this process to find the distribution of the cluster sizes.

We define the *total size* up to generation $n$ of a branching process $(Z_n)_{n\in\mathbb{N}}$ as

$$Y_n = \sum_{k=0}^{n} Z_k, \tag{2.1}$$

for every $n \geq 0$. Letting $n \to \infty$, we define to the limiting random variable $Y_\infty = \lim_{n\to\infty} Y_n$ as the *total progeny* of $(Z_n)_{n\in\mathbb{N}}$.

The probability mass function is given by the following theorem. A proof can be found in [18] and is a result of the hitting time theorem for random walks.

**Theorem 1** (Law of total progeny)**.** *For a branching process with i.i.d. offspring distribution $X$*

$$\mathbb{P}(Y_\infty = n) = \frac{1}{n}\,\mathbb{P}(X_1 + \ldots + X_n = n-1), \tag{2.2}$$

*where $(X_i)_{i=1}^n$ are i.i.d. copies of $X$.*

Hence given an offspring distribution $X$, we can determine the probability mass function for the final sizes of a cluster of identical labels with Theorem 1.

In [25] multiple offspring distributions are considered and fitted on epidemiological data. Here they assume that every infected individual has an individual reproduction value given by a random variable $\rho$ with mean $R$, which denotes the average number of secondary cases. The secondary infections are modelled with a Poisson process such that the number of infections has distribution $\text{POIS}(\rho)$, with $\rho$ being random. When $\rho$ is assumed to have a Gamma distribution with mean $R$ and parameter $k$ such that $\text{Var}(\rho) = R^2/k$, then $\text{POIS}(\rho) \sim \text{NEGBIN}(k, q)$, where $q = k/(R+k)$. This choice of offspring distribution includes the conventional $\text{POIS}(R)$ and $\text{GEO}(q)$ when $k \to \infty$ and $k = 1$ respectively.

Now let $X \sim \text{NEGBIN}(k, q)$, where $k > 0$ and $q \in (0, 1)$. We use a generalized parametrization of the negative binomial which allows positive real values for $k$ instead of just integers. The probability mass function is then given by

$$\mathbb{P}(X = n) = \frac{\Gamma(k+n)}{n!\,\Gamma(k)} q^k (1-q)^n, \tag{2.3}$$

for $n \in \mathbb{N}$. The usual properties of the negative binomial still hold and are given in the following claim.

**Claim 1** (Basic properties of negative binomial distributions)**.** *The probability mass function given in* (2.3) *sums up to* 1 *and the random variable $X$ with* $\text{NEGBIN}(k, q)$ *distribution has mean $\mathbb{E}(X) = k\frac{1-q}{q}$ for all $k > 0$ and $q \in (0, 1)$. Moreover if $X_1$ and $X_2$ are two identical copies of $X$, then $X_1 + X_2 \sim \text{NEGBIN}(2k, q)$.*

*Proof.* The result follows from using the identity $Q_{X_1+X_2}(s) = Q_{X_1}(s)Q_{X_2}(s)$, where $Q_{X_1}$ and $Q_{X_2}$ are the probability generating functions of $X_1$ and $X_2$ respectively. □

The estimates for $k$ in [35] are often smaller than 1, which is why it is necessary to include these cases in the parameter space. The claim ensures that we have defined a proper distribution. Moreover, adding two negative binomials with the same success parameter still results in a negative binomial distribution. The intuition stems from the fact that a negative binomial random variable is the sum of $k$ i.i.d. geometric random random variables, but generalizes to the case where $k$ is not an integer.

Assuming a negative binomial offspring distribution comes with two useful properties in the computation for the probability mass function in (2.2). Since the branching process where each individual carries the ancestral label has offspring distribution $\widetilde{X} \sim \text{BIN}(X, 1-\nu)$, we need to apply Theorem 1 to the offspring distribution $\widetilde{X}$. It turns out that $\widetilde{X}$ is again a negative binomial, and by Claim 1 we also have that the sum of i.i.d. copies of $\widetilde{X}$ is a negative binomial.

**Claim 2.** *Let $X \sim \text{NEGBIN}(r, q)$ with $k > 0$ and $q \in (0, 1)$ and let $\widetilde{X} \sim \text{BIN}(X, 1-\nu)$ be the thinning of $X$, with $\nu \in (0, 1)$. Then $\widetilde{X} \sim \text{NEGBIN}(k, \tilde{q})$ with $\tilde{q} = \frac{q}{1-\nu(1-q)}$.*

The reproduction number $R$ is the average number of secondary case infections and the inference of this parameter is the goal of many studies [5, 7, 9, 10, 15, 25, 28, 35]. In our model this coincides with the mean of $X$ and since we have $\mathbb{E}(X) = k\frac{q}{1-q} = R$, we often refer to a negative binomial with mean $R$ and *dispersion parameter* $k$. In this case we have $q = \frac{k}{R+k}$ and from Claim 1 is follows that $\tilde{q} = \frac{k}{(1-\nu)R+k}$ for the thinned negative binomial, as $\mathbb{E}(\widetilde{X}) = (1-\nu)R$. Now Claim 1 and 2 can be combined to show the following corollary of Theorem 1.

**Corollary 1.** *Let $(Z_n)_{n\in\mathbb{N}}$ be a branching process with offspring distribution $X \sim \mathrm{NEGBIN}(k, q)$ with $k > 0$ and $q = \frac{k}{R+k}$ for some $R > 0$, and let $\nu$ be the probability of mutation. Then the probability mass function for the final size of a cluster of identical sequences $\widetilde{Y}_\infty$ is given by*

$$\mathbb{P}\Big(\widetilde{Y}_\infty = n\Big) = \frac{\Gamma(nk + (n-1))}{n!\,\Gamma(nk)} \left(\frac{k}{(1-\nu)R+k}\right)^{nk} \left(\frac{(1-\nu)R}{(1-\nu)R+k}\right)^{n-1}, \tag{2.4}$$

*which is proper if $(1-\nu)R \leq 1$. Otherwise, (2.4) is defective and $\mathbb{P}(\widetilde{Y}_\infty = \infty)$ is the survival probability.*

## 2.3 Inferring the reproduction number from identical genetic sequence cluster sizes

Following the ideas of [25], the probability mass function in (2.4) can be used construct a likelihood function. Given a set of observed cluster sizes, we can optimize the likelihood function to obtain a maximum likelihood estimation for $(R, k)$. In [35] they directly infer $R$ and $k$ from the result obtained in Corollary 1. However, they assume that each cluster is a branching process standing on its own, while in fact they are embedded in the bigger infection tree. This changes the observed cluster size distribution, and the formula in (2.4) needs to be modified, as we explain now.

Suppose that we observe the whole branching process up to some generation $n \in \mathbb{N}$. Moreover, we also know the label of each node generated by the infinite alleles model, and therefore we know the sizes of each cluster with identical labels. While observing clusters which have reached their final size (the total progeny), we also observe newly started clusters of smaller sizes that have yet to reach their final size (total size between its time of birth and $n$), see Figure 1.1. Thus as $n$ gets large, a proportion of the small clusters we observe were started only a few generations ago, but their potential final size is as least as large as the one observed. Hence we expect that the cumulative distribution function corresponding to (2.4), is stochastically dominated[1] by the true distribution function.

The goal of the following sections is to incorporate into the model that the clusters are observed in an exponentially growing tree, where some clusters might not have reached their final sizes yet. This requires us to consider a more general approach to model branching processes, which uses the Ulam-Harris family space of trees. With the help of this model we can define functions (characteristics), which count for example how many cluster of size $n$ we observe at some time $t > 0$ on a sampled tree. We then show that the stochastic processes that result from counting these characteristics on a branching process, converge under the appropriate scaling, to a computable limit. We use this limit to derive a likelihood function which better fits the model. Moreover, this limit can be used to estimate the probability of mutation $\nu$ from genetic sequence data.

[1] For two real valued random variables $X$ and $Y$, not necessarily living on the same probability space, $Y$ is stochastically dominated by $X$ if for all $x \in \mathbb{R}$, $\mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x)$.

# Chapter 3

# General single-type branching processes

## 3.1 The family space

Consider a branching process that starts at the root and gives birth to new individuals throughout its lifetime. Each individual is given a name instead of a label, since the latter term is already reserved for the infinite alleles model. The naming of the individuals is done according to the Ulam-Harris family history space, and our notation is based on that of [21]. The root is named 0 and is the only element of $\mathbb{N}^0 = \{0\}$. All of the root's possible children receive the names $i \in \mathbb{N}$, where $\mathbb{N} = \{1, 2, \ldots\}$. The second generation, which are the grandchildren of the root, are assigned the names $x \in \mathbb{N}^2$, meaning that if $x = (i, j)$, then $j$ is the $j$-th child of $i \in \mathbb{N}$. Continuing with this logic, we can express any individual $x \in \mathbb{N}^{n+1}$ for some $n \geq 0$ as $x = (x_1, \ldots, x_{n+1}) = yj$, where $y \in \mathbb{N}^n$ and $j \in \mathbb{N}$. Here we used the concatenation $xy = (x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{N}^{n+m}$ for some $x \in \mathbb{N}^n$ and $y \in \mathbb{N}^m$. The previous notation is consistent for any $n, m \geq 0$ if we also introduce the notation $x = x0 = 0x$. A schematic overview of a realisation of the branching process with only four generations is given in Figure 3.1.

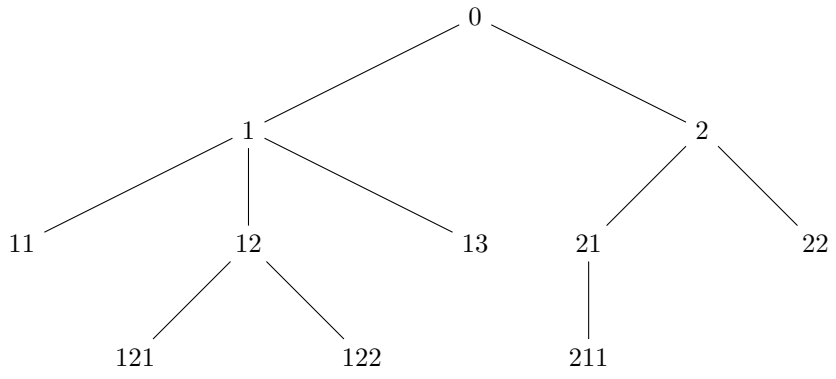


Figure 3.1: A schematic representation of the ancestral tree of a general branching process. The vertices represent the individuals born over time. If an edge is drawn between two vertices, then the top vertex gave birth to the bottom one over the course of it's life. The naming of the vertices follows the naming described in Section 3.1.

We build the sample space from which each realization of the branching process is drawn, first constructing a space for the life career of a single individual and expanding it to include all (possible conceivable) individuals.

**Definition 1.** Each possible individual is contained in the set of all *individuals* which is defined as

$$I = \bigcup_{n=0}^{\infty} \mathbb{N}^n.$$

All possible life careers are contained in the *life space* $\Omega$ equipped with a $\sigma$-algebra $\mathcal{A}$.

All the individuals $x \in I$ are assigned a life career $\omega_x$ which is an element from the life space $\Omega$. The information given by $\omega \in \Omega$ depends on the model assumptions. For example, if we consider a multi-type branching process in which individuals are associated with a certain type, then $\omega$ must describe which type is assigned to each child of the individual with life career $\omega$. The life careers should in any case contain enough information to define the functions $\tau_i(\omega)$, for each $i \in \mathbb{N}$, which gives the age of the parent with life $\omega$ at the moment of giving birth to the $i$-th child. So $\tau_1(\omega)$ is the first moment of childbearing. We define this formally now.

**Definition 2.** The *age of childbearing* of $i$-th child is defined as the real-valued function $\tau_i$ which is measurable with respect to $(\Omega, \mathcal{A})$. Moreover, it must satisfy

$$0 \leq \tau_1 \leq \tau_2 \leq \ldots \leq \infty, \tag{3.1}$$

where we set $\tau_i(\omega) = \infty$ if the life $\omega$ describes less than $i$ births. The *reproduction process* $\xi$ is defined as the point process

$$\xi(B) = |\{i : \tau_i \in B\}|,$$

for any borel set $B \in \mathcal{B}(\mathbb{R}^+)$. We abbreviate

$$\xi(t) = \xi([0, t]). \tag{3.2}$$

Given an individual $x \in I$ with life $\omega_x \in \Omega$, we denote its reproduction process by

$$\xi_x(B) = |\{i : \tau_i(\omega_x) \in B\}|.$$

In the preceding definition, we also defined the reproduction process, which is an important object for the application of the results in the following sections. Now that we have the ages of childbearing of the mothers of the individuals, we can add up the ages of childbearing of the ancestors to determine an individual's time of birth in the process, provided we have knowledge of the lives of these ancestors. For this reason, we now define the space from which the realizations of the branching process are sampled.

**Definition 3.** The *population process* is defined as $\left(\Omega^I, \mathcal{A}^I\right)$, where

$$\Omega^I = \bigtimes_{x \in I} \Omega, \quad \mathcal{A}^I = \bigotimes_{x \in I} \mathcal{A}, \tag{3.3}$$

which respectively denotes the Cartesian product of $(\Omega)_{x \in I}$ and the $\sigma$-algebra generated by the cylinder sets of $(\mathcal{A})_{x \in I}$.

The elements of $\Omega^I$ are of the form $\{\omega_x\}_{x \in I}$ and we denote $\underline{\omega} = \{\omega_x\}_{x \in I}$, which is a realisation of the branching process. Hence a life career $\omega_x$ can be identified for each individual $x \in I$. This gives us enough information to inductively define the times of birth for the individuals.

**Definition 4.** The *birth times* $\sigma_x$ of the individuals are defined by setting the birth time of the root

$$\sigma_0(\underline{\omega}) = 0,$$

and for any individual $x = yi$, where $y \in \mathbb{N}^n$ and $i \in \mathbb{N}$ for $n \geq 0$, as

$$\sigma_x(\underline{\omega}) = \sigma_y(\underline{\omega}) + \tau_i(\omega_y).$$

Note that in the definition of the birth times $\sigma_x$ and $\sigma_y$ are functions of $\underline{\omega}$ and $\tau_i$ is a function the element $\omega_y$. This ensures that $i$ is really a child of $y$. If we refer back to Figure 3.1, we see that individual 1 descended from 0 and gave birth to 11, which in turn did not produce any offspring. We thus have $\sigma_{11}(\{\underline{\omega}\}) = \sigma_1(\underline{\omega}) + \tau_1(\omega_1)$.

Whenever an individual $x = yi$ is not born, i.e. $\tau_i(\omega_y) = \infty$, it then follows from the definition that also the birth times $\sigma_{xz}(\underline{\omega}) = \infty$ for every $z \in I$. The branching process inherits its characteristic properties, e.g. the additive property given in Proposition 1, from the basic assumption that all individuals produce i.i.d. offspring. For this reason we make the following natural assumption to capture these properties for all population processes considered.

**Assumption 1.** *We assume the existence of a probability space* $(\Omega^I, \mathcal{A}^I, \mathbb{P})$ *such that the reproduction processes* $\xi_x$ *are i.i.d. for each* $x \in I$*.*

## 3.2 Counting general branching processes

Now that we have constructed the population process along with the birth times and a probability measure, we have defined the branching process. We want to derive measurable functions of the branching process. For example, the total size up to time $t$ can be counted via the function $Y_t$ which is measurable with respect to the probability space $(\Omega^I, \mathcal{A}^I, \mathbb{P})$. This can be done by computing the sum

$$Y_t = \sum_{x \in I} \mathbb{1}_{[\sigma_x, \infty)}(t), \tag{3.4}$$

which counts every individual whose time of birth occurred before $t$. If we want to count how many individuals gave birth to at least one individual by time $t$, we need to replace the indicator in (3.4) by $\mathbb{1}_{[\sigma_{x1}, \infty)}(t)$, since we want to count all individuals $x$ which gave birth to their firstborn $x1$. Giving the measurable function

$$N_t^{\text{mother}} = \sum_{x \in I} \mathbb{1}_{[\sigma_{x1}, \infty)}(t). \tag{3.5}$$

The indicators in the expressions (3.4) and (3.5) evaluate the contribution of each individual to outcome of the functions $Y_t$ and $N_t^{\text{mother}}$. We refer to the contribution of the root, in the examples these are the indicators $\mathbb{1}_{[0,\infty)}(t)$ and $\mathbb{1}_{[\sigma_1, \infty)}(t)$, as characteristics. We make this formal in the following definition.

**Definition 5.** A *random characteristic* is any real-valued process $(\chi(t))_{t \in \mathbb{R}}$ defined on $\Omega^I$ and for which the map $(t, \{\omega_x\}_{x \in I}) \mapsto \chi\big(t, \{\omega_x\}_{x \in I}\big)$ is measurable with respect to the product $\sigma$-algebra $\mathcal{B}(\mathbb{R}) \times \mathcal{A}^I$ and vanishes for negative values in the first argument.

*Remark* 1. Both characteristics $\mathbb{1}_{\mathbb{R}^+}$ and $\mathbb{1}_{[\sigma_1, \infty)}$ considered in (3.4) and (3.5) only look at the root itself or at it's reproduction process. For this reason we refer to characteristics that depend on $\omega_0$ as individual characteristics. In more generality, a characteristic could also depend on a larger (or the whole) subtree of the root. There is also an extension where one is allowed to look upwards in the tree for finitely many generations, see e.g. [21].

To be able to formally describe sums like (3.4) and (3.5), we introduce for any $x \in I$ the shift operator $S_x : \Omega^I \to \Omega^I$, which is defined as the map

$$S_x\Big(\{\omega_y\}_{y\in I}\Big) = \{\omega_{xy}\}_{y\in I}. \tag{3.6}$$

It maps any element from $\Omega^I$ to the subprocess where $x$ is the new root, thus making it a branching process started at $x$.

**Definition 6.** Let $\underline{\omega} \in \Omega^I$ and $x \in I$. The *daughter process* of $x$ is defined as the subprocess obtained by applying shift operator $S_x$ from (3.6) on $\underline{\omega}$.

*Remark* 2. In the preceding definition we have used the term daughter process of $x$ to describe the subtree with root $x$. This terminology is taken from [27] which assumes that everyone is female, and is equivalent to assuming that the reproduction process is asexual. The latter aligns with our framework regarding virus spread, which is why we choose to adapt this naming.

We can evaluate the characteristic on the daughter process of $x$, which we denote by

$$\chi_x(a) = \chi(a, S_x) = \chi \circ S_x(a) \tag{3.7}$$

and can be interpreted as the score of $x$ at age $a$. In order to count the scores at some time $t$, we need to make the time shift $t - \sigma_x$ to make sure that the time at which the characteristic $\chi_x$ is evaluated coincides with the age of $x$. If we set $\chi(t) = \mathbb{1}_{\mathbb{R}^+}(t)$, which corresponds to the case of (3.4), then we indeed have $\chi_x(t - \sigma_x) = \mathbb{1}_{\mathbb{R}^+}(t - \sigma_x) = \mathbb{1}_{[\sigma_x,\infty)}(t)$. With the tools at hand we now define the general branching process counted by any characteristic satisfying Definition 5.

**Definition 7.** Suppose that $\chi$ is a random characteristic, we define the *branching process counted by the random characteristic* $\chi$ as

$$Z_t^\chi = \sum_{x\in I} \chi_x(t - \sigma_x).$$

We conclude the section by stating an intuitive lemma that gives us the property that the daughter processes are indeed i.i.d. copies of the branching process started at the root. We denote by $\mathcal{A}_n$ the $\sigma$-algebra which contains the lives up to generation $n$, which is

$$\mathcal{A}_n = \sigma\left(\left\{\omega_x : x \in \bigcup_{k=0}^{n} \mathbb{N}^k\right\}\right). \tag{3.8}$$

We now give the following lemma known as the Generation branching lemma, as stated in [21].

**Lemma 2** (Generation branching lemma)**.** *Given $\mathcal{A}_n$, $n \in \mathbb{N}$, all daughter processes $S_x$, $x \in \mathbb{N}^{n+1}$, are conditionally independent and*

$$\mathbb{P}(S_x \in A \mid \mathcal{A}_n) = \mathbb{P}(A), \tag{3.9}$$

*for any $x \in \mathbb{N}^{n+1}$ and $A \in \mathcal{A}^I$.*

## 3.3 Some results from renewal theory

We want to study the asymptotics of $Z_t^\chi$. First, we study the convergence of the mean under appropriate scaling. This scaling is also helpful in determining whether a limiting distribution

exists. In this section, we derive some elementary results with the help of renewal theory. We start with the mean of the process

$$m_t^\chi = \mathbb{E}[Z_t^\chi]. \tag{3.10}$$

Results from renewal theory are relevant for us since they arise from the distribution of the arrival of points on the real line, which in our case is the arrival of newborn children, described by the reproduction process $\xi$. The expectation of $\xi$ is a useful object to study, which leads us to the following definition.

**Definition 8.** Suppose that $\xi$ is a reproduction process as in Definition 2. The *intensity measure* $\mu$ of $\xi$ is defined as $\mu(B) = \mathbb{E}[\xi(B)]$ for every $B \in \mathcal{B}(\mathbb{R}^+)$. We write

$$\mu(t) = \mathbb{E}[\xi(t)].$$

with $\xi$ as in (3.2)

We use the intensity measure $\mu$ to derive the renewal equation for the mean $m_t^\chi$ given in (3.10), which is stated in the following theorem.

**Theorem 2.** *Let $Z_t^\chi$ be a branching process counted by $\chi$ and $\mu$ be the intensity measure for the reproduction process. Then the mean $m_t^\chi = \mathbb{E}[Z_t^\chi]$ satisfies the renewal equation*

$$m_t^\chi = \mathbb{E}[\chi(t)] + \int_0^t m_{t-u}^\chi \, \mu(\mathrm{d}u). \tag{3.11}$$

We prove this theorem by deriving a (3.11) from a decomposition for $Z_t^\chi$. We make use of the basic decomposition, which we state in the following lemma.

**Lemma 3** (Basic decomposition of a branching process)**.** *The basic decomposition of a branching process if given by*

$$Z_t^\chi = \chi(t) + \sum_{i \in \mathbb{N}} Z_{t-\sigma_i}^\chi(i), \tag{3.12}$$

*where $Z_t^\chi(i) = Z_t^\chi \circ S_i$ denotes the branching process counted by $\chi$ with individual $i$ as the root.*

*Proof.* We split the sum $x \in I$ over the different generations $\mathbb{N}^n$ to arrive at

$$Z_t^\chi = \sum_{x \in I} \chi_x(t - \sigma_i) = \sum_{n=0}^{\infty} \sum_{x \in \mathbb{N}^n} \chi_x(t - \sigma_x) = \chi_0(t) + \sum_{n=1}^{\infty} \sum_{x \in \mathbb{N}^n} \chi_x(t - \sigma_x),$$

and point out that $\chi_0(t) = \chi(t)$. By writing every individual $x \in \mathbb{N}^n$ as $x = iy$ with $y \in \mathbb{N}^{n-1}$ and $i \in \mathbb{N}$, the summation can be rewritten as

$$\sum_{n=1}^{\infty} \sum_{y \in \mathbb{N}^{n-1}} \sum_{i \in \mathbb{N}} \chi_{iy}(t - \sigma_{iy}) = \sum_{m=0}^{\infty} \sum_{y \in \mathbb{N}^m} \sum_{i \in \mathbb{N}} \chi_{iy}(t - \sigma_{iy}) = \sum_{i \in \mathbb{N}} \sum_{y \in I} \chi_{iy}(t - \sigma_{iy}).$$

Since we have $S_{iy} = S_y \circ S_i$ and $\sigma_{iy} = \sigma_i + \sigma_y = \sigma_i + \sigma_y \circ S_i$, we arrive at

$$\sum_{i \in \mathbb{N}} \sum_{y \in I} \chi_y \circ S_i(t - \sigma_i - \sigma_y \circ S_i) = \sum_{i \in \mathbb{N}} Z_{t-\sigma_i}^\chi(i),$$

which gives the result. □

We are now ready to give the proof of Theorem 2.

*Proof of Theorem 2.* By taking expectations on both sides of (3.12) we obtain by monotonicity

$$m_t^\chi = \mathbb{E}[\chi(t)] + \sum_{i\in\mathbb{N}} \mathbb{E}\big[Z^\chi_{t-\sigma_i}(i)\big].$$

Since $S_i$ is independent of $\mathcal{A}_0$ by Lemma 2 and $\sigma_i$ is measurable with respect to $\mathcal{A}_0$, we see

$$\begin{aligned}\mathbb{E}\big[Z^\chi_{t-\sigma_i}(i)\big] &= \mathbb{E}\big[\mathbb{E}\big[Z^\chi_{t-\sigma_i}(i)\,\mathbb{1}_{\sigma_i\in[0,t]} \,\big|\, \mathcal{A}_0\big]\big] = \mathbb{E}\big[\mathbb{E}\big[Z^\chi_{t-\sigma_i}(i) \,\big|\, \mathcal{A}_0\big]\,\mathbb{1}_{\sigma_i\in[0,t]}\big]\\ &= \mathbb{E}\big[\mathbb{E}\big[Z^\chi_{t-\sigma_i}\big]\,\mathbb{1}_{\sigma_i\in[0,t]}\big] = \mathbb{E}\big[m^\chi_{t-\sigma_i}\,\mathbb{1}_{\sigma_i\in[0,t]}\big].\end{aligned}$$

If we again exchange the summation and expectation, we obtain

$$\mathbb{E}\left[\sum_{i\in\mathbb{N}} m^\chi_{t-\sigma_i}\,\mathbb{1}_{\sigma_i\in[0,t]}\right] = \mathbb{E}\left[\int_0^t m^\chi_{t-u}\,\xi(\mathrm{d}u)\right].$$

By properties of the random measure $\xi$ we have

$$\mathbb{E}\left[\int_0^t m^\chi_{t-u}\,\xi(\mathrm{d}u)\right] = \int_0^t m^\chi_{t-u}\;\mathbb{E}[\xi(\mathrm{d}u)] = \int_0^t m^\chi_{t-u}\,\mu(\mathrm{d}u),$$

since we defined $\mu(t)$ as the intensity function of $\xi$. □

Before stating the Renewal theorem, we give two definitions that are required for the statement of the theorem. The first one puts a restriction on the integrability of $\mathbb{E}[\chi(t)]$.

**Definition 9.** A non-negative function $h : D \subseteq \mathbb{R} \to [0,\infty)$ is *directly Riemann integrable* if the upper and lower Riemann sums of $h$ over the whole domain converge to the same limit as the mesh of the partition vanishes.

*Remark* 3. The above definition is more restrictive than the usual Riemann integrability. When $D = [0,\infty)$, $h$ is said to be Riemann integrable if the upper and lower Riemann sums converge on $[0,t]$ to a common value, and the integral is defined as the the limit of these values when $t \to \infty$. For the direct Riemann integrability we require that the upper and lower sums directly converge on $[0,\infty)$ as opposed to on $[0,t]$ for all $t > 0$.

The next definition concerns the intensity measure $\mu$ and makes a distinction between the cases where $\mu$ has all of its mass concentrated on $\lambda\mathbb{Z}$ for some $\lambda > 0$.

**Definition 10.** A probability measure $G$ on $[0,\infty)$ is called *lattice* with span $\lambda$, if for some $r \geq 0$, $\lambda$ is the largest number such that $\sum_{j=0}^\infty G(\{j\lambda + r\}) = 1$. We call $G$ *nonlattice* if no such number exists.

An example of such a lattice measure is the one for the Bienaymé-Galton-Watson process, where all mass is concentrated at 1. Now we state the Renewal theorem. This version of the theorem can be found in [13] together with a proof.

**Theorem 3** (Renewal theorem)**.** *Let $G$ be a probability measure on $[0,\infty)$ such that $\int u\,G(\mathrm{d}u) < \infty$ and $h$ is a directly Riemann integrable function on $[0,\infty)$. Suppose that $H$ is the solution of the following equation*

$$H(t) = h(t) + \int_0^t H(t-u)\,G(\mathrm{d}u). \tag{3.13}$$

1. *If $G$ is nonlattice, then*

$$\lim_{t\to\infty} H(t) = \frac{\int_0^\infty h(u)\,\mathrm{d}u}{\int_0^\infty u\,G(\mathrm{d}u)}. \tag{3.14}$$

*2. If $G$ is lattice with span $\lambda$, then*

$$\lim_{n\to\infty} H(x+n\lambda) = \lambda \frac{\sum_{u=0}^{\infty} h(x+u\lambda)}{\int_0^\infty u\, G(\mathrm{d}u)}. \tag{3.15}$$

We can apply Theorem 3 to the (transformed) equation (3.11) if we scale $\mu$ such that it gives a probability measure. Fisher describes in [14, Chapter 2] that given a density for the expected offspring, in our case this corresponds to $\mu(\mathrm{d}u)$, we can determine the stable age distribution. Note that $\mu$ does in general not need have a density with respect to the Lebesgue measure. In this case the expression $\int f(u)\,\mu(\mathrm{d}u))$ denotes the Lebesgue-Stieltjes integration of $f$ with respect to the measure $\mu$. When the unique rate $\alpha$ such that $\int_0^\infty e^{-\alpha u}\,\mu(\mathrm{d}u)$ integrates to 1 is computed, the cumulative distribution function for the stable age distribution is given by $\int_0^x e^{-\alpha u}\,\mu(\mathrm{d}u)$. The parameter $\alpha$ captures the exponential growth of the process and is generally known as the Malthusian parameter. For branching processes this is formally defined in the following way.

**Definition 11.** Suppose that $\mu$ is an intensity measure as in Definition 2. We define the *Malthusian parameter* (if it exists) as the $\alpha \in \mathbb{R}$ such that

$$\hat{\mu}(\alpha) = \int_0^\infty e^{-\alpha u}\,\mu(\mathrm{d}u) = 1. \tag{3.16}$$

*Remark* 4. As pointed out in [1], the Malthusian parameter always exists if $\mu([0,\infty)) \geq 1$ and satisfies $\alpha \geq 0$ by monotonicity. If $\mu([0,\infty)) < 1$, then we always have $\alpha < 0$ provided that it exists. When $\alpha > 0$, we speak of a supercritical branching process, and is a requirement for most of the following statements. The process is called subcritical is $\alpha < 0$ and critical otherwise.

By definition, $e^{-\alpha u}\,\mu(\mathrm{d}u)$ now gives a probability measure on $[0,\infty)$, which we use in the following corollary of Theorem 3.

**Corollary 2.** *Let $Z_t^\chi$ be a branching process counted by $\chi$ and $\mu$ be the intensity measure for the reproduction process with Malthusian parameter $\alpha$. If $\mathbb{E}[\chi(t)] \geq 0$ is continuous almost everywhere as a function of $t$ and satisfies*

$$\sum_{k=0}^{\infty} \sup_{k\leq a\leq k+1} e^{-\alpha a}\,\mathbb{E}[\chi(a)] < \infty, \tag{3.17}$$

*then as a result of Theorem 3, if $\mu$ is nonlattice we have*

$$\lim_{t\to\infty} e^{-\alpha t}\,\mathbb{E}[Z_t^\chi] = \frac{\int_0^\infty e^{-\alpha u}\,\mathbb{E}[\chi(u)]\,\mathrm{d}u}{\int_0^\infty u e^{-\alpha u}\,\mu(\mathrm{d}u)}. \tag{3.18}$$

*If $\mu$ is lattice with span $\lambda$ we have*

$$\lim_{n\to\infty} e^{-\alpha n}\,\mathbb{E}[Z_n^\chi] = \lambda \frac{\sum_{u=0}^{\infty} e^{-\alpha u\lambda}\,\mathbb{E}[\chi(u\lambda)]}{\int_0^\infty u e^{-\alpha u}\,\mu(\mathrm{d}u)}. \tag{3.19}$$

*Proof.* The result follows from multiplying both sides of equation (3.11) with $e^{-\alpha t}$. By rewriting the following

$$e^{-\alpha t}\int_0^t m_{t-u}^\chi\,\mu(\mathrm{d}u) = \int_0^t e^{-\alpha(t-u)} m_{t-u}^\chi e^{-\alpha u}\,\mu(\mathrm{d}u),$$

we see that Theorem 3 applies for $H(t) = e^{-\alpha t} m_t^\chi$, $h(t) = e^{-\alpha t}\,\mathbb{E}[\chi(t)]$ and $G(\mathrm{d}t) = e^{-\alpha t}\,\mu(\mathrm{d}t)$, provided that $e^{-\alpha t}\,\mathbb{E}[\chi(t)]$ is directly Riemann integrable. It follows from the assumption (3.17) that the upper Riemann sum is bounded from above. The assumption that $\mathbb{E}[\chi(t)]$ is continuous almost everywhere gives us then that the upper and lower sums converge to the same value. Now (3.18) follows from (3.14) immediately and (3.19) follows from evaluating (3.15) at $x = 0$. □

*Remark* 5. Corollary 2 is stated under the same conditions as Theorem 3.4 in [21], except they restrict $\mu$ to be nonlattice. We aim to extend the results obtained in Section 2.2 with the use of characteristics, which is why we include the lattice case. Jagers and Nerman also prove the convergence of $e^{-\alpha t} Z_t^\chi$ to a limiting random variable in probability and $L^1$ under similar conditions. Moreover, under a more restricting assumption, but still valid in our application, they provide almost sure convergence.

We want to use the results of this section in the following sections, where we construct characteristics for statistical analysis. Since we need stronger results than the convergence of the mean. It turns out that the scaled process $e^{-\alpha t} Z_t^\chi$ converges to a random variable $W_\infty$ scaled by a factor of the limit in (3.19). We skip the details and conclude this section with a statement on the almost sure convergence of the process. The proof can be found in [27] and is based on the construction of a martingale $\{W_t\}_{t\geq 0}$ which converges to $W_\infty$. The more restrictive assumption (3.20) is the cost for having the almost sure convergence.

**Theorem 4.** *Let $Z_t^\chi$ be a branching process counted by characteristic $\chi$, where $\chi$ is nonnegative and with paths in $D([0,\infty))$[1]. Suppose that $\mu$ is nonlattice with Malthusian parameter $\alpha \in (0,\infty)$ and*

$$\int_0^\infty e^{-ru}\,\mu(\mathrm{d}u) < \infty, \tag{3.20}$$

*for some $r < \alpha$. Moreover, assume that $\chi$ satisfies*

$$\mathbb{E}\left[\sup_{u\geq 0} e^{-ru}\chi(u)\right] < \infty, \tag{3.21}$$

*for some $r < \alpha$. Then, as $t \to \infty$,*

$$\frac{Z_t^\chi}{Y_t} \to \int_0^\infty e^{-\alpha u}\,\mathbb{E}[\chi(u)]\,\mathrm{d}u \quad \textit{almost surely,} \tag{3.22}$$

*conditional on $\{Y_t \to \infty\}$ where $Y_t$ is as* (3.4).

*Remark* 6. We stress that analogues for the lattice case also hold, as the results are intuitively interchangeable in the sense of equations (3.18) and (3.19), i.e. the Lebesgue differential $\mathrm{d}u$ can be replaced by $\lambda\, m(\mathrm{d}u)$, where $m(u) = 1$ for every integer $u \geq 0$. Conditions such as (3.20) remain unchanged if interpreted as a Lebesgue-Stieltjes integral and (3.20) can be relaxed by changing the supremum over positive real numbers to the positive integers. However, starting from the preceding theorem, the analogues are not stated any more for the sake of readability and to avoid repetition.

## 3.4 Counting cluster sizes on an infinite alleles single-type branching process

At the end of Section 2.2 we argued that the observed sizes of the observed clusters are influenced by the exponential growth of the process. With the results from Section 3.3, we can formalize this conjecture.

We do not yet define the classical Bienaymé-Galton-Watson process as a general branching process, because the characteristics we construct do not depend on the underlying reproduction

[1] $D([0,\infty))$ is the space of functions defined on $[0,\infty)$ with left and right limits everywhere.

process. We define the infinite alleles model again as a process independent of the reproduction process $\xi$. We introduce the indicator function $\gamma : \mathbb{N} \times \Omega \to \{0, 1\}$ as

$$\gamma(i, \omega_y) = \mathbb{1}(\text{the } i\text{-th child of an individual with life career } \omega_y \text{ inherits the label from the individual}),$$

With probability $\nu$ a new label is introduced for every born individual, making $\gamma(i, \omega_y)$ a independent Bernoulli random variable with success probability $1 - \nu$ for every $i \in \mathbb{N}$ and $\omega_y \in \Omega$. We assume that the outcome of $\gamma$ is i.i.d. for every pair $(i, \omega_y) \in \mathbb{N} \times \Omega$ and independent of the reproduction process.

We wish to derive an empirical measure of cluster sizes with the use of characteristics. When gathering clusters of identical labels, e.g. genetic sequences, such a measure is exactly what can be estimated from the data. The goal is to construct characteristics $\chi^n$ and $\chi^{\mathbb{N}}$ such that

$$\frac{Z_t^{\chi^n}}{Z_t^{\chi^{\mathbb{N}}}} = \frac{\#\text{clusters of size } n \text{ up to time } t}{\#\text{clusters up to time } t}, \tag{3.23}$$

and let $t \to \infty$. We are again computing a probability mass function for the observed cluster sizes. But we expect to see a different result than (2.4), because we take into account that an observed cluster can potentially attain a larger size than its size at the time of the observation.

Suppose that an individual $y \in I$ produces an offspring $yi$ which carries a new label, i.e. $\gamma(i, \omega_y) = 0$. Then the number of individuals carrying the same label as $yi$ in the subtree started by $yi$, is equal to the size of this cluster. Hence we define a new reproduction process $\tilde{\xi}$ embedded in $\xi$, which only keeps individuals carrying the parent's label. This new process can be expressed as follows

$$\tilde{\xi}(B, \omega) = \sum_{i \in \mathbb{N}} \gamma(i, \omega) \mathbb{1}_B(\tau_i(\omega)), \tag{3.24}$$

for $B \in \mathcal{B}(\mathbb{R})$. Intuitively, this embedded process is the binomial thinning of original reproduction process. All individuals in the branching process with the reproduction process $\tilde{\xi}$ carry the same label as the root, which is referred to as the ancestral label, analogous to Section 2.2. We denote by $\widetilde{Y}_t$ the total size of the branching process carrying the ancestral label up to time $t$, in the same way as (3.4). The characteristic $\chi^n$ which counts how many cluster of size $n$ result from the root, can be defined as

$$\chi^n(t) = \sum_{i \in \mathbb{N}} \mathbb{1}_{\mathbb{R}^+}(t - \sigma_i)(1 - \gamma(i, \omega_0)) \mathbb{1}_{\{n\}}\left(\widetilde{Y}_{t-\sigma_i}(i)\right). \tag{3.25}$$

Here $\mathbb{1}_{\mathbb{R}^+}(t-\sigma_i)$ is the indicator that the individual $i$ is born by time $t$, $(1-\gamma(i, \omega_0))$ is the indicator that a new label is introduced with $i$ and $\mathbb{1}_{\{n\}}(\widetilde{Y}_{t-\sigma_i}(i))$ is the indicator that the subtree started by $i$ with the label of $i$ is of size $n$. Thus with $Z_t^{\chi^n}$ we count for every individual the number of children that satisfy the following condition: a child is born, a mutation happened between the child and the individual, and $\widetilde{Y}_{t-\sigma_i}(i)$ is equal to $n$. If we want to count the total number of clusters, then we leave out the last indicator, which gives

$$\chi^{\mathbb{N}}(t) = \sum_{i \in \mathbb{N}} \mathbb{1}_{\mathbb{R}^+}(t - \sigma_i)(1 - \gamma(i, \omega_0)). \tag{3.26}$$

In order to compute the limit of (3.23) as $t \to \infty$, we state the following corollary of Theorem 4.

**Corollary 3.** *Let $Z_t^{\chi}$ and $Z_t^{\chi'}$ be branching process counted by $\chi$ and $\chi'$ respectively, such that both $\chi$ and $\chi'$ and $\mu$ satisfy the conditions of Theorem 4. Then, as $t \to \infty$,*

$$\frac{Z_t^{\chi}}{Z_t^{\chi'}} \to \frac{\int_0^\infty e^{-\alpha u} \, \mathbb{E}[\chi(u)] \, \mathrm{d}u}{\int_0^\infty e^{-\alpha u} \, \mathbb{E}[\chi'(u)] \, \mathrm{d}u} \quad \textit{almost surely,}$$

*conditional on* $\{Y_t \to \infty\}$.

In the next two section we apply the corollary under two different assumptions. The first case coincides with the Bienaymé-Galton-Watson process, which means that the age of childbirth is one for every birth that takes place in the life $\omega$. The second case is the age-dependent branching process, which means that the age of childbearing follows a general distribution $G$ on $[0, \infty)$ which is nonlattice, and all births in the life $\omega$ take place at this time. In both cases we wish to compute the probability mass function for the cluster sizes observed in an exponentially growing tree. It turns out that Corollary 3 can immediately give us a probability mass function with the characteristics we constructed. We state this in the following definition.

**Definition 12.** The limiting empirical cluster size distribution $C^\alpha$ is defined as the random variable with the probability mass function

$$\mathbb{P}(C^\alpha = n) = \frac{\int_0^\infty e^{-\alpha u}\, \mathbb{E}[\chi^n(u)]\, \mathrm{d}u}{\int_0^\infty e^{-\alpha u}\, \mathbb{E}[\chi^{\mathbb{N}}(u)]\, \mathrm{d}u}, \tag{3.27}$$

where $\chi^n$ and $\chi^{\mathbb{N}}$ are as in (3.25) and (3.26) respectively.

### 3.4.1 Empirical cluster size distribution for the Bienaymé-Galton-Watson process

As in Section 2.2 every individual produces offspring upon death according to a distribution $X$. The lifetime of each individual is of length 1. The reproduction process for an individual with life $\omega$ is given by

$$\xi(t, \omega) = X_\omega \mathbb{1}_{[1,\infty)}(t), \tag{3.28}$$

where $X_\omega$ is a copy of $X$. The intensity measure $\mu(t) = \mathbb{E}[\chi(t)]$ takes the form

$$\mu(t) = m\delta_1(\{t\}),$$

where $\mathbb{E}[X] = m$ and $\delta_1$ is the Dirac measure with $\delta_1(A)$=1 if $1 \in A$ and 0 otherwise. Observe that all the mass of $\mu$ is concentrated at 1, which makes $\mu$ lattice with $\lambda = 1$ according to Definition 10. According to Definition 11, the Malthusian parameter is the solution to

$$\int_0^\infty e^{-\alpha u}\, \mu(\mathrm{d}u) = me^{-\alpha} = 1,$$

which is $\alpha = \log(m)$. Observe that $\alpha > 0$ only if $m > 1$, and corresponds to the supercritical regime. We now wish to compute the probability mass function $\mathbb{P}(C^\alpha = n)$ as in (3.27). Because of the lattice property, the limit of the ratio (3.23) takes the form

$$\mathbb{P}(C^\alpha = n) = \frac{\sum_{u=0}^\infty e^{-\alpha u}\, \mathbb{E}[\chi^n(u)]}{\sum_{u=0}^\infty e^{-\alpha u}\, \mathbb{E}[\chi^{\mathbb{N}}(u)]}, \tag{3.29}$$

where we point out that $e^{-\alpha u} = m^{-u}$. We need to compute the expectation of $\chi^n(u)$, by the independence of the mutation process we have

$$\begin{aligned}
\mathbb{E}[\chi^n(u)] &= \sum_{i \in \mathbb{N}} \mathbb{E}[(1 - \gamma(i, \omega_0)]\, \mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u - \sigma_i)\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-\sigma_i}(i)\Big)\Big] \\
&= \nu \sum_{i \in \mathbb{N}} \mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u - \sigma_i)\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-\sigma_i}(i)\Big)\Big],
\end{aligned}$$

where $\nu$ is the probability that a child does not inherit the label of the parent. By conditioning on $X_{\omega_0}$ using the law of total expectation, we know that $\mathbb{1}_{\mathbb{R}^+}(u-\sigma_i) = \mathbb{1}_{\mathbb{R}^+}(u-1)$ if $X \geq i$ and 0 otherwise. This gives

$$\begin{aligned}\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-\sigma_i)\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-\sigma_i}(i)\Big)\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-\sigma_i)\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-\sigma_i}(i)\Big)\;\Big|\; X_{\omega_0}\Big]\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{1}(X_{\omega_0}\geq i)\mathbb{1}_{\mathbb{R}^+}(u-1)\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-1}(i)\Big)\;\Big|\; X_{\omega_0}\Big]\Big],\end{aligned}$$

where we can take the indicator measurable with respect to $X_{\omega_0}$ out. By Lemma 2 we know that the branching process started at $i$ is independent of the reproduction process of the root, and therefore of $X_{\omega_0}$. Rewriting this leads us to

$$\begin{aligned}&\mathbb{1}(u\geq 1)\,\mathbb{E}\Big[\mathbb{1}(X_{\omega_0}\geq i)\,\mathbb{E}\Big[\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-1}(i)\Big)\Big]\Big] \\ &\qquad = \mathbb{1}(u\geq 1)\,\mathbb{E}\Big[\mathbb{1}(X_{\omega_0}\geq i)\,\mathbb{P}\Big(\widetilde{Y}_{u-1}(i)=n\Big)\Big].\end{aligned}$$

We again apply Lemma 2 to argue that $\widetilde{Y}_{u-1}(i)$ and $\widetilde{Y}_{u-1}$ are equal in distribution, and therefore have

$$\mathbb{1}(u\geq 1)\,\mathbb{P}\Big(\widetilde{Y}_{u-1}=n\Big)\,\mathbb{E}[\mathbb{1}(X_{\omega_0}\geq i)] = \mathbb{1}(u\geq 1)\,\mathbb{P}\Big(\widetilde{Y}_{u-1}=n\Big)\,\mathbb{P}(X_{\omega_0}\geq i).$$

At last we take out all terms that do not depend on $i$ out the sum and recognize the expectation of the offspring distribution $X$, to obtain

$$\mathbb{E}[\chi^n(u)] = \mathbb{1}(u\geq 1)\,\mathbb{P}\Big(\widetilde{Y}_{u-1}=n\Big)\sum_{i\in\mathbb{N}}\mathbb{P}(X_{\omega_0}\geq i) = \mathbb{1}(u\geq 1)\,\mathbb{P}\Big(\widetilde{Y}_{u-1}=n\Big)\,\mathbb{E}[X].$$

The computation for $\mathbb{E}[\chi^{\mathbb{N}}(u)]$ is along the same lines but with the indicator $\mathbb{1}_{\{n\}}(\widetilde{Y}_{u-1}(i))$ substituted by 1, and arrives at the expression

$$\mathbb{E}\big[\chi^{\mathbb{N}}(u)\big] = \mathbb{1}(u\geq 1)\,\mathbb{E}[X].$$

We conclude that the probability mass function in equation (3.29) is equal to

$$\mathbb{P}(C^\alpha = n) = \frac{\sum_{u=1}^{\infty} m^{-u}\,\mathbb{P}\Big(\widetilde{Y}_{u-1}=n\Big)}{\sum_{u=1}^{\infty} m^{-u}} = \left(1-\frac{1}{m}\right)\sum_{u=0}^{\infty} m^{-u}\,\mathbb{P}\Big(\widetilde{Y}_u=n\Big). \tag{3.30}$$

The expression in terms of $\widetilde{Y}_u$ in (3.30) allows us to compare the probability mass function with the one derived in Corollary (1). The corollary is a specific case where the offspring distribution is a negative binomial random variable, but it easily follows from Theorem 1 that a more general formula can be stated.

**Theorem 5.** *Consider an infinite alleles Bienaymé-Galton-Watson branching process with offspring distribution $X$ and mutation parameter $\nu \in (0,1)$. Then the total size $\widetilde{Y}_\infty$ of the final cluster sizes with (possible defective) probability mass function*

$$\mathbb{P}\Big(\widetilde{Y}_\infty = n\Big) = \frac{1}{n}\,\mathbb{P}(\mathrm{BIN}(nX, 1-\nu) = n-1),$$

*is stochastically dominated by $C^\alpha$ with probability mass function* (3.30).

*Proof.* As mention at the end of Section 2.3, $\widetilde{Y}_\infty$ is stochastically dominated by $C^\alpha$ if for all $n \in \mathbb{N}$, $\mathbb{P}(C^\alpha \leq n) \geq \mathbb{P}(\widetilde{Y}_\infty \leq n)$. Observe that the event

$$\Big\{\widetilde{Y}_\infty \leq n\Big\} = \Big\{\widetilde{Y}_u \leq n\Big\} \cap \Big\{\widetilde{Y}_t \leq n, \forall t > u\Big\},$$

is contained in $\{\widetilde{Y}_u \leq n\}$ for any $u \in \mathbb{N}$. It then follows that

$$\mathbb{P}\Big(\widetilde{Y}_\infty \leq n\Big) \leq \mathbb{P}\Big(\widetilde{Y}_u \leq n\Big),$$

which gives the result

$$\begin{aligned}\mathbb{P}(C^\alpha \leq n) = \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{P}\Big(\widetilde{Y}_u \leq n\Big) &\geq \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{P}\Big(\widetilde{Y}_\infty \leq n\Big) \\ &\geq \mathbb{P}\Big(\widetilde{Y}_\infty \leq n\Big).\end{aligned}$$

□

The next quantity of interest is the mean of $C^\alpha$. We state and prove the expression in the following lemma which we use later to prove a more general statement regarding an estimator for $\nu$.

**Lemma 4.** *For a Bienaymé-Galton-Watson process with Malthusian parameter $\alpha > 0$ equipped with the infinite alleles model with mutation probability $\nu \in (0, 1)$, the mean of empirically observed cluster size is*

$$\mathbb{E}[C^\alpha] = \nu^{-1} \tag{3.31}$$

*Proof.* Expanding and interchanging summation gives

$$\begin{aligned}\mathbb{E}[C^\alpha] = \sum_{n=1}^{\infty} n\, \mathbb{P}(C^\alpha = n) &= \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u} \sum_{n=1}^{\infty} n\, \mathbb{P}\Big(\widetilde{Y}_u = n\Big) \\ &= \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{E}\Big[\widetilde{Y}_u\Big].\end{aligned}$$

We can use the basic decomposition (3.12), for $\widetilde{Y}_u$ where $\chi(t) = \mathbb{1}_{\mathbb{R}^+}(t)$ and $\tilde{\xi}(1, \omega_0) = \widetilde{X}_{\omega_0}$ denotes the number of individual produced by the root carrying the ancestral label. The decomposition is as follows

$$\widetilde{Y}_u = 1 + \sum_{i=1}^{\widetilde{X}_{\omega_0}} \widetilde{Y}_{u-1}(i),$$

where we use the convention that $\widetilde{Y}_{-1} = 0$ such that we correctly have $\widetilde{Y}_0 = 1$. Substituting this decomposition into the preceding expression gives

$$\begin{aligned}\left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{E}\Big[\widetilde{Y}_u\Big] &= \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{E}\left[1 + \sum_{i=1}^{\widetilde{X}_{\omega_0}} \widetilde{Y}_{u-1}(i)\right] \\ &= \left(1 - \frac{1}{m}\right) \left(\sum_{u=0}^{\infty} m^{-u} + \sum_{u=0}^{\infty} m^{-u}\, \mathbb{E}\left[\sum_{i=1}^{\widetilde{X}_{\omega_0}} \widetilde{Y}_{u-1}(i)\right]\right) \\ &= 1 + \left(1 - \frac{1}{m}\right) \sum_{u=0}^{\infty} m^{-u}\, \mathbb{E}\Big[\widetilde{X}_{\omega_0}\Big]\, \mathbb{E}\Big[\widetilde{Y}_{u-1}\Big],\end{aligned} \tag{3.32}$$

where we used Wald's identity and the fact that $\widetilde{Y}_{u-1}(i)$ is a copy of $\widetilde{Y}_{u-1}$ in the last step. Since $\widetilde{X}_{\omega_0}$ is the binomial thinning of $X_{\omega_0}$ it has mean $(1-\nu)m$ and $\mathbb{E}\Big[\widetilde{Y}_{-1}(i)\Big] = 0$, we obtain

$$\begin{aligned}\sum_{u=0}^{\infty} m^{-u}\,\mathbb{E}\Big[\widetilde{X}_{\omega_0}\Big]\,\mathbb{E}\Big[\widetilde{Y}_{u-1}\Big] &= (1-\nu)m\sum_{u=0}^{\infty} m^{-u}\,\mathbb{E}\Big[\widetilde{Y}_{u-1}\Big] \\ &= (1-\nu)\sum_{u=0}^{\infty} m^{-u}\,\mathbb{E}\Big[\widetilde{Y}_{u}\Big].\end{aligned}$$

where we moved the index of $\widetilde{Y}_u$. Substituting this into (3.32) gives

$$\mathbb{E}[C^\alpha] = 1 + (1-\nu)\Big(1-\frac{1}{m}\Big)\sum_{u=0}^{\infty} m^{-u}\,\mathbb{E}\Big[\widetilde{Y}_u\Big] = 1 + (1-\nu)\,\mathbb{E}[C^\alpha],$$

which simplifies to (3.31). □

### 3.4.2 Empirical cluster size distribution for the age-dependent process

The age-dependent branching process is a stochastic process in continuous time. Similarly to the Bienaymé-Galton-Watson process, individuals produce a random number of offspring with distribution $X$ upon death. However, lifetimes are not of length 1, but follow a general age distribution $G$ which takes values in $[0,\infty)$ and is of nonlattice type. The reproduction process has a similar form as the Bienaymé-Galton-Watson case (3.28), but is due to the random waiting times defined as

$$\xi(t,\omega) = X_\omega\,\mathbb{1}_{[T_\omega,\infty)}(t),$$

where $X_\omega$ is a copy of $X$ and $T_\omega \sim G$ which are independent. The intensity measure is given by

$$\mu(t) = \mathbb{E}\big[X\,\mathbb{1}_{[T,\infty)}(t)\big] = \mathbb{E}[X]\,\mathbb{P}(T>t) = m(1-G(t)). \tag{3.33}$$

We find the Malthusian parameter again by computing $\hat{\mu}(\alpha)$, which equals

$$\int_0^\infty e^{-\alpha u}\,\mu(\mathrm{d}u) = \int_0^\infty e^{-\alpha u} m\,G(\mathrm{d}u), \tag{3.34}$$

where $\alpha$ solves $\hat{\mu}(\alpha) = 1$ as in (3.16).

Since the support of $\mu$ is not a lattice in the age-dependent setting, we have to compute $\mathbb{P}(C^\alpha = n)$ in the form of (3.27). Starting with the numerator, we obtain in a similar way as the lattice case that

$$\begin{aligned}\mathbb{E}[\chi^n(u)] &= \nu\sum_{i\in\mathbb{N}}\mathbb{E}\Big[\mathbb{1}(X_{\omega_0}\geq i)\,\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-T_{\omega_0})\,\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-T_{\omega_0}}\Big)\;\Big|\;X_{\omega_0}\Big]\Big] \\ &= \nu\sum_{i\in\mathbb{N}}\mathbb{P}(X\geq i)\,\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-T_{\omega_0})\,\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-T_{\omega_0}}\Big)\Big] \\ &= \nu\,\mathbb{E}[X]\,\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-T_{\omega_0})\,\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-T_{\omega_0}}\Big)\Big].\end{aligned}$$

We can again apply the law of total expectation conditioning on $T_{\omega_0}$

$$\begin{aligned}&\mathbb{E}\Big[\mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-T_{\omega_0})\,\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-T_{\omega_0}}\Big)\;\Big|\;T_{\omega_0}\Big]\Big] \\ &\quad= \int_0^\infty \mathbb{E}\Big[\mathbb{1}_{\mathbb{R}^+}(u-T_{\omega_0})\,\mathbb{1}_{\{n\}}\Big(\widetilde{Y}_{u-T_{\omega_0}}\Big)\;\Big|\;T_{\omega_0}=t\Big]\,G(\mathrm{d}t) \\ &\quad= \int_0^u \mathbb{P}\Big(\widetilde{Y}_{u-t}=n\Big)\,G(\mathrm{d}t).\end{aligned}$$

At last we compute

$$\begin{aligned}\int_0^\infty e^{-\alpha u}\,\mathbb{E}[\chi^n(u)]\,\mathrm{d}u &= \nu\,\mathbb{E}[X]\int_{u=0}^\infty e^{-\alpha u}\int_0^u \mathbb{P}\Big(\widetilde{Y}_{u-t}=n\Big)\,G(\mathrm{d}t)\,\mathrm{d}u\\ &= \nu m\int_{t=0}^\infty\int_{u=t}^\infty e^{-\alpha u}\,\mathbb{P}\Big(\widetilde{Y}_{u-t}=n\Big)\,\mathrm{d}u\,G(\mathrm{d}t),\end{aligned}$$

and with the substitution $v = u - t$ we arrive at

$$\begin{aligned}&\nu m\int_{t=0}^\infty\int_{v=0}^\infty e^{-\alpha(v+t)}\,\mathbb{P}\Big(\widetilde{Y}_v=n\Big)\,\mathrm{d}v\,G(\mathrm{d}t)\\ &\qquad= \nu\int_{t=0}^\infty e^{-\alpha t}m\,G(\mathrm{d}t)\int_{v=0}^\infty e^{-\alpha v}\,\mathbb{P}\Big(\widetilde{Y}_v=n\Big)\,\mathrm{d}v\\ &\qquad= \nu\int_0^\infty e^{-\alpha v}\,\mathbb{P}\Big(\widetilde{Y}_v=n\Big)\,\mathrm{d}v.\end{aligned}$$

Analogous to the Bienaymé-Galton-Watson case, we obtain the result for $\chi^{\mathbb{N}}$ by replacing $\mathbb{1}_{\{n\}}(\widetilde{Y}_{u-\sigma_i}(i))$ with 1, giving

$$\int_0^\infty e^{-\alpha u}\,\mathbb{E}\big[\chi^{\mathbb{N}}(u)\big]\,\mathrm{d}u = \nu\int_{t=0}^\infty e^{-\alpha t}m\,G(\mathrm{d}t)\int_{v=0}^\infty e^{-\alpha v}\,\mathrm{d}v = \frac{\nu}{\alpha}.$$

The probability mass function for the observed cluster sizes in the age-dependent setting is therefore given by

$$\mathbb{P}(C^\alpha = n) = \alpha\int_0^\infty e^{-\alpha u}\,\mathbb{P}\Big(\widetilde{Y}_u=n\Big)\,\mathrm{d}u. \tag{3.35}$$

There is an analogue for Lemma 4 in age-dependent case. Exactly the same phenomena occurs, the average observed cluster size is given by $\nu^{-1}$. The proof uses many of the same details from the derivation of (3.35) and the proof Lemma 4.

**Lemma 5.** *For an age-dependent branching process with Malthusian parameter $\alpha > 0$ equipped with the infinite alleles model with mutation probability $\nu \in (0,1)$, the mean of empirically observed cluster size is*

$$\mathbb{E}[C^\alpha] = \nu^{-1}$$

*Proof.* First we exchange the integral and summation, and make use of the basic decomposition to obtain

$$\mathbb{E}[C^\alpha] = \alpha\int_0^\infty e^{-\alpha u}\,\mathbb{E}\Big[\widetilde{Y}_u\Big]\,\mathrm{d}u = \alpha\int_{u=0}^\infty e^{-\alpha u}\left(1+\mathbb{E}\Big[\widetilde{X}\Big]\int_{t=0}^u \mathbb{E}\left[\widetilde{Y}_{u-t}\right]G(\mathrm{d}t)\right)\mathrm{d}u. \tag{3.36}$$

We evaluate and again exchange integrals,

$$\begin{aligned}&1+\mathbb{E}\Big[\widetilde{X}\Big]\alpha\int_{u=0}^\infty e^{-\alpha u}\int_{t=0}^u \mathbb{E}\left[\widetilde{Y}_{u-t}\right]G(\mathrm{d}t)\,\mathrm{d}u\\ &\qquad= 1+(1-\nu)m\alpha\int_{t=0}^\infty\int_{u=t}^\infty e^{-\alpha u}\,\mathbb{E}\left[\widetilde{Y}_{u-t}\right]\mathrm{d}u\,G(\mathrm{d}t),\end{aligned} \tag{3.37}$$

perform the substitution $v = u - t$, and recognize that $\hat{\mu}(\alpha) = 1$ and the expression in (3.36), which brings us to

$$1+(1-\nu)\int_{t=0}^\infty e^{-\alpha t}m\,G(\mathrm{d}t)\alpha\int_{v=0}^\infty e^{-\alpha v}\,\mathbb{E}\left[\widetilde{Y}_v\right]\mathrm{d}v = 1+(1-\nu)\,\mathbb{E}[C^\alpha], \tag{3.38}$$

and concludes the proof. □

### 3.4.3 Parameter estimation for single-type branching processes and the infinite alleles model based on empirical cluster sizes

In this section we cover the estimation of model parameters based on the results derived earlier in Sections 3.4.1 and 3.4.2. We ended both sections with the computation of the mean for the observed cluster sizes. It appeared that in both cases the mean equals $\nu^{-1}$, which implies that we can use the empirically observed cluster sizes to obtain an estimator for $\nu$. Due to the simplicity of the expression, we can obtain a strong result with the help of some basic limit theorems. In the following theorem we state and prove that empirical mean gives a strongly consistent estimator for the mutation probability $\nu$.

**Theorem 6.** *Consider a branching process equipped with the infinite alleles model where each individual gives birth to a random number of children upon death according to a random variable $X$. Moreover, assume that death occurs at age 1 (Galton-Watson) or after a random time with nonlattice distribution $G$ on $[0,\infty)$ (age-dependent) and the Malthusian parameter satisfies $\alpha \in (0,\infty)$. If $\nu \in (0,1)$ is the probability that a new allele is introduced, then*

$$\hat{\nu} = 1/\overline{C_n^\alpha},$$

*is a (strongly) consistent estimator for $\nu$, where $(C_i^\alpha)_{i=1}^n$ are the observed cluster sizes with probability mass function* (3.27)*.*

*Proof.* We showed in Lemmas 4 and 5 that $\mathbb{E}[C_i^\alpha] = \nu^{-1}$. From the Strong Law of Large Numbers it follows that $\overline{C_n^\alpha} \to \nu^{-1}$ almost surely as $n \to \infty$, as the observed cluster sizes are assumed to be i.i.d. daughter process of the branching process. From the Continuous mapping theorem it follows that as $n \to \infty$, $1/\overline{C_n^\alpha} \to \nu$ almost surely, making $\overline{C_n^\alpha}$ a strongly consistent estimator. □

Observe how the only requirement for the process on the reproduction process is that $\alpha \in (0,\infty)$ which is for the Bienaymé-Galton-Watson process and age-dependent process equivalent to being supercritical. We refer to a standard probability textbook for the limit results used in the proof.

We use the rest of the section to derive expressions for the probability mass functions obtained in equations (3.30) and (3.35). Due to the dependence on the total size $\widetilde{Y}_u$, it is hard to compute probability mass function for large $n$. We restrict ourselves to analytically computing the probabilities of observing clusters of size 1 and 2, resulting in a distribution $\mathbb{P}(C^\alpha = 1)$, $\mathbb{P}(C^\alpha = 2)$ and $\mathbb{P}(C^\alpha \geq 3)$. Furthermore, we assume the offspring distribution $X$ to satisfy $\mathbb{E}[X] = m > 1$. In Chapter 5, such that we are in the same setting as Section 2.2. Now $\widetilde{X} \sim \mathrm{BIN}(X, 1-\nu)$ gives the offspring distribution for the branching process carrying the ancestral label, which can be found in Claim 2.

#### Bienaymé-Galton-Watson case

In the discrete time, we have that $\{\widetilde{Y}_u = 1\}$ only occurs for $u \geq 1$ when the root does not produce offspring with the same label, i.e. $\{\widetilde{X}_{\omega_0} = 0\}$. Thus $\mathbb{P}(\widetilde{Y}_0 = 1) = 1$ and $\mathbb{P}(\widetilde{Y}_u = 1) = \mathbb{P}(\widetilde{X}_{\omega_0} = 0)$ for $u \geq 1$. This gives

$$\mathbb{P}(C^\alpha = 1) = \left(1 - \frac{1}{m}\right)\left(1 + \sum_{u=1}^{\infty} m^{-u}\,\mathbb{P}\Big(\widetilde{X}_{\omega_0} = 0\Big)\right) = 1 - \frac{1}{m}\Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big).$$

For $\{\widetilde{Y}_u = 2\}$ we need that the root gives birth to exactly one individual with the same label, and that individual needs to produce no offspring with the ancestral label. That is $\{\widetilde{X}_{\omega_0} = 1, \widetilde{X}_{\omega_1} =$

0}. We have

$$\begin{aligned}\mathbb{P}(C^\alpha = 2) &= \left(1 - \frac{1}{m}\right)\left(\frac{1}{m}\,\mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1\Big) + \sum_{u=2}^{\infty} m^{-u}\,\mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1\Big)\,\mathbb{P}\Big(\widetilde{X}_{\omega_1} = 0\Big)\right) \\ &= \mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1\Big)\left(1 - \frac{1}{m}\right)\left(\frac{1}{m} + \mathbb{P}\Big(\widetilde{X}_{\omega_1} = 0\Big)\frac{1}{m^2}\sum_{u=0}^{\infty} m^{-u}\right),\end{aligned}$$

which can be rewritten in the following form

$$\begin{aligned}\mathbb{P}(C^\alpha = 2) = \mathbb{P}\Big(\widetilde{X} = 1\Big)\left(\frac{1}{m} - \frac{1}{m^2} + \mathbb{P}\Big(\widetilde{X} = 0\Big)\frac{1}{m^2}\right) &= \frac{1}{m}\,\mathbb{P}\Big(\widetilde{X} = 1\Big)\left(1 - \frac{1}{m}\Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\right) \\ &= \frac{1}{m}\,\mathbb{P}\Big(\widetilde{X} = 1\Big)\,\mathbb{P}(C^\alpha = 1).\end{aligned}$$

**Age-dependent case**

We consider an age-dependent branching process with general age distribution $G$ and offspring distribution $X$, where $\mathbb{E}[X] = m > 1$, such that the branching process has Malthusian parameter $\alpha \in (0, \infty)$, according to [1].

We can have $\{Y_u = 1\}$ if the root did not reach the end of its life yet at time $u$, or it produces no offspring carrying the same label at the end of its life. Thus

$$\begin{aligned}\mathbb{P}\Big(\widetilde{Y}_u = 1\Big) &= \mathbb{P}(T_{\omega_0} > u) + \mathbb{P}\Big(\widetilde{X} = 0, T_{\omega_0} < u\Big) \\ &= \int_{t=u}^{\infty} G(\mathrm{d}t) + \mathbb{P}\Big(\widetilde{X} = 0\Big)\int_{t=0}^{u} G(\mathrm{d}t),\end{aligned}$$

where $T_{\omega_0} \sim G$ gives the root's age of death. We simplify the equation further into

$$1 - \int_{t=0}^{u} G(\mathrm{d}t) + \mathbb{P}\Big(\widetilde{X} = 0\Big)\int_{t=0}^{u} G(\mathrm{d}t) = 1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t=0}^{u} G(\mathrm{d}t). \tag{3.39}$$

This can then be substituted in

$$\begin{aligned}\mathbb{P}(C^\alpha = 1) &= \alpha\int_{u=0}^{\infty} e^{-\alpha u}\left(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t=0}^{u} G(\mathrm{d}t)\right)\mathrm{d}u \\ &= 1 + \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\alpha\int_{u=0}^{\infty} e^{-\alpha u}\int_{t=0}^{u} G(\mathrm{d}t)\,\mathrm{d}u,\end{aligned} \tag{3.40}$$

since $\alpha e^{-\alpha u}$ is the density of an EXP($\alpha$) random variable. We continue with the integral, where we exchange the integrals such that

$$\begin{aligned}\alpha\int_{u=0}^{\infty} e^{-\alpha u}\int_{t=0}^{u} G(\mathrm{d}t)\,\mathrm{d}u &= \int_{t=0}^{\infty}\alpha\int_{u=t}^{\infty} e^{-\alpha u}\,\mathrm{d}u\,G(\mathrm{d}t) \\ &= \int_{t=0}^{\infty} e^{-\alpha t}\alpha\int_{v=0}^{\infty} e^{-\alpha v}\,\mathrm{d}v\,G(\mathrm{d}t).\end{aligned}$$

We again recognize the density of an exponential. Moreover, the other integral is a multiple of $\hat{\mu}(\alpha)$, which equals 1. For this reason we have

$$\int_{t=0}^{\infty} e^{-\alpha t}\alpha\int_{v=0}^{\infty} e^{-\alpha v}\,\mathrm{d}v\,G(\mathrm{d}t) = \frac{1}{m}\int_{t=0}^{\infty} e^{-\alpha t}m\,G(\mathrm{d}t) = \frac{1}{m}. \tag{3.41}$$

Returning to (3.40), we arrive at

$$\mathbb{P}(C^\alpha = 1) = 1 + \frac{1}{m}\Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big),$$

which coincides with the Bienaymé-Galton-Watson case.

We have again two cases where $\{\widetilde{Y}_u = 2\}$ occurs. The root gives birth to one individual with the same label before time $u$, but the newborn individual is still alive, which is the event $\{T_{\omega_0} < u, \widetilde{X}_{\omega_0} = 1, T_{\omega_0} + T_{\omega_1} > u\}$. Or the root also gives birth to one individual with the same label, but that individual produces no offspring with the ancestral label before time $u$. This is given by $\{\widetilde{X}_{\omega_0} = 1, \widetilde{X}_{\omega_1} = 0, T_{\omega_0} + T_{\omega_1} < u\}$. Thus

$$\begin{aligned}\mathbb{P}\Big(\widetilde{Y}_u = 2\Big) &= \mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1, T_{\omega_0} < u, T_{\omega_0} + T_{\omega_1} > u\Big) + \mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1, \widetilde{X}_{\omega_1} = 0, T_{\omega_0} + T_{\omega_1} < u\Big) \\ &= \mathbb{P}\Big(\widetilde{X}_{\omega_0} = 1\Big)\Big(\mathbb{P}(T_{\omega_0} < u, T_{\omega_0} + T_{\omega_1} > u) + \mathbb{P}\Big(\widetilde{X}_{\omega_1} = 0\Big)\,\mathbb{P}(T_{\omega_0} + T_{\omega_1} < u)\Big).\end{aligned}$$

We continue with the terms within the parentheses and rewrite them using the same step as in (3.39), which gives

$$\begin{aligned}&\mathbb{P}(T_{\omega_0} < u, T_{\omega_0} + T_{\omega_1} > u) + \mathbb{P}\Big(\widetilde{X}_{\omega_1} = 0\Big)\,\mathbb{P}(T_{\omega_0} + T_{\omega_1} < u) \\ &\quad= \int_{t_0=0}^{u}\int_{t_1=u-t_0}^{u} G(\mathrm{d}t_1)\,G(\mathrm{d}t_0) + \mathbb{P}\Big(\widetilde{X} = 0\Big)\int_{t_0=0}^{u}\int_{t_1=0}^{u-t_0} G(\mathrm{d}t_1)\,G(\mathrm{d}t_0) \\ &\quad= \int_{t_0=0}^{u}\Big(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t_1=0}^{u-t_0} G(\mathrm{d}t_1)\Big)\,G(\mathrm{d}t_0).\end{aligned}$$

We substitute it in the following expression, and exchange integrals once again

$$\begin{aligned}\mathbb{P}(C^\alpha = 2) &= \alpha\int_{u=0}^{\infty} e^{-\alpha u}\int_{t_0=0}^{u}\Big(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t_1=0}^{u-t_0} G(\mathrm{d}t_1)\Big)\,G(\mathrm{d}t_0)\,\mathrm{d}u \\ &= \int_{t_0=0}^{\infty}\alpha\int_{u=t_0}^{\infty} e^{-\alpha u}\Big(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t_1=0}^{u-t_0} G(\mathrm{d}t_1)\Big)\,\mathrm{d}u\,G(\mathrm{d}t_0),\end{aligned}$$

we can do a substitution again, which gives

$$\begin{aligned}&\int_{t_0=0}^{\infty} e^{-\alpha t_0}\alpha\int_{v=0}^{\infty} e^{-\alpha v}\Big(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t_1=0}^{v} G(\mathrm{d}t_1)\Big)\,\mathrm{d}v\,G(\mathrm{d}t_0) \\ &\quad= \int_{t_0=0}^{\infty} e^{-\alpha t_0}\,G(\mathrm{d}t_0)\alpha\int_{v=0}^{\infty} e^{-\alpha v}\Big(1 - \Big(1 - \mathbb{P}\Big(\widetilde{X} = 0\Big)\Big)\int_{t_1=0}^{v} G(\mathrm{d}t_1)\Big)\,\mathrm{d}v,\end{aligned}$$

where the first integral is equal to (3.41) and the second one is from (3.40). We arrive at

$$\mathbb{P}(C^\alpha = 2) = \frac{1}{m}\,\mathbb{P}\Big(\widetilde{X} = 1\Big)\,\mathbb{P}(C^\alpha = 1),$$

which is again the same as in the Bienaymé-Galton-Watson case.

We have shown for the two models we considered, that the values $\mathbb{P}(C^\alpha = n)$ coincide for $n = 1, 2$. Due to a series of simulations, it is suspected that the probability mass functions for the two models are equal. This would mean that the choice for the age distribution, either general or constant, is irrelevant for the empirical cluster size distribution, as it would only depend on the

choice of $X$. Moreover, if the conjecture is true, then an explicit expression for each $n \in \mathbb{N}$ holds for both models. By assuming that the relation

$$\mathbb{P}(C^\alpha = n) = \frac{1}{m}\,\mathbb{P}\left(1 + \sum_{i=1}^{\widetilde{X}} C_i^\alpha = n\right),$$

holds for $n \geq 2$, where the $C_i^\alpha$'s are i.i.d. copy's of $C^\alpha$, one can derive the implicit formula

$$Q_{C^\alpha}(z) = z(1 - \frac{1}{m}) + z\frac{1}{m}Q_{\widetilde{X}}(Q_{C^\alpha}(z)),$$

where $Q_{C^\alpha}$ and $Q_{\widetilde{X}}$ are the probability generating functions of $C^\alpha$ and $\widetilde{X}$ respectively. Using the Lagrange Inversion Formula [16], the coefficients of $Q_{C^\alpha}$ are given explicitly, i.e. we can give an explicit expression for $\mathbb{P}(C^\alpha = n)$ for all $n \in \mathbb{N}$. However, this formula already fails for $n = 3$, but only by a small margin. This effort did not yield the desired result, but this may inspire a different approach that will pay off. For this reason, we formulate the following conjecture.

**Conjecture 1.** *The probability mass function for the empirically observed cluster sizes is the same whether the branching process follows a Bienaymé-Galton-Watson process or an age-dependent process, provided that $\mathbb{E}[X] = m > 1$ holds for the offspring distribution $X$ and the Malthusian parameter satisfies $\alpha \in (0, 1)$. Moreover, for all $n \in \mathbb{N}$ there exist an explicit expression for $\mathbb{P}(C^\alpha = n)$ in terms of a finite sum, depending only on $m$ and the probability mass function of $\widetilde{X}$.*

### 3.4.4 Observing cluster sizes on a downsampled tree

In real-world spreading processes, it is not always the case that every infected individual can be observed. This is known as partial sampling. We assume that each individual is independently sampled with probability $s$. Referring back to Section 2.2, where the assumption is that each cluster is a branching process on its own, the observed cluster size was denoted by $\widetilde{Y}_\infty$ with probability mass function (2.4). Since each individual is independently sampled, the observed downsampled cluster size follows a $\mathrm{BIN}(\widetilde{Y}_\infty, s)$ distribution. Hence the probability of observing a downsampled cluster of size $k$ is given by

$$\mathbb{P}\Big(\mathrm{BIN}(\widetilde{Y}_\infty, s) = n\Big) = \sum_{k=n}^{\infty} \mathbb{P}\Big(\widetilde{Y}_\infty = k\Big)\,\mathbb{P}(\mathrm{BIN}(k, s) = n), \tag{3.42}$$

where we include the case $n = 0$. Since clusters of size 0 can not be observed in general, this has to be taken into account. The expression in (3.42) can be normalized by $1 - \mathbb{P}(\mathrm{BIN}(\widetilde{Y}_\infty, s) = 0)$, for an applicable statistical model.

In this section we show that we obtain an analogue of (3.42), for the case where we equip the model in Section 3.4 with downsampling. Intuitively, this result follows from replacing the term $\mathbb{P}(\widetilde{Y}_u = n)$ with $\mathbb{P}(\mathrm{BIN}(\widetilde{Y}_u, s) = n)$ in (3.35) and exchange the integral and sum. By constructing an appropriate characteristic, we show that this intuition is correct, up to a normalization factor.

Before we can construct the characteristic, we need to introduce a function on $\Omega$ that determines whether an individual is sampled. When we defined the reproduction process $\tilde{\xi}$ in (3.24) for the total size $\widetilde{Y}_t$ up to time $t$, we implied the existence of birth times $\tilde{\sigma}_x$, for $x \in I$ generated by $\tilde{\xi}$. These birth times are used to define $\widetilde{Y}_t$ in the same way as in (3.4).

We introduce the indicator function $\phi : \Omega \to \{0, 1\}$ as

$$\phi(\omega_y) = \mathbb{1}(\text{an individual with life career } \omega_y \text{ is sampled}).$$

We stated at the beginning of this section that we assume to sample each individual independently with probability $s$. Moreover, we assume that the outcome $\phi$ is independent of the reproduction process and $\gamma$, which assigns the labels. This makes $\{\phi(\omega_y)\}_{y\in I}$ a sequence of i.i.d. Bernoulli random variables with probability $s$.

Now, we construct the process that counts the number of individuals which are both sampled and wearing the ancestral label. The process is defined as

$$\widetilde{Y}_t^{\phi} = \sum_{x\in I} \mathbb{1}_{[\tilde{\sigma}_x,\infty)}(t)\phi(\omega_x),$$

which is a branching process counted by the characteristic

$$\chi^{\phi}(t) = \mathbb{1}_{\mathbb{R}^+}(t)\phi(\omega_0) \tag{3.43}$$

With $\widetilde{Y}_t^{\phi}$ we are able to construct the characteristic which counts the number of empirically observed downsampled clusters of size $n$. The characteristic is the same as in (3.25), but $\widetilde{Y}_t^{\phi}$ plays the role of $\widetilde{Y}_t$. We have

$$\chi^{n,\phi}(t) = \sum_{i\in\mathbb{N}} \mathbb{1}_{\mathbb{R}^+}(t-\sigma_i)(1-\gamma(i,\omega_0))\mathbb{1}_{\{n\}}\Big(\widetilde{Y}^{\phi}_{t-\sigma_i}(i)\Big),$$

and for the total number of empirically observed downsampled clusters

$$\chi^{\mathbb{N},\phi}(t) = \sum_{i\in\mathbb{N}} \mathbb{1}_{\mathbb{R}^+}(t-\sigma_i)(1-\gamma(i,\omega_0))\mathbb{1}_{\mathbb{N}}\Big(\widetilde{Y}^{\phi}_{t-\sigma_i}(i)\Big).$$

Note that the term $\mathbb{1}_{\mathbb{N}}(\widetilde{Y}^{\phi}_{t-\sigma_i}(i))$ implies that we are not able observe to clusters of size 0. Similar to Definition 3.29, we define the limiting distribution of the empirically observed downsampled clusters as the random variable $C^{\alpha,\phi}$ with probability mass function

$$\mathbb{P}\big(C^{\alpha,\phi}=n\big) = \frac{\int_0^\infty e^{-\alpha u}\,\mathbb{E}\big[\chi^{n,\phi}(u)\big]\,\mathrm{d}u}{\int_0^\infty e^{-\alpha u}\,\mathbb{E}[\chi^{\mathbb{N},\phi}(u)]\,\mathrm{d}u}, \tag{3.44}$$

where $\chi^{n,\phi}$ and $\chi^{\mathbb{N},\phi}$ are as in (3.4.4) and (3.4.4) respectively. We skip the derivations leading to the expressions

$$\int_0^\infty e^{-\alpha u}\,\mathbb{E}\big[\chi^{n,\phi}(u)\big]\,\mathrm{d}u = \int_0^\infty e^{-\alpha u}\,\mathbb{P}\Big(\widetilde{Y}_u^{\phi}=n\Big)\,\mathrm{d}u, \tag{3.45}$$

and

$$\int_0^\infty e^{-\alpha u}\,\mathbb{E}\big[\chi^{\mathbb{N},\phi}(u)\big]\,\mathrm{d}u = \int_0^\infty e^{-\alpha u}\,\mathbb{P}\Big(\widetilde{Y}_u^{\phi}\geq 1\Big)\,\mathrm{d}u, \tag{3.46}$$

as they run analogous to the computations in Section 3.4.1 and 3.4.2. In order to arrive at the analogue of (3.42), we do have to show that $\widetilde{Y}_u^{\phi} \sim \mathrm{BIN}(\widetilde{Y}_u, s)$. Which prove in the following claim.

**Claim 3.** *The branching process $\widetilde{Y}_t^{\phi}$ counted by $\chi^{\phi}$ as in* (3.43) *has probability mass function*

$$\mathbb{P}\Big(\widetilde{Y}_t^{\phi}=n\Big) = \mathbb{P}\Big(\mathrm{BIN}(\widetilde{Y}_t,s)=n\Big) = \sum_{k=n}^{\infty}\mathbb{P}\Big(\widetilde{Y}_t=k\Big)\,\mathbb{P}(\mathrm{BIN}(k,s)=n). \tag{3.47}$$

*Proof.* We start with applying the law of total probability by conditioning on $\widetilde{Y}_t$

$$\mathbb{P}\Big(\widetilde{Y}_t^{\phi} = n\Big) = \sum_{k=n}^{\infty} \mathbb{P}\Big(\widetilde{Y}_t^{\phi} = n \;\Big|\; \widetilde{Y}_t = k\Big)\, \mathbb{P}\Big(\widetilde{Y}_t = k\Big).$$

We proceed with the conditional probability. Since the characteristic $\chi^{\phi}$ vanishes for negative arguments, we can apply the decomposition

$$\widetilde{Y}_t^{\phi} = \sum_{x \in I} \chi_x^{\phi}(t - \sigma_x) = \sum_{\tilde{\sigma}_x \leq t} \chi_x^{\phi}(t - \sigma_x) = \sum_{\tilde{\sigma}_x \leq t} \phi(\omega_x),$$

as birth times with $\tilde{\sigma} - t \geq 0$ imply that $\chi_x^{\phi}(t - \sigma_x) = 0$. Conditional on $\{\widetilde{Y}_t = k\}$, we see that $|\{\tilde{\sigma}_x \leq t\}| = k$. This gives

$$\mathbb{P}\Big(\widetilde{Y}_t^{\phi} = n \;\Big|\; \widetilde{Y}_t = k\Big) = \mathbb{P}\left(\sum_{\tilde{\sigma}_x \leq t} \phi(\omega_x) = n \;\middle|\; \widetilde{Y}_t = k\right) = \mathbb{P}\left(\sum_{i=1}^{k} \phi(\omega_i) = n\right),$$

as the outcome of $\phi$ is independent of the reproduction process. Since $\{\phi(\omega_i)\}_{i=1}^{k}$ is an i.i.d. sequence of Bernoulli random variables with parameter $s$, we arrive at

$$\mathbb{P}\left(\sum_{i=1}^{k} \phi(\omega_i) = n\right) = \mathbb{P}(\mathrm{BIN}(k, s) = n),$$

which concludes the proof. □

We are almost done with deriving the expression for (3.44). With the preceding claim we have computed the integrand of the numerator. The denominator gives a normalization term which is not as simple as the $1/\alpha$ we obtained in (3.35). This is due to the fact that $\{\widetilde{Y}_u^{\phi} = 0\}$ occurs when every individual up to time $u$ is not sampled.

Observe that (3.47) is valid for $n \geq 0$, so we can apply it to (3.46) to obtain

$$\begin{aligned}\int_0^{\infty} e^{-\alpha u}\, \mathbb{P}\Big(\widetilde{Y}_u^{\phi} \geq 1\Big)\, \mathrm{d}u &= \int_0^{\infty} e^{-\alpha u}\Big(1 - \mathbb{P}\Big(\widetilde{Y}_u^{\phi} = 0\Big)\Big)\, \mathrm{d}u \\ &= \frac{1}{\alpha} - \int_0^{\infty} e^{-\alpha u} \sum_{k=1}^{\infty} \mathbb{P}\Big(\widetilde{Y}_u = k\Big)\, \mathbb{P}(\mathrm{BIN}(k, s) = 0)\, \mathrm{d}u.\end{aligned}$$

We compute the probability that the outcome of a binomial is 0 and arrive by exchanging sum and integration at

$$\begin{aligned}\frac{1}{\alpha} - \int_0^{\infty} e^{-\alpha u} \sum_{k=1}^{\infty} \mathbb{P}\Big(\widetilde{Y}_u = k\Big)\, \mathbb{P}(\mathrm{BIN}(k, s) = 0)\, \mathrm{d}u &= \frac{1}{\alpha} - \sum_{k=1}^{\infty} \int_0^{\infty} e^{-\alpha u}\, \mathbb{P}\Big(\widetilde{Y}_u = k\Big)\, \mathrm{d}u (1 - s)^k \\ &= \frac{1}{\alpha} - \frac{1}{\alpha} G_{C_\alpha}(1 - s),\end{aligned} \tag{3.48}$$

where we recognize the expression in (3.35) up to a factor $\alpha$. Here, $G_{C_\alpha}$ denotes the probability generating function of $C_\alpha$. At last we compute the numerator. We exchange sum and integration by substituting (3.47) in (3.45), which gives

$$\int_0^{\infty} e^{-\alpha u} \sum_{k=n}^{\infty} \mathbb{P}\Big(\widetilde{Y}_u = k\Big)\, \mathbb{P}(\mathrm{BIN}(k, s) = n)\, \mathrm{d}u = \sum_{k=n}^{\infty} \int_0^{\infty} e^{-\alpha u}\, \mathbb{P}\Big(\widetilde{Y}_u = k\Big)\, \mathrm{d}u\, \mathbb{P}(\mathrm{BIN}(k, s) = n). \tag{3.49}$$

We recognize again the term $\mathbb{P}(C^\alpha = k)$ up to a factor $\alpha$ in the summand. Substituting both (3.46) and (3.49) in (3.44) gives

$$\mathbb{P}\big(C^{\alpha,\phi} = n\big) = \frac{\sum_{k=n}^{\infty} \mathbb{P}(C^\alpha = k)\, \mathbb{P}(\mathrm{BIN}(k,s) = n)}{1 - G_{C_\alpha}(1-s)}, \tag{3.50}$$

where we multiplied both sides with $\alpha$.

Deriving an expression for the denominator turns out to be a challenge, as one needs to compute the Laplace transform of the probability generating function of $\widetilde{Y}_u$. However, one can derive an implicit formula.

**Proposition 2.** *For an age-dependent branching process with offspring distribution $X$ with mean $m$, lifetime distribution $G$, and Malthusian parameter $\alpha \in (0,\infty)$, the following formula holds*

$$\mathcal{L}(Q_{Y_\bullet}(z)) = z\mathcal{L}(Q_X(Q_{Y_\bullet}(z))), \quad z \geq 0,$$

*where $\mathcal{L}$ denotes the Laplace transform and $Q_X$ and $Q_{Y_\bullet}$ are the probability generating functions of $X$ and $Y_\bullet$ respectively.*

*Proof.* We write $G_{Y_u}(z)$ as $\mathbb{E}[z^{Y_u}]$ and express $Y_u$ as $Y_u = \mathbb{1}_{\{T_{\omega_0} < u\}}\Big(1 + \sum_{i=0}^{X_{\omega_0}} Y^{(i)}_{u-T_{\omega_0}}\Big) + \mathbb{1}_{\{T_{\omega_0} > u\}}$, where the $Y_u^{(i)}$ are i.i.d. copies of $Y_u$, $X_{\omega_0} \sim X$ and $T_{\omega_0} \sim G$. We apply the law of total expectation and split the integral at $t = u$

$$\begin{aligned}
\mathbb{E}[z^{Y_u}] &= \int_{t=0}^{u} \mathbb{E}\big[z^{Y_u} \;\big|\; T_{\omega_0} = t\big]\, G(\mathrm{d}t) + \int_{t=u}^{\infty} \mathbb{E}\big[z^{Y_u} \;\big|\; T_{\omega_0} = t\big]\, G(\mathrm{d}t) \\
&= \int_{t=0}^{u} \mathbb{E}\Big[z^{1+\sum_{i=0}^{X_{\omega_0}} Y^{(i)}_{u-t}}\Big]\, G(\mathrm{d}t) + \int_{t=u}^{\infty} \mathbb{E}[z]\, G(\mathrm{d}t),
\end{aligned}$$

and consider the first integral. We take out $z$ and condition on $X_{\omega_0}$ which gives

$$\begin{aligned}
\int_{t=0}^{u} \mathbb{E}\Big[z^{1+\sum_{i=0}^{X_{\omega_0}} Y^{(i)}_{u-t}} \;\Big|\; T = t\Big]\, G(\mathrm{d}t) &= z\int_{t=0}^{u} \mathbb{E}\Big[\mathbb{E}\Big[z^{\sum_{i=0}^{X_{\omega_0}} Y^{(i)}_{u-t}} \;\Big|\; X_{\omega_0}\Big]\Big]\, G(\mathrm{d}t) \\
&= z\int_{t=0}^{u} \mathbb{E}\left[\prod_{i=1}^{X_{\omega_0}} \mathbb{E}\Big[z^{Y^{(i)}_{u-t}}\Big]\right] G(\mathrm{d}t),
\end{aligned}$$

since $Y^{(i)}_{u-t}$ is independent of $X_{\omega_0}$. Since the $Y^{(i)}_{u-t}$ are i.i.d. copies of $Y_{u-t}$ and $X_{\omega_0} \sim X$ we obtain

$$\begin{aligned}
z\int_{t=0}^{u} \mathbb{E}\left[\prod_{i=1}^{X_{\omega_0}} \mathbb{E}\Big[z^{Y^{(i)}_{u-t}}\Big]\right] G(\mathrm{d}t) = z\int_{t=0}^{u} \mathbb{E}\Big[\mathbb{E}\big[z^{Y_{u-t}}\big]^X\Big]\, G(\mathrm{d}t) &= z\int_{t=0}^{u} \mathbb{E}\big[Q_{Y_{u-t}}(z)^X\big]\, G(\mathrm{d}t) \\
&= z\int_{t=0}^{u} Q_X\big(Q_{Y_{u-t}}(z)\big)\, G(\mathrm{d}t).
\end{aligned}$$

□

This proposition is related to Conjecture 1, as $\mathcal{L}(Q_{Y_\bullet}(z)) = Q_{C^\alpha}(z)$. By solving either the conjecture or the formula given in the proposition, the denominator in (3.50) can be explicitly computed. As a result, this model would be more applicable to real data, than the model that assumes observation of the entire tree, since it is often the case that only a fraction of a spreading process is visible.

# Chapter 4

# General multi-type branching processes

This chapter serves as an extension to Chapter 3, where the individuals are assumed to be indistinguishable. There are various reasons why we want to relax this assumption. For example, in the context of epidemic spread, where a population can consist of multiple distinguishable types. The transmission does not have to be homogeneous among the various types, resulting in different reproduction processes. Another example is a varying mutation probability $\nu$. In the preceding chapters, it is assumed that $\nu$ is constant, but for each newly introduced allele, we can sample a mutation probability $\nu$ from some distribution. This can be made precise using the multi-type model. Each individual inherits the parents type if it inherits the same label, but an individual assumes a new unique type if a new label is introduced. The mutation rate is then determined by the type of the individual.

## 4.1 An extension of the general single-type branching process

The aforementioned examples indicate that the type space can be broad. Jagers gives in [20] a construction of a general branching with an abstract type space, providing us with the machinery to handle the extension to the multi-type branching process. The naming of the individuals is as in Section 3.1. The life space $(\Omega, \mathcal{A})$ as defined in Definition 1, and thus the times of childbearing $\tau_i$, remain unchanged. The types are elements of some abstract type space $S$ with a countably generated $\sigma$-algebra $\mathcal{S}$. We require that the types are given by the measurable functions $\rho_i : \Omega \to S$, $i \in \mathbb{N}$, where $\rho_i(\omega)$ denotes the type of the $i$-th child of an individual with life $\omega \in \Omega$.

As hinted in the beginning of this section, an individual can produce offspring of different types, where the distribution might depend on the type of the individual itself. For this reason the reproduction process in the multi-type case is defined as an extension of (2), which is

$$\xi(A \times B) = |\{i : \rho_i \in A, \tau_i \in B\}|, \tag{4.1}$$

where $A \in \mathcal{S}$ and $B \in \mathcal{B}(\mathbb{R}^+)$. The appointment of each individuals type is a result of its parent's reproduction process, which is determined by the parent's type. This can recursively be traced back to the type of the root. Hence the population process which was previously defined in Definition 3, now also has to include the type space $S$ and the $\sigma$-algebra $\mathcal{S}$. By including the

root's type in the sampling of the trees, we arrive at the population process

$$(S \times \Omega^I, \mathcal{S} \times \mathcal{A}^I), \tag{4.2}$$

where $\Omega^I$ and $\mathcal{A}^I$ are as in (3.3) and the products in (4.2) represent the Cartesian product and the $\sigma$-algebra generated by the cylinder sets respectively.

A probability measure on (4.1) can be constructed by assuming the existence of a probability measure $P(s, \cdot)$ on the life space $(\Omega, \mathcal{A})$ for each $s \in S$. The functions $s \mapsto P(s, A)$ should be measurable for every $A \in \mathcal{A}$. In [20] they show that this defines a unique probability measure $\mathbb{P}_s$ on the population space (4.2) for each $s \in S$. The $\mathbb{P}_s$-expectation is denoted by $\mathbb{E}_s$. We refer to $\mathbb{E}_r$ as the $r$-expectation. Moreover, in [20], Jagers gives a stronger version of Lemma 2, which implies branching, i.e. the independence of disjoint daughter processes. We skip most of the details, but point out that the conditional independence of $S_x$ for all $x \in \mathbb{N}^{n+1}$ is in the multi-type case with respect to $\mathcal{S} \times \mathcal{A}_n$ for some $n \in \mathbb{N}$, where $\mathcal{A}_n$ is as in (3.8). Furthermore, for each $s \in S$ and $n \in \mathbb{N}$, (3.9) now takes the form

$$\mathbb{P}_s(S_x \in A \mid \mathcal{S} \times \mathcal{A}_n) = \mathbb{P}_{\rho_x}(A),$$

for any $x \in \mathbb{N}^{n+1}$ and $A \in \mathcal{S} \times \mathcal{A}^I$. Here, we use the convention that $\rho_x$ gives the type of individual $x \in I$.

In Section 3.2 the notion of characteristics is introduced as measurable functions on the sampled trees. For this reason, we require in Definition 5 that they are measurable with respect to the product $\sigma$-algebra $\mathcal{B}(\mathbb{R}) \times \mathcal{A}^I$. Because the sampling of trees happens in the multi-type scenario with respect to the population process $(S \times \Omega^I, \mathcal{S} \times \mathcal{A}^I)$, it is only natural to define a multi-type characteristic as any real-valued process $(\chi(t))_{t \in \mathbb{R}}$ for which the map $\chi : \mathbb{R} \times S \times \Omega^I \to \mathbb{R}^+$ is measurable with respect to the product $\sigma$-algebra $\mathcal{B}(\mathbb{R}) \times \mathcal{S} \times \mathcal{A}^I$ and vanishes for $t < 0$. The definition of a branching process counted by characteristic $\chi$ remains unchanged.

The intensity measure of $\xi$ given in (4.1) now takes the following form

$$\mu(r, \mathrm{d}s \times \mathrm{d}u) = \mathbb{E}_r[\xi(\mathrm{d}s \times \mathrm{d}u)],$$

where $r \in S$ denotes the type of the individual. The process is said to be Malthusian with parameter $\alpha \in \mathbb{R}$, if the kernel

$$\hat{\mu}(r, \mathrm{d}s; \alpha) = \int_0^\infty e^{-\alpha u} \mu(r, \mathrm{d}s \times \mathrm{d}u), \tag{4.3}$$

has Perron root 1. How this condition is satisfied is made clear depending on the context.

According to [22], there now exists, if $\alpha > 0$, a $\sigma$-finite measure $\pi$ on the type space $(S, \mathcal{S})$ and a with respect to $\pi$-almost everywhere finite, strictly positive, measurable function $h$ on the same space, such that

$$\int_S \hat{\mu}(r, \mathrm{d}s; \alpha)\, \pi(\mathrm{d}r) = \pi(\mathrm{d}s),$$

$$\int_S h(s)\, \hat{\mu}(r, \mathrm{d}s; \alpha) = h(r).$$

It is assumed that

$$0 < \beta = \int_0^\infty u e^{-\alpha u} h(s) \mu(r, \mathrm{d}s \times \mathrm{d}u)\, \pi(\mathrm{d}r) < \infty,$$

which implies that we can norm $h$ such that $\int h(s)\, \pi(\mathrm{d}s) = 1$. Moreover, under the assumption that $\inf h > 0$, we can also norm $\pi$ to a probability measure. At last we assume that there exists $\varepsilon > 0$ such that

$$\sup_{r \in S} \mu(r, S \times [0, \varepsilon]) < 1. \tag{4.4}$$

To be able to use the techniques of Chapter 3, we need multi-type analogues for Theorem 4 and as a result Corollary 3. Unfortunately, we do not have these convergence results in the almost sure sense for the multi-type case. Because we can restate Theorem 4 with $L^1$-convergence at most, the analogue of Corollary 3 can only be as strong as convergence in probability. However, the conditions to obtain $L^1$-convergence are more relaxed in some sense. The most notable difference is that we require the reproduction process $\xi$ to satisfy what is known as the $x\log(x)$-condition, which we define now.

**Definition 13.** A branching process satisfies the $x\log(x)$-condition if the following holds for the reproduction process

$$\mathbb{E}_\pi\big[{}_\alpha\xi \log^+({}_\alpha\xi)\big] < \infty, \tag{4.5}$$

where $\log^+(x) = \max(0, \log(x))$ and

$${}_\alpha\xi = \int_{S\times\mathbb{R}^+} e^{-\alpha u} h(s)\, \xi(\mathrm{d}s \times \mathrm{d}u).$$

*Remark* 7. In Remark 5 it is pointed out that $e^{-\alpha t} Z_t^\chi$ converges to a multiple of the limiting random variable $W_\infty$. In [27], we find that the condition (4.5) is equivalent to $W > 0$ almost surely on $\{Y_t \to \infty\}$. This allows us to imply the convergence of ratios. Moreover, in the age-dependent case defined in Section 3.4.2, the $x\log(x)$-condition is equivalent to

$$\mathbb{E}\big[\xi \log^+(\xi)\big] < \infty.$$

We are ready to state the following theorem from [22]

**Theorem 7.** *Consider a branching process $Z_t^\chi$ counted by characteristic $\chi$ with reproduction process $\xi$. Assume that $\xi$ satisfies the $x\log(x)$-condition and $\xi(S\times\mathbb{R}^+) < \infty$, and the nonlattice intensity measure $\mu$ has Malthusian parameter $\alpha \in (0,\infty)$ and satisfies (4.4). Moreover, assume that $\chi$ is bounded and $e^{-\alpha u}\,\mathbb{E}_r[\chi(u)]$ as a function of $r$ and $u$ is directly Riemann integrable with respect to $\pi(\mathrm{d}r)\times\mathrm{d}u$[1]. Then, as $t\to\infty$,*

$$e^{-\alpha t} Z_t^\chi \to \int_0^\infty e^{-\alpha u} \frac{W_\infty\, \mathbb{E}_\pi[\chi(u)]}{\alpha\beta}\,\mathrm{d}u \text{ in } L^1(\mathbb{P}_r),$$

*for $\pi$-almost all $r\in S$. Here $W_\infty$ is a nonnegative random variable such that $\mathbb{E}_r[W_\infty] = h(r)$.*

In the single-type case, the $x\log(x)$-condition is enough to ensure that $W_\infty > 0$ on $\{Y_t\to\infty\}$, as mentioned in Remark 7. However, for the multi-type branching process we need to make this assumption stronger by stating the analogue of Theorem 4 in the following way.

**Theorem 8.** *Suppose that the assumptions of Theorem 7 hold, add the assumption that $\inf_{r\in S}\mathbb{P}_s(W_\infty > 0)$, then, as $t\to\infty$,*

$$\frac{Z_t^\chi}{Y_t} \to \int_0^\infty e^{-\alpha u}\,\mathbb{E}_\pi[\chi(u)]\,\mathrm{d}u \text{ in } \mathbb{P}_r\text{-probability},$$

*conditional on $\{Y_t\to\infty\}$, for $\pi$-almost all $r\in S$.*

*Remark* 8. Observe the difference with (3.22), apart from the mode of convergence, that the expectation here is what we denote as the $\pi$-expectation, i.e. we integrate over $\mathbb{E}_r$ on the type space with respect to the stationary measure $\pi(\mathrm{d}r)$.

[1] According to [33], a measurable function $g$ is called directly Riemann integrable if for any $\varepsilon > 0$ we can find $\delta > 0$ and functions $g^-, g^+ \in L^1(\pi(\mathrm{d}r)\times\mathrm{d}u)$ such that for $\pi$-almost all $r\in S$, $g^-(r,\cdot) \le g(r,\cdot) \le g^+(r,\cdot)$, $g^\pm(r,u) = g^\pm(r,n\delta)$ for $n\delta \le u < (n+1)\delta$, and the difference in $L^1(\pi(\mathrm{d}r)\times\mathrm{d}u)$ between $g^-$ and $g^+$ is less than $\varepsilon$.

## 4.2 Varying the probability of mutation

As an extension to the infinite alleles model, we can assume that with every newly introduced allele $a$, we associate a probability of mutation $\nu(a) \in (0,1)$ with all individuals carrying the allele. Rather than using the conventional labelling system, where the first introduced allele is associated with label 1, the second with 2, and so forth, we instead uniformly sample a value from the interval $S = (0,1)$. In this way we can sample a new probability of mutation according to the density $\nu : S \to (0,1)$.

The varying probability of mutation per allele influences the reproduction process $\xi$, as opposed to the assumption where $\nu$ is constant. Individuals that carry an allele which is associated with a large probability of mutation, are more likely to produce children which carry a new allele. For this reason, we need the multi-type extension, and the labels the alleles are associated with, are referred to as types from this point onward.

According to the lines of the beginning of this chapter, the $i$-th child's type is given by $\rho_i$. For an individual with type $r \in S$, we want to assign to each child a new type uniformly from $S$ with probability $\nu(r)$, or assign type $r$ with probability $1 - \nu(r)$. Hence for each $x \in I$ we need to introduce $U_x \sim U \sim U(0,1)$ and $M_{xi} \sim M_{\rho_x} \sim \mathrm{BER}(\nu(\rho_x))$, for all children $i \in \mathbb{N}$. We assume that the sequences $(U_x)_{x\in I}$ and $(M_x)_{x\in I}$ are independent of each other and independent of the ages of childbearing $(\tau_i(\omega_x))_{i\in\mathbb{N}}$ for each $x \in I$. The random variables $M_x$ indicate if a mutation has occurred, and $U_x$ samples a new type. Hence for an individual with life $\omega_x$, the functions $\rho_i$ take the following form

$$\rho_i(\omega_x) = U_{xi}M_{xi} + \rho_x(1 - M_{xi}).$$

It must be pointed out that we assume $\mathcal{S} = \mathcal{B}(S)$.

We immediately consider the application to an age-dependent process with offspring distribution $X$ and a general age distribution $G$, as defined in Section 3.4.2. The reproduction process for an individual with life $\omega$ and type $r \in S$, takes the following form

$$\xi(r, A \times [0,t], \omega) = \left[\sum_{i=1}^{X_\omega} \mathbb{1}_A(U_iM_i + r(1 - M_i))\right] \mathbb{1}_{[T_\omega,\infty)}(t), \tag{4.6}$$

for $A \in \mathcal{S}$, where $X_\omega$ is a copy of $X$ and $T_\omega \sim G$. Taking the expectation of (4.6) gives the intensity measure

$$\mu(r, B \times [0,t]) = \mathbb{E}_r[X]\,\mathbb{P}_r(U_iM_i + r(1 - M_i) \in B)G(t).$$

Computing the Malthusian parameter requires finding $\alpha \in \mathbb{R}$ in equation (4.3). With some abuse of notation, we have

$$\hat{\mu}(r, \mathrm{d}s; \alpha) = \int_0^\infty e^{-\alpha u}\mu(r, \mathrm{d}s \times \mathrm{d}u) = \mathbb{E}_r[X] \int_0^\infty e^{-\alpha u}\, G(\mathrm{d}u)\, \mathbb{P}_r(UM_{\rho_0} + r(1 - M_{\rho_0}) \in \mathrm{d}s),$$

where we recognize the expression (3.34), considering that the number of offspring is independent of the root's type, i.e. $\mathbb{E}[X] = \mathbb{E}_r[X] = m$. For any distribution $\pi$, we find that

$$\begin{aligned}\int_S \mathbb{P}_r(UM_{\rho_0} + r(1 - M_{\rho_0}) \in A)\,\pi(\mathrm{d}r) &= \int_S \mathbb{P}(U \in A)\nu(r)\,\pi(\mathrm{d}r) + \int_S \mathbb{1}_A(r)(1 - \nu(r))\,\pi(\mathrm{d}r) \\ &= |A| \int_S \nu(r)\,\pi(\mathrm{d}r) + \pi(A) - \int_A \nu(r)\,\pi(\mathrm{d}r),\end{aligned}$$

and therefore for any eigendistribution, we must have

$$\pi(A) = \hat{\mu}(\alpha)\left(|A| \int_S \nu(r)\,\pi(\mathrm{d}r) + \pi(A) - \int_A \nu(r)\,\pi(\mathrm{d}r)\right). \tag{4.7}$$

Assuming $\pi(A) = \hat{\mu}(\alpha)\pi(A)$, we obtain that the Malthusian parameter is the same as in Section 3.4.2.

*Remark* 9. This assumption only holds if the kernel has Perron root $\hat{\mu}(\alpha)$, which we did not show. However, we motivate this assumption by consulting the intuition which argues that the infinite allele labelling process does influence the offspring distribution.

A suitable candidate which gives the equality in (4.7) is $\pi(\mathrm{d}r) = \frac{N_\nu}{\nu(r)}\,\mathrm{d}r$, such that

$$\int_A \nu(r)\frac{N_\nu}{\nu(r)}\,\mathrm{d}r = N_\nu \int_A \mathrm{d}r = N_\nu |A| \int_S \mathrm{d}r = |A| \int_S \nu(r)\frac{N_\nu}{\nu(r)}\,\mathrm{d}r,$$

where

$$N_\nu^{-1} = \int_S \nu(r)^{-1}\,\mathrm{d}r.$$

Note that we made an implicit assumption about $\nu$. Specifically, the inverse $\nu^{-1}$ must be integrable over $S$ with respect to the Lebesgue measure.

We may again use the characteristics $\chi^n$ and $\chi^{\mathbb{N}}$ defined in (3.25) and (3.26). We obtain the multi-type analogue for the probability mass function in (3.27) defined in Definition 12, by applying the following corollary of Theorem 8.

**Corollary 4.** *Let $Z_t^\chi$ and $Z_t^{\chi'}$ be branching process counted by $\chi$ and $\chi'$ respectively, such that both $\chi$ and $\chi'$ and $\mu$ satisfy the conditions of Theorem 7. Then, as $t \to \infty$,*

$$\frac{Z_t^\chi}{Z_t^{\chi'}} \to \frac{\int_0^\infty e^{-\alpha u}\,\mathbb{E}_\pi[\chi(u)]\,\mathrm{d}u}{\int_0^\infty e^{-\alpha u}\,\mathbb{E}_\pi[\chi'(u)]\,\mathrm{d}u} \text{ in } \mathbb{P}_r\text{-probability},$$

*conditional on $\{Y_t \to \infty\}$, for $\pi$-almost all $r \in S$.*

In order to compute

$$\mathbb{P}(C^\alpha = n) = \frac{\int_0^\infty e^{-\alpha u}\,\mathbb{E}_\pi[\chi^n(u)]\,\mathrm{d}u}{\int_0^\infty e^{-\alpha u}\,\mathbb{E}_\pi[\chi^{\mathbb{N}}(u)]\,\mathrm{d}u}, \tag{4.8}$$

we must evaluate the expectations in the integrands. This is similar to the computations that lead to (3.35). However, when we take the expectation of $1-\gamma(i,\omega_0)$ in the single-type case, this gives the constant mutation rate $\nu$. We are now under the assumption that a mutation occurs with probability $\nu(r)$, if the root is of type $r \in S$. Hence, taking the $\pi$-expectation of $1-\gamma(i,\omega_0)$ gives

$$\mathbb{E}_\pi[1-\gamma(i,\omega_0)] = \int_S \mathbb{E}_r[1-\gamma(i,\omega_0)]\,\pi(\mathrm{d}r) = \int_S \nu(r)\,\pi(\mathrm{d}r),$$

where $\nu(r)$ is the probability of mutation for the root. Writing $\pi(\mathrm{d}r)$ as a density with respect to the Lebesgue measure then gives

$$\int_S \nu(r)\frac{N_\nu}{\nu(r)}\,\mathrm{d}r = \int_S N_\nu\,\mathrm{d}r = N_\nu. \tag{4.9}$$

As a result we have

$$\mathbb{E}_\pi[\chi^n(u)] = N_\nu\,\mathbb{E}_\pi[X],$$

and

$$\mathbb{E}_\pi[\chi^{\mathbb{N}}(u)] = N_\nu\,\mathbb{E}_\pi[X].$$

Substituting these equations in (4.8) gives similarly to the single-type case the following expression

$$\mathbb{P}(C^\alpha = n) = \alpha \int_0^\infty e^{-\alpha u}\,\mathbb{P}_\pi\Big(\widetilde{Y}_u = n\Big)\,\mathrm{d}u. \tag{4.10}$$

Computing the mean of $C^\alpha$ with the probability mass function given above is again similar to the single-type case. Along the lines of equations (3.36)-(3.38), we obtain

$$\mathbb{E}[C^\alpha] = \alpha \int_0^\infty e^{-\alpha u}\, \mathbb{E}_\pi\left[\widetilde{Y}_u\right] \mathrm{d}u = 1 + \mathbb{E}_\pi\left[\widetilde{X}\right] \int_{t=0}^\infty e^{-\alpha t}\, G(\mathrm{d}t)\alpha \int_{v=0}^\infty e^{-\alpha v}\, \mathbb{E}_\pi\left[\widetilde{Y}_v\right] \mathrm{d}v,$$

where we used the fact that

$$\mathbb{E}_\pi\left[\widetilde{Y}_v(i)\right] = \int_S \mathbb{E}_\pi\left[\widetilde{Y}_v(i) \;\middle|\; \rho_i = r\right] \mathbb{P}_\pi(\mathrm{d}r) = \int_S \mathbb{E}_r\left[\widetilde{Y}_v\right] \pi(\mathrm{d}r) = \mathbb{E}_\pi\left[\widetilde{Y}_v\right]. \tag{4.11}$$

We compute $\mathbb{E}_\pi\left[\widetilde{X}\right]$ by observing that the $r$-expectation of $\widetilde{X}$ is equal to the expectation of a $\mathrm{BIN}(X, 1-\nu(r))$ random variable, which gives

$$\mathbb{E}_\pi\left[\widetilde{X}\right] = \int_S \mathbb{E}_\pi\left[\widetilde{X}\right] \pi(\mathrm{d}r) = m \int_S (1-\nu(r))\, \pi(\mathrm{d}r).$$

Identical to the computation in (4.9), we obtain

$$m \int_S (1-\nu(r))\, \pi(\mathrm{d}r) = m(1-N_\nu),$$

and

$$\begin{aligned}\mathbb{E}[C^\alpha] &= 1 + (1-N_\nu) \int_{t=0}^\infty e^{-\alpha t} m\, G(\mathrm{d}t)\alpha \int_{v=0}^\infty e^{-\alpha v}\, \mathbb{E}_\pi\left[\widetilde{Y}_v\right] \mathrm{d}v \\ &= 1 + (1-N_\nu)\, \mathbb{E}[C^\alpha].\end{aligned}$$

The mean of $C^\alpha$ is therefore given by

$$\mathbb{E}[C^\alpha] = N_\nu^{-1} = \int_S \nu(r)^{-1}\, \mathrm{d}r,$$

which includes the case where $\nu$ is constant.

## 4.3 Multi-type age-dependent model

We consider the age-dependent model in the multi-type case, as an extension of Section 3.4.2. The goal of this section is to show that Lemma 5 also holds when the offspring distribution is not homogeneous. As mentioned in the introduction in Chapter 1, we aim to eventually extend the results towards inferring the age-contact matrices. As this requires us to divide the population into a finite number of age groups, we define our type space to be $S = \{1, 2, \ldots, N\}$, for some $N \in \mathbb{N}$. We equip $S$ with the $\sigma$-algebra $\mathcal{S} = 2^S$. We assume in contrary to the previous section, that the probability of mutation $\nu \in (0,1)$ is constant.

For an individual of type $r \in S$, we assume that it produces a random vector of offspring $(X_{rs})_{s\in S}$, where $X_{rs}$ denotes the number of offspring with type $s$. Note that the $X_{rs}$'s are not necessarily independent. One could assume a type-dependent age distribution $G_r$ for each $r \in S$, but we impose the restriction that $G_r \equiv G$ for each $r \in S$, for some general age distribution $G$. The matrix $M = (m_{rs})_{(r,s)\in S^2}$ whose elements are given by $m_{rs} = \mathbb{E}[X_{rs}]$, is denoted as the mean matrix of the branching process.

The reproduction process for an individual with life $\omega$ and type $r \in S$, is of the form

$$\xi(r, \{s\} \times [0,t], \omega) = X_{rs,\omega} \mathbb{1}_{[T_\omega, \infty)}(t),$$

for $s \in S$. Here, $X_{rs,\omega} \sim X_{rs}$ and $T_\omega \sim G$. According to [1], the Malthusian parameter is given by the number $\alpha$ such that the matrix $\hat{M}(\alpha) = (\hat{m}_{rs})_{(r,s)\in S^2}$, whose elements are $\hat{m}_{rs} = m_{rs} \int e^{-\alpha u} G_r(\mathrm{d}u)$, has largest eigenvalue 1. Since we assume that $G_r \equiv G$ for each $r \in S$, the matrix $\hat{M}(\alpha)$ is given by

$$\hat{M}(\alpha) = \int_0^\infty e^{-\alpha u} G(\mathrm{d}u) M.$$

Let $\lambda_i$ denote the eigenvalues of $M$, for $1 \leq i \leq N$. We define $\rho = \max_i \lambda_i$ as the largest eigenvalue of $M$. Hence, the eigenvalues of $\hat{M}(\alpha)$ are given by $\int e^{-\alpha u} G(\mathrm{d}u)\lambda_i$. In order to find the Malthusian parameter, we must therefore compute $\alpha$ such that

$$\int e^{-\alpha u} \rho \, G(\mathrm{d}u) = 1.$$

We point out that $\rho$ has taken over the role of $\mathbb{E}[X] = m$ in (3.34).

The eigendistribution, $\pi$, is now given by a vector in $\mathbb{R}^{1\times N}$. More specifically, it is given by the left eigenvector $\pi$ such that $\pi\hat{M}(\alpha) = \pi$, or equivalently,

$$\pi M = \rho\pi.$$

As mentioned in Section 4.1, we require that the vector $\pi = (\pi(r))_{r\in S}$ sums up to 1.

We continue with the computation of $\mathbb{P}(C^\alpha = n)$, as expressed in (4.8), in order to derive the mean of the average observed cluster size for a multi-type age-dependent process. The computations are again similar to the derivations in Section 3.4.2. However, one arrives at

$$\mathbb{E}_\pi[\chi^n(u)] = \nu \sum_{i\in\mathbb{N}} \sum_{s\in S} \mathbb{E}_\pi[\mathbb{1}(X_{\rho_0 s} \geq i)] \int_0^u \mathbb{P}_\pi\Big(\widetilde{Y}_t = n\Big) \, G(\mathrm{d}t).$$

By separately considering the double sum, we observe that

$$\sum_{i\in\mathbb{N}} \sum_{s\in S} \mathbb{E}_\pi[\mathbb{1}(X_{\rho_0 s} \geq i)] = \sum_{i\in\mathbb{N}} \sum_{s\in S} \sum_{r\in S} \mathbb{E}_r[\mathbb{1}(X_{\rho_0 s} \geq i)]\pi(r) = \sum_{s\in S} \sum_{r\in S} \sum_{i\in\mathbb{N}} \mathbb{E}[\mathbb{1}(X_{rs} \geq i)]\pi(r),$$

where we recognize that the inner sum over $i$ gives the mean of $X_{rs}$. Which leads to

$$\sum_{s\in S} \sum_{r\in S} \mathbb{E}[X_{rs}]\pi(r) = \sum_{s\in S} \sum_{r\in S} \pi(r) m_{rs} = \sum_{s\in S} \rho\pi(s) = \rho, \tag{4.12}$$

as $\pi$ is an eigenvector of the mean matrix $M$ corresponding to eigenvalue $\rho$. In the last step we use that $\pi$ is normed to 1. We obtain

$$\mathbb{E}_\pi[\chi^n(u)] = \nu\rho \int_0^u \mathbb{P}_\pi\Big(\widetilde{Y}_t = n\Big) \, G(\mathrm{d}t),$$

which gives

$$\int_0^\infty e^{-\alpha u} \, \mathbb{E}_\pi[\chi^n(u)] \, \mathrm{d}u = \nu \int_0^\infty e^{-\alpha u} \, \mathbb{P}_\pi\Big(\widetilde{Y}_u = n\Big) \, \mathrm{d}u,$$

as $e^{-\alpha u}\rho \, G(\mathrm{d}u)$ integrates to 1. As a result we obtain the same expression for $\mathbb{P}(C^\alpha = n)$ as in (4.10).

We may argue that the computation for the mean of $C^\alpha$, leads to the expression

$$\mathbb{E}_\pi[C^\alpha] = \alpha \int_0^\infty e^{-\alpha u} \, \mathbb{E}_\pi\Big[\widetilde{Y}_u\Big] \, \mathrm{d}u = 1 + \sum_{s\in S} \mathbb{E}_\pi\Big[\widetilde{X}_{\rho_0 s}\Big] \int_{t=0}^\infty e^{-\alpha t} \, G(\mathrm{d}t) \, \mathbb{E}_\pi[C^\alpha], \tag{4.13}$$

using a similar argument as in (4.11), where $\widetilde{X}_{\rho_0 s} \sim \text{BIN}(X_{\rho_0 s}, 1-\nu)$. The summation can be simplified to

$$\sum_{s\in S} \mathbb{E}_\pi \left[\widetilde{X}_{\rho_0 s}\right] = (1-\nu) \sum_{s\in S} \mathbb{E}[X_{rs}]\pi(r) = (1-\nu)\rho,$$

which was derived in (4.12). Substituting this in (4.13), gives

$$\mathbb{E}[C^\alpha] = 1 + (1-\nu) \int_{t=0}^{\infty} e^{-\alpha t} \rho \, G(\mathrm{d}t) \, \mathbb{E}[C^\alpha],$$

yielding

$$\mathbb{E}[C^\alpha] = \nu^{-1}.$$

# Chapter 5

# Application: A simulation study

In this chapter, the models derived in earlier chapters are applied in two simulation studies, one which strengthens the importance Theorem 5 and the other which demonstrates Theorem 6.

The goal of the first study is to compare the random variables $\widetilde{Y}_\infty$ and $C^\alpha$ as a model for the empirically observed cluster sizes in a branching process, equipped with the infinite alleles model. The probability mass functions of the two random variables are respectively given by (2.4) and (3.30). The probability mass function of $\widetilde{Y}_\infty$ is used by [35] to infer $(R, k)$ on real epidemiological data, assuming an offspring distribution $X \sim \text{NEGBIN}(k, q)$, with $\mathbb{E}[X] = R$. We argue in Section 2.3 that the empirically observed cluster size distribution is influenced by the exponential growth of the branching process. This specifically happens when the clusters are observed in a supercritical tree, i.e. with Malthusian parameter $\alpha \in (0, \infty)$. Because this is merely a demonstration, we restrict ourselves to only showing a comparison for one combination of parameters $R$, $k$ and $\nu$. To be able to apply the probability mass function of $C^\alpha$, we need the branching process to be supercritical, which requires $R > 1$. The probability mass function of $\widetilde{Y}_\infty$ is proper if $(1 - \nu)R \leq 1$. In order to be in the regime where both models are applicable, which is also consistent with COVID-19 estimates, we choose the parameters $R = 1.75$, $k = 0.5$ and $\nu = 0.5$.

In Section 3.4.3, we explicitly computed the $\mathbb{P}(C^\alpha = 1)$ and $\mathbb{P}(C^\alpha = 2)$. Continuing this reasoning, one can also show that

$$\mathbb{P}(C^\alpha = 3) = \frac{1}{m}\,\mathbb{P}\Big(\widetilde{X} = 2\Big)\left(1 - \frac{1}{m}\left(1 - \mathbb{P}\Big(\widetilde{X}_{\omega_0} = 0\Big)^2\right)\right) + \frac{1}{m}\,\mathbb{P}\Big(\widetilde{X} = 1\Big)\,\mathbb{P}(C^\alpha = 2).$$

We use the convention that all clusters of size $n \geq 4$, are observed with probability $\mathbb{P}(C^\alpha \geq 4) = 1 - \mathbb{P}(C^\alpha \leq 3)$. We perform maximum likelihood estimation with this probability mass function, which is referred to as the "Malthusian" model, because this model captures the emerging of cluster sizes under Malthusian growth. The probabilities $\mathbb{P}(\widetilde{Y}_\infty = n)$ can explicitly be computed for all $n \geq 1$, and this model is referred to as "Progeny", as it is given by the law of total progeny in Theorem 1.

The results obtained by the simulations are shown in the box plots in Figure 5.1. The Progeny model consistently underestimates both values, with very few outliers. Additionally, the distance between the first and third quartiles is small compared to the Malthusian model. In contrast, the Malthusian model shows a higher variance, but it demonstrates higher accuracy. Despite the variance in the Malthusian model appearing to be at least an order of magnitude larger, its accuracy is very promising. Especially considering the fact that only the first three values of the probability mass function were computed, which is a plausible explanation for the high variance.

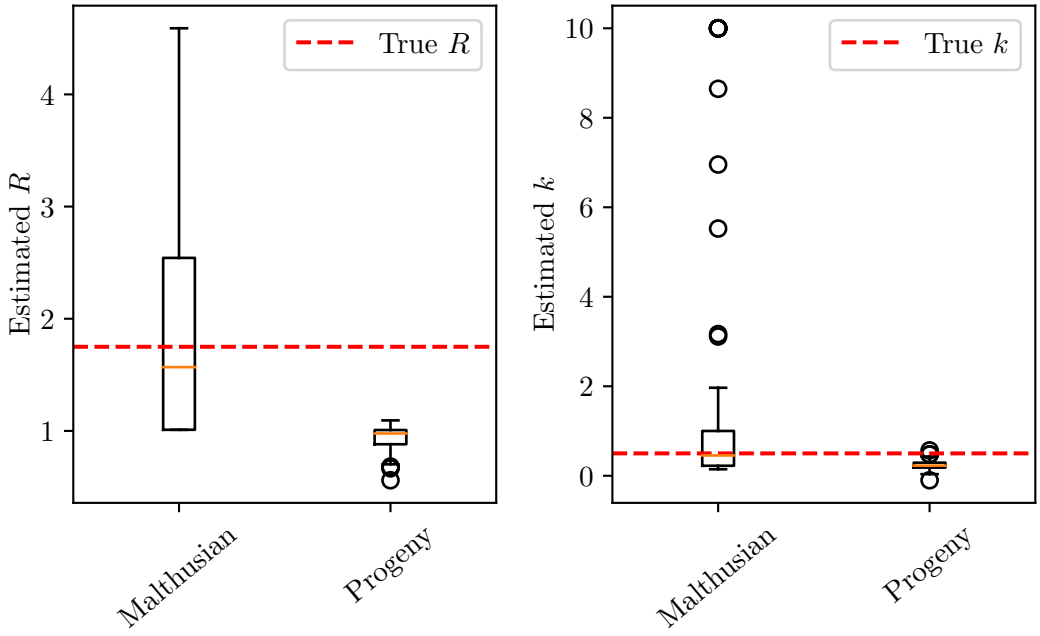


Figure 5.1: Box plots of simultaneously estimated values for $R$ and $k$ from $N = 100$ simulations. In each simulation, an infinite alleles single-type Bienaymé-Galton-Watson process with offspring distribution $X \sim \text{NEGBIN}(k, q)$ is simulated up to the tenth generation ($t = 10$), where $k = 0.5$ and $q = \frac{k}{R+k}$, such that $R = \mathbb{E}[X] = 1.75$. The probability of mutation $\nu = 0.5$ is chosen such that the distribution in (2.4) is proper. The sample is used for the maximum likelihood estimation if the size of the branching at time $t = 10$, $y_{10}$, satisfies $y_{10} > \frac{1-R^{11}}{2(1-R)} = \mathbb{E}[Y_{10}]$. We impose this threshold to exclude trees that have gone extinct or have a disproportional small size which does not show the behaviour due to exponential growth. Maximum likelihood estimation is performed for $(R, k)$ assuming two different models. The first model has the probability mass function given by (3.30) for $n = 1, 2, 3$, and for $n \geq 4$ we compute $1 - \mathbb{P}(C^\alpha \leq 3)$. We refer to this model as "Malthusian" on the x-axis. The second model has the probability mass function given by (2.4), and is referred to as "Progeny". For both models the log-likelihood function $LL(R, k)$ is maximized simultaneous in both arguments, using the L-BFGS-B method [6].

The Progeny model seems to be consistently estimating a reproduction value close to 1. This can be explained with Figure 1.1. In the caption it is argued that assuming all clusters have reached their final sizes, i.e. under the Progeny model, small cluster sizes are underrepresented. Clusters that eventually reach a large size, might be represented in the sample at the time of observation by a small value, further increasing the imbalance and decreasing the observed reproduction number $R$ under the Progeny model. This has a similar effect on the estimated dispersion parameter $k$, which is proportional to the variance of offspring distribution.

The following simulation study is demonstrating the potential of Theorem 6 for real world applications. The goal is estimate $\nu$, by computing the empirical mean of the observed cluster sizes of an infinite alleles single type Bienaymé-Galton-Watson process. We again assume $X \sim \text{NEGBIN}(k, q)$, such that $R = \mathbb{E}[X]$, for the offspring distribution. However, we let $R \in \{1.05, 1.75, 2.45\}$, such that we have an almost critical branching process with $R = 1.05$,

the same setting as the previous simulation study for $R = 1.75$, and a relatively high reproduction number with $R = 2.45$. For the probability of mutation we choose $\nu \in \{0.1, 0.5, 0.9\}$ to test two extreme values, $\nu = 0.1$ and $\nu = 0.9$, and the same value as in the previous study, $\nu = 0.5$. The results are shown in the box plots in Figure 5.2.

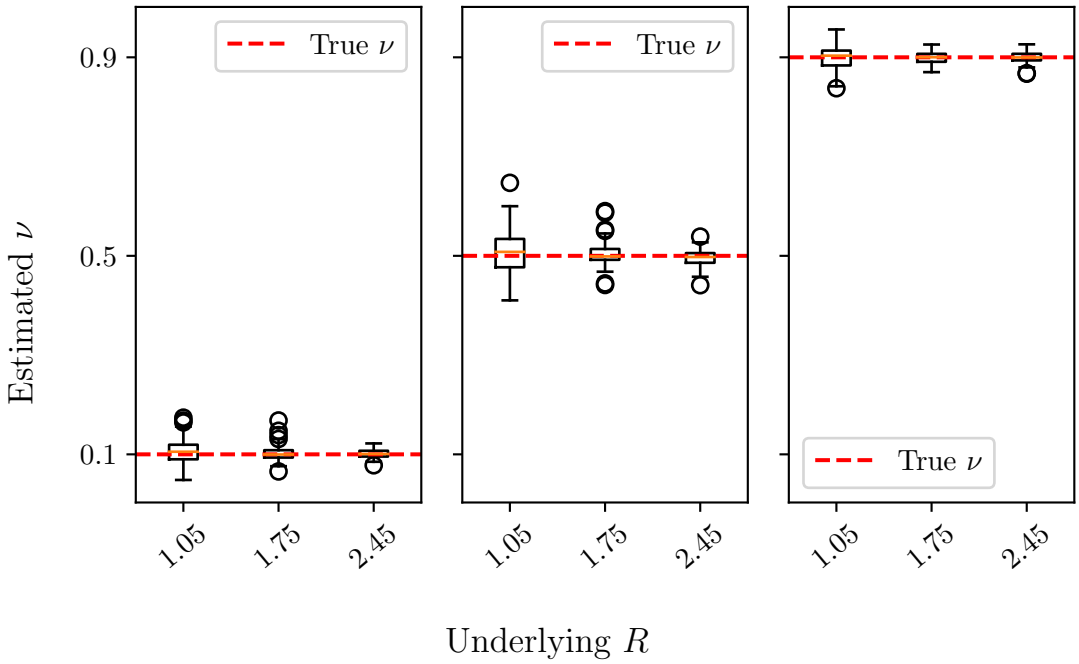


Figure 5.2: Box plots of estimates for $\nu$. In each simulation, an infinite alleles single-type Bienaymé-Galton-Watson process with offspring distribution $X \sim \text{NEGBIN}(k, q)$ is simulated up to the $t_{max}$-th generation ($t = t_{max}$), where $k = 0.5$ and $q = \frac{k}{R+k}$, such that $R = [X] \in \{1.05, 1.75, 2.45\}$. The value of $t_{max}$ is $t_{max} = 11$, $t_{max} = 9$ and $t_{max} = 7$, for $R = 1.05$, $R = 1.75$ and $R = 2.45$ respectively to reduce computation time for the branching processes with $R = 2.45$ specifically. The probability of mutation $\nu$ takes values $\nu \in \{0.1, 0.5, 0.5\}$. The sample is used for the estimation if the size of the branching at time $t = t_{max}$, $y_{t_{max}}$, satisfies $y_{t_{max}} > \max(\frac{1-R^{11}}{2(1-R)}, 100) = \mathbb{E}[Y_{t_{max}}]$. For each combination of $R$ and $\nu$, a total of $N = 100$ simulations are performed. It must be pointed out that the estimates for $\nu$ with different underlying $R$ are not directly comparable, as a sample from a branching process with $R = 2.45$ often contains hundreds of clusters, as $2.45^7 > 500$. Whereas we only have $1.05^{11} = 1.71$, for $R = 1.05$. The results show indeed that the average of the cluster sizes is not dependent on the parameters of the offspring distribution, but only depends on $\nu$.

The results show what is expected, provided that the whole tree is known up to the time of observation, the average empirically observed cluster size is approaching $\nu^{-1}$, independent from the parameters of the offspring distribution. The estimations with $R = 1.05$ exhibit a larger variance, likely due to the branching processes achieving much smaller sizes compared to those with higher underlying reproduction numbers. These deviations are not unexpected, since a consistent estimator converges as the sample size increases.

# Chapter 6

# Conclusion

Multiple models of the infinite alleles branching process were considered, all with the same goal; to count the sizes of cluster with identical labels. The classical Bienaymé-Galton-Watson process was considered in Chapter 2, for which we derived a probability mass function $\mathbb{P}(\widetilde{Y}_\infty = n)$, counting the sizes of empirically observed clusters. Here the cluster sizes are modelled as the final size of a branching process $(\widetilde{Z}_n)_{n\in\mathbb{N}}$. We argued that this model is not suitable under the assumption that the branching process is supercritical, i.e. the mean offspring distribution $X$ satisfies $\mathbb{E}[X] > 1$. This reasoning led to the next chapter.

In Chapter 3 we considered general branching processes counted by characteristics in order to derive a probability mass function $\mathbb{P}(C^\alpha = n)$. By utilizing the properties of these characteristics, we count the cluster sizes as they emerge in the exponentially growing branching process. As a result, we showed in Theorem 5, that the obtained random variable $C^\alpha$ stochastically dominates $\widetilde{Y}_\infty$. This implies that, by modelling a sample of observed cluster sizes as $\widetilde{Y}_\infty$, smaller clusters are underrepresented in the sample. Applying this model to any spreading event therefore leads to biased estimates.

The random variable $C^\alpha$ also provided a notably straightforward method for estimating the probability of mutation, $\nu$. In Theorem 6, we demonstrated that the inverse of the empirical mean of the observed cluster sizes gives a (strongly) consistent estimator for $\nu$. However, one should be careful with the interpretation of $\nu$, as mutations in genetic codes are often modelled as a continuous process, characterized by rates [32]. In [35], the probability of mutation $\nu$ is modelled as $1 - p$, where $p$ is the probability that transmission occurs before substitution[1]. The estimation of $\nu$ could therefore be of interest for the inference of transmission or substitution parameters.

We also studied the expression for the probability mass function of $C^\alpha$, in the two cases where the underlying branching process is either modelled by a Bienaymé-Galton-Watson process or an age-dependent process. This lead to Conjecture 1, which states that $\mathbb{P}(C^\alpha = n)$ coincides for all $n \in \mathbb{N}$, for the two models we mentioned. This would mean that, the empirically observed cluster sizes are not affected by the general age distribution $G$. At last, the chapter was concluded with the derivation of a probability mass function for cluster sizes which underwent downsampling. Here, it is necessary to know the probability generating function of $C^\alpha$, which is unknown, as it would partly solve the conjecture. The importance of solving the conjecture is strengthened by the applied viewpoint of this section.

In Chapter 4, the multi-type extension of the general branching process is constructed upon the fundamentals of its single-type analogue. The machinery provided by the multi-type model

[1] Also referred to as a substitution mutation, which is the event where a character in the genetic code gets replaced by another character.

is employed in two different setups, which both aim to derive analogues of Theorem 6. In the first model, it is assumed that the probability of mutation for each cluster is given by $\nu(r)$, where a type $r$ is uniformly sampled from $(0,1)$ and $\nu : (0,1) \to (0,1)$ a function. This allows the probability of mutation to vary among different individuals in the branching process. However, individuals carrying the same label, or which are of the same type in this case, have the same probability of mutation. We show that, under the condition that $\nu^{-1}$ is integrable on $(0,1)$, the mean of $C^\alpha$ is equal to $N_\nu^{-1}$. Here, $N_\nu$ is the normalization factor such that $N_\nu^{-1}\nu^{-1}$ is a probability density function with respect to the Lebesgue measure.

We also considered the case where the population is divided into a finite number of groups, and the offspring distribution might not be homogeneous among the different types. For the purpose of extending the developed theory towards inferring age-contact matrices, which was briefly discussed in the introduction in Chapter 1, it is of interest to know whether Theorem 6 is still valid. We have shown that the mean of $C^\alpha$ is still equal to $\nu^{-1}$, assuming that the branching process has Malthusian parameter $\alpha \in (0, \infty)$.

The results of Chapter 3 are applied in a simulation study in Chapter 5. Here, $C^\alpha$ and $\widetilde{Y}_\infty$, are compared as a model for the empirically observed cluster sizes. A maximum likelihood estimation is performed for the parameters $R$ and $k$, where the offspring distribution is given by $X \sim \text{NEGBIN}(k, q)$ such that $\mathbb{E}[X] = R$. We have shown that the simulation is consistent with the obtained results. That is, when $\widetilde{Y}_\infty$ is used as a model, the estimates show a significant bias. This implies that in the context of virus spread, the severity of an epidemic might be underestimated, which is undesirable from a public health perspective.

In light of the comparison of these two models, which indirectly compares to the model defined in [35], we note that the application of $C^\alpha$ as a model still requires careful consideration. For example, $\widehat{Y}_\infty$ has been applied as a model by [5] to perform exactly the same inference, a maximum likelihood estimation for $(R, k)$. However, the context which supports the research in [5], is a disease that is introduced from an external source, but is too weak to support epidemic spread, i.e. they assume $R \in (0, 1)$. It is not possible to compare $C^\alpha$ and $\widehat{Y}_\infty$ under this assumption, as most of our results do not hold.

Concluding Chapter 5, using the same simulations as mentioned before, but for varying values of $R$, we demonstrated the capabilities of Theorem 6. Under multiple extreme setups, the estimation of $\nu$ appeared to be accurate, highlighting the potential for further investigation into the inference of parameters determining the probability of mutation.

It is important to point out, that we are not the first to derive an expression of the form in (3.35). A similar observation has been made by Taïb in [34]. Whereas we look at the proportion of observed $n$-sized clusters up to some time $t$, Taïb investigates the proportion of alleles which are represented by $n$ individuals, exactly at $t$. The resulting expressions are almost identical, as we can achive Taïb's formula by replacing $\widetilde{Y}_u$ with $\widetilde{Z}_u$ in (3.35), but their interpretations are quite different. Our approach assumes that data is continuously gathered throughout time until $t$, which is a valid assumption in the context of epidemiological data. When this assumption fails, and data is only available at a single point in time, Taïb's model is more fitting. However, Taïb claims that it is not possible to be more specific about his formula, which we do not believe to be true for (3.35).

Because finding an explicit expression for $C^\alpha$ would be highly valuable, we set out to make further attempts to discover this expression. We expect that the computation of this expression will be as computationally expensive as that of (2.4), resulting in more accurate estimator without losing computational efficiency. This would be worth a publication on its own. If we do not succeed in this, a larger simulation study will be conducted. This study will involve maximum likelihood estimation using $\mathbb{P}(C^\alpha = n)$ with explicit terms up to some $n \geq 4$, with a possible application to real epidemiological data.

# References


[1] K. B. Athreya and P. E. Ney. *Branching Processes*. Dover Books on Mathematics. Dover Publications, 2004.

[2] M. S. Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):211–229, 1949.

[3] D. Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Mémoires de l'Académie Royale des Sciences à Paris*, 1760.

[4] I.J. Bienaymé. De la loi de multiplication et de la durée des familles. *Soc. Philomat. Paris Extraits, Sér*, 5(37-39):4, 1845.

[5] S. Blumberg and J. O. Lloyd-Smith. Inference of r0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS computational biology*, 9(5):e1002993, 2013.

[6] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

[7] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. M. Hyman. The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. *Journal of Theoretical Biology*, 229(1):119–126, 2004.

[8] P. L. Delamater, E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen. Complexity of the basic reproduction number (r0). *Emerging Infectious Diseases*, 25(1):1–4, 1 2019.

[9] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.

[10] K. Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2(1):23–41, 1993.

[11] R. Durrett. *Probability models for DNA sequence evolution*, volume 2. Springer, 2008.

[12] C. P. Farrington, M. N. Kanaan, and N. J. Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2):279–295, 2003.

[13] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[14] R. A. Fisher. *The genetical theory of natural selection.* Clarendon Press, 1930.

[15] C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, et al. Pandemic potential of a strain of influenza a (h1n1): early findings. *Science*, 324(5934):1557–1561, 2009.

[16] I. M. Gessel. Lagrange inversion. *Journal of Combinatorial Theory, Series A*, 144:212–249, November 2016.

[17] R. C. Griffiths and A. G. Pakes. An infinite-alleles version of the simple branching process. *Advances in Applied Probability*, 20(3):489–524, 1988.

[18] R. van der Hofstad. *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016.

[19] C. Jacob. Branching processes: their role in epidemiology. *International Journal of Environmental Research and Public Health*, 7(3):1186–1204, 2010.

[20] P. Jagers. General branching processes as markov fields. *Stochastic Processes and their Applications*, 32(2):183–212, August 1989.

[21] P. Jagers and O. Nerman. The growth and composition of branching populations. *Advances in Applied Probability*, 16(2):221–259, 1984.

[22] P. Jagers and O. Nerman. The asymptotic composition of supercritical, multi-type branching populations. *Seminaire de probabilités de Strasbourg*, 30:40–54, 1996.

[23] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927.

[24] M. Kimura and J. F. Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725, 1964.

[25] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005.

[26] T. O. McDonald and M. Kimmel. A multitype infinite-allele branching process with applications to cancer evolution. *Journal of Applied Probability*, 52(3):864–876, 2015.

[27] O. Nerman. On the convergence of supercritical general (c-m-j) branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 48(3):241–252, 1979.

[28] H. Nishiura, P. Yan, C. K. Sleeman, and C. J. Mode. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *Journal of Theoretical Biology*, 294:48–55, 2012.

[29] A. G. Pakes. An infinite alleles version of the markov branching process. *Journal of the Australian Mathematical Society*, 46(1):146–169, 1989.

[30] R. Ross. The prevention of malaria. *Journal of the American Medical Association*, LVII:1715–1716, 01 1911.

[31] R. Sanjuán and P. Domingo-Calap. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*, 73(23):4433–4448, July 2016.

[32] B. Shapiro, A. Rambaut, and A. J. Drummond. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular biology and evolution*, 23(1):7–9, 2006.

[33] V. M. Shurenkov. *Ergodicheskie protsessy Markova*. Nauka, 1989.

[34] Z. Taïb. *Branching processes and neutral evolution*, volume 93. Springer Science & Business Media, 2013.

[35] C. Tran-Kiem and T. Bedford. Estimating the reproduction number and transmission heterogeneity from the size distribution of clusters of identical pathogen sequences. *medRxiv*, 2023.

[36] H. W. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.

[37] X. Wu and M. Kimmel. Modeling neutral evolution using an infinite-allele markov branching process. *International Journal of Stochastic Analysis*, 2013(1):963831, 2013.