Journal Pre-proofs

The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data

Marc Manceau, Ankit Gupta, Timothy Vaughan, Tanja Stadler

PII: S0022-5193(20)30255-1
DOI: https://doi.org/10.1016/j.jtbi.2020.110400
Reference: YJTBI 110400

To appear in: *Journal of Theoretical Biology*

Received Date: 28 November 2019
Revised Date: 7 May 2020
Accepted Date: 3 July 2020

Please cite this article as: M. Manceau, A. Gupta, T. Vaughan, T. Stadler, The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data, *Journal of Theoretical Biology* (2020), doi: https://doi.org/10.1016/j.jtbi.2020.110400

The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data

Marc Manceau*, Ankit Gupta*, Timothy Vaughan*, Tanja Stadler**

Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland.

Abstract

We consider a homogeneous birth-death process with three different sampling schemes. First, individuals can be sampled through time and included in a reconstructed phylogenetic tree. Second, they can be sampled through time and only recorded as a point 'occurrence' along a timeline. Third, extant individuals can be sampled and included in the reconstructed phylogenetic tree with a fixed probability. We further consider that sampled individuals can be removed or not from the process, upon sampling, with fixed probability.

We derive the probability distribution of the population size at any time in the past conditional on the joint observation of a reconstructed phylogenetic tree and a record of occurrences not included in the tree. We also provide an algorithm to simulate ancestral population size trajectories given the observation of a reconstructed phylogenetic tree and occurrences.

This distribution can be readily used to draw inferences about the ancestral population size in the field of epidemiology and macroevolution. In epidemiology, these results will allow data from epidemiological case count studies to be used in conjunction with molecular sequencing data (yielding reconstructed phylogenetic trees) to coherently estimate prevalence through time. In macroevolution, it will foster the joint examination of the fossil record and extant taxa to reconstruct past biodiversity.

Keywords: birth-death process, fossilized birth-death model, epidemiology, macroevolution, phylogenetics

1. Introduction

Owing to seminal papers by Yule (1925), Kendall (1948), and much later by Nee et al. (1994), birth-death models have become ubiquitous in evolutionary biology. They are used as a population dynamic model, parameterized via a birth and death rate, in studies spanning fields as diverse as paleontology,

*Corresponding author, marc.manceau@bsse.ethz.ch
**Corresponding author, tanja.stadler@bsse.ethz.ch

macroevolution, linguistics, and epidemiology (see e.g. Foote (2000); Heath et al. (2014); Gray et al. (2009); Stadler et al. (2013)). A major aim when using these models is to reliably estimate the ancestral number of species, languages or infected individuals, i.e. past biodiversity, past prevalence, or more general past population sizes. In both macroevolution and epidemiology, population dynamics inferences can rely on occurrence data, i.e. the fossil record and the case counts record. This data is modeled as a sampling of individuals from the full population through time (Foote, 2000; Starrfelt and Liow, 2016).

In recent years, impressive sequencing efforts targeting present-day species and pathogens have enabled the reconstruction of phylogenies. Two main modeling approaches allow to quantify past population sizes in the past using these trees. First, phylodynamics tools have been developed to fit the birth and death rates of a birth-death process on the reconstructed phylogenetic tree of interest, while integrating over past population sizes (Stadler, 2011; Morlon et al., 2011). In order to quantify past population sizes, typically the expected population sizes based on these estimated birth and death rates are calculated (Morlon et al., 2011; Ratmann et al., 2016; Billaud et al., 2019). Thus, such population sizes are not directly conditioned on the reconstructed phylogenetic tree. Instead, the statistical signal in the tree is only used to compute rate estimates. Second, phylodynamic tools have been developed to fit the expected population size of a coalescent model on a reconstructed phylogenetic tree. This modeling approach may appear as a better alternative, for it is directly parametrized with the population size that we wish to estimate. However, this comes at the cost of ignoring stochastic fluctuations in small populations (Morlon et al., 2010; Ratmann et al., 2016).

Statistical approaches stemming from the analysis of case count data or from the analysis of reconstructed evolutionary trees have been part of separate bodies of work for many years, historically yielding conflicts between biodiversity estimates based on the fossil record and estimates based on reconstructed phylogenies of extant taxa (Quental and Marshall, 2010 but see also Morlon et al., 2011). A first path towards merging these disparate data was introduced by the fossilized birth-death model of Stadler (2010), which considered a birth-death model with sampling and inclusion of individuals in the tree through time. This allowed taking into account infection trees reconstructed from pathogen sequences sampled throughout an epidemic (Stadler et al., 2011). In macroevolution, it paved the way to more precise phylogenetic dating using well-conserved fossil taxa which could be placed on a reconstructed phylogeny using morphological characters (Gavryushkina et al., 2016). Not so well-conserved fossils (i.e. occurrences) have also been used with this model, using a Markov Chain Monte Carlo (MCMC) scheme to integrate over all possible placements along a fixed tree (Heath et al., 2014). Analytical developments around this new model have been made by Gupta et al. (2019), which derived an analytical formula for the probability density of an outcome of the process, which consists of a reconstructed phylogenetic tree along with a record of occurrences. Again, all these methods do not quantify population sizes directly, but estimate birth and death rates while analytically

integrating over population sizes.

Very recently, Vaughan et al. (2019) introduced a Monte-Carlo particle filtering algorithm allowing direct quantification of past population sizes and birth- and death rates conditioned on reconstructed phylogenetic trees and occurrences (see Andrieu et al., 2010 for details about particle filtering methods). As such, it can produce more accurate population size estimates than the methods mentioned above as the estimates directly condition on all data, i.e. the occurrence record (e.g. poorly preserved fossils, or case count epidemiological record) and the reconstructed phylogenetic tree.

In this paper, we build on the analytical developments presented by Gupta et al. (2019), to calculate the past population size distribution as originally targeted by Vaughan et al. (2019). Our approach here is more analytic, leading to much faster numerical calculations compared to the particle filtering method previously developed. The efficiency of our method paves the way towards considering much bigger datasets, and towards extending the method to multi-type or density-dependent birth-death processes.

In Section 2, we present the model, notation, and an overview of the strategy to express the targeted distribution. In Section 3, we adapt the main results of Gupta et al. (2019) to compute the probability density of observations made after a given time, conditioned on the past population size. In Section 4, we provide a way to compute the joint density of the past population size and observations made before a given time. Combining results of Sections 3 and 4 in Section 5, we compute the distribution of past population sizes conditional on the full outcome of the process,and perform sanity checks against previously published methods achieving similar tasks (Stadler, 2010; Vaughan et al., 2019; Gupta et al., 2019). We finally discuss applications and potential extensions of the model.

2. Model and notation

2.1. Parameters of the process

We consider a population of individuals, any of which can give birth to another individual at rate λ or die at rate μ. The process starts at time t_{or} in the past with one individual, and evolves until reaching present time 0, i.e. time is oriented from the present towards the past. In the rest of the manuscript, something *happening at time t* will thus always refer to an event taking place *t units before present*.

We superimpose to this background population dynamics three different sampling schemes. First, individuals can be ψ-*sampled* at rate ψ throughout their lifetime. When ψ-sampled, the individual will be included in the reconstructed phylogenetic tree. Second, individuals can be ω-*sampled* at rate ω throughout their lifetime. When ω-sampled, the individual is not included in the reconstructed phylogenetic tree, but its sampling time is nevertheless recorded and called 'an occurrence'. Last, the process finishes upon reaching

the present time 0, and each extant individual at that time is *ρ-sampled* with fixed probability ρ, leading to their inclusion in the reconstructed phylogenetic tree. The sum of all per-capita rates will be called for short $\gamma = \lambda + \mu + \psi + \omega$.

Following Vaughan et al. (2019), we also include in the model an effect of the ψ- and ω-sampling through time on the population dynamics. We consider that, upon sampling, an individual is either removed from the process with probability $r \in (0, 1)$, or is unaffected by the sampling with probability $(1-r)$. The overall number of individuals, denoted (I_t), thus follows a linear birth-death process with birth rate λ and death rate $\mu + (\psi + \omega)r$. Note that, because the ρ-sampling step occurs here at the end of the process, it does not matter whether or not individuals are removed upon ρ-sampling.

2.2. Introducing useful probabilities

Some aspects of this process have been previously investigated thoroughly. We now use two key probabilities. First, we will call u_t the probability that a process starting at time t with only one individual remains unsampled up to and including the present time (time 0).We recall that u_t satisfies the ordinary differential equation (ODE) (Maddison et al., 2007)

$$\begin{aligned} u_0 &= z \\ \dot{u}_t &= \lambda u_t^2 - \gamma u_t + \mu \quad . \end{aligned} \tag{2.1}$$

The solution of this for a particular initial condition z being the following

$$u(t, z) = \frac{x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}}{(x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}} \tag{2.2}$$

where $\Delta = \gamma^2 - 4\lambda\mu > (\lambda + \mu)^2 - 4\lambda\mu \geq (\lambda - \mu)^2 > 0$ and x_1, x_2 are the two roots of the polynomial $\lambda x^2 - \gamma x + \mu$,

$$x_1 = \frac{\gamma - \sqrt{\Delta}}{2\lambda} \quad \text{and} \quad x_2 = \frac{\gamma + \sqrt{\Delta}}{2\lambda} \quad .$$

Second, we call p_t the probability that a process starting at time t with one individual precisely leads to one sampled individual at present time 0. Writing the ODE governing the evolution of this quantity leads to

$$\begin{aligned} p_0 &= 1 - z \\ \dot{p}_t &= (2\lambda u(t, z) - \gamma)p_t \quad . \end{aligned} \tag{2.3}$$

The solution of this being the following

$$p(t, z) = (1 - z)\frac{\Delta}{\lambda^2}\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-2} e^{-\sqrt{\Delta}t} \quad . \tag{2.4}$$

These formulas are well known, and correspond respectively to quantities called $p_0(t)$ and $p_1(t)$ in Stadler (2010). When $z = 1 - \rho$, we will drop the dependence on z and use the shorter notation u_t, p_t. We recall standard ways to derive these expressions in AppendixA.

2.3. Strategy of the paper

The process with sampling leads to the observation of two distinct objects $(\mathcal{T}, \mathcal{O})$ illustrated in Figure 1. The reconstructed phylogenetic tree $\mathcal{T}$, on the one hand, represents the evolutionary relationships between all ψ-sampled and ρ-sampled individuals. We further consider that ψ-sampled individuals are labeled either as 'removed' or 'non-removed'. All ψ-sampled removed individuals are necessarily leaves of $\mathcal{T}$, whereas ψ-sampled non-removed ones can either stand as leaves (when the descent of the individual is not sampled) or as vertices along a branch (when the descent of the individual is further sampled), in which case they are referred to as *sampled ancestors*.
The record of occurrences $\mathcal{O}$, on the other hand, is an ordered list of all ω-sampling times. We also consider that these sampling times are labeled as either 'removed' or 'non-removed'.

In this paper, we are interested in computing the probability distribution of the number of individuals in the past, conditioned on the observed outcome $(\mathcal{T}, \mathcal{O})$ of the process. If k_t denotes the number of sampled lineages in $\mathcal{T}$ at time t, we call our target distribution,

$$\forall t \geq 0, \forall i \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}, \quad K_t^{(i)} := \mathbb{P}(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}) \quad . \tag{2.5}$$

We will refer to *epochs* as the maximal time slices within which no sampling event in $\mathcal{O}$, nor branching event in $\mathcal{T}$, happened. These epochs are delimited by the union of sampling times in $\mathcal{O}$, branching times in the tree $\mathcal{T}$, and sampling times of leaves and sampled ancestors in $\mathcal{T}$. All pooled together, we call these ordered times $(t_h)_{h=0}^n$, starting at present time $t_0 = 0$ and ending at the origin time $t_n = t_{or}$.

At any time $t \geq 0$ we also introduce:

$$\begin{aligned}
\mathcal{T}_t^{\uparrow} &:= \text{the tree } \mathcal{T} \text{starting at the origin time } t_{or} \text{ and cut at time } t \\
\mathcal{T}_t^{\downarrow} &:= \text{the collection of trees (or forest) obtained by cutting } \mathcal{T} \\
&\qquad \text{at time } t, \text{ and considering all subtrees descending from cut lineages} \\
\mathcal{O}_t^{\uparrow} &:= \mathcal{O}_{|(t, t_{or})} \\
\mathcal{O}_t^{\downarrow} &:= \mathcal{O}_{|(0,t)}
\end{aligned}$$

The general strategy – and outline – of the paper is the following. We will traverse the tree and record of occurrences *breadth-first*, i.e. level-by-level through time. In a *backward traversal* we will compute the

Figure 1: General setting of the method. a) the full process with sampling. Pink dots translate as dots in $\mathcal{O}$ and correspond to ω-sampling (sampling through time without sequencing). Blue dots translate as dots in $\mathcal{T}$ and correspond to ψ-sampling (sampling through time with sequencing). Yellow dots correspond to all present-day ρ-sampling events. Filled or unfilled dots correspond respectively to sampling with or without removal. b) Population size through time. c) Observed occurrences through time. d) Reconstructed phylogenetic tree. e) Number of individuals in reconstructed phylogenetic tree through time.

probability density of observations made between time t and 0 conditioned on the population size at time t. We call this probability density,

$$\forall i \in \mathbb{N}_0, \quad L_t^{(i)} := \mathbb{P}\left(\mathcal{T}_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i\right) \quad . \tag{2.6}$$

In a *forward traversal* we will then compute the joint probability density of the observations made prior to time t and the population size at time t. We call this density,

$$\forall i \in \mathbb{N}_0, \quad M_t^{(i)} := \mathbb{P}\left(\mathcal{T}_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i\right) \quad . \tag{2.7}$$

Provided we get expressions of $(L_t)_{t=0}^{t_{or}}$ and $(M_t)_{t=0}^{t_{or}}$, our target distribution can then be expressed by combining both, noting that

$$\begin{aligned} K_t^{(i)} &:= \mathbb{P}\left(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}\right) \\ &\propto \mathbb{P}\left(I_t = k_t + i, \mathcal{T}_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, \mathcal{T}_t^{\downarrow}, \mathcal{O}_t^{\downarrow}\right) \\ &= \mathbb{P}\left(\mathcal{T}_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, \mathcal{T}_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \ \mathbb{P}\left(I_t = k_t + i, \mathcal{T}_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &= L_t^{(i)} M_t^{(i)} \end{aligned} \tag{2.8}$$

where the last line holds because, conditionally on $I_t = k_t + i$, the future of the (Markov) process is independent of what happened before.

In the process of getting the probability density of $\mathcal{T}, \mathcal{O}$ under the same model, Gupta et al. (2019) provided an analytical formula and an algorithm to compute the first ingredient L_t in the case where all individuals are removed upon sampling (i.e. $r = 1$). We thus recall their main result, and adapt it to our slightly different framework, in the next section.

3. Calculation of L_t – The density of observations below t conditioned on past population size

We start this section by presenting the ODEs satisfied by the probability density L_t. This provides us with a numerical algorithm to compute L_t, which we subsequently simplify with analytical results for specific sets of parameters.

3.1. Set of ODEs satisfied by L_t

We can derive the probability density L_t by studying its evolution through time. First, observe that we can express L_0 at present time 0. Indeed, provided we know the exact number of individuals living at time 0, the probability to see the tips of the tree is directly driven by the ρ-sampling,

$$\forall i \in \mathbb{N}_0,\ L_0^{(i)} = \rho^{k_0}(1-\rho)^i \quad . \tag{3.9}$$

We now derive the ODE driving the evolution of L_t through time across any given epoch. We consider an infinitesimal time step δt and list the events which could have happened in the full process between $t+\delta t$ and t, leading to our observations. Suppose the number of observed lineages in this epoch is k, and the total number of individuals alive is $k+i$. We emphasize three cases, illustrated in Figure 2:

1. nothing happened with probability $(1-\gamma(k+i)\delta t)$
2. a birth event happened
   (a) among the k sampled lineages in $\mathcal{T}_t^{\downarrow}$, and it leads to an extinct or unsampled subtree to the left or to the right, with probability $2\lambda k\delta t$.
   (b) among the i other individuals, with probability $\lambda i\delta t$.
3. a death event happened among the i particles, with probability $\mu i\delta t$.

Figure 2: Four unobservable scenarios taken into account to derive the ODEs 3.10 and 4.24.

These allow us to write, $\forall i \in \mathbb{N}_0$,

$$L_{t+\delta t}^{(i)} = (1-\gamma(k+i)\delta t)\, L_t^{(i)} + \lambda(2k+i)\delta t L_t^{(i+1)} + \mu i \delta t L_t^{(i-1)} \quad .$$

Note that for $i=0$, $L_t^{(i-1)}$ is not defined, but the term cancels out thanks to the factor i.

Subtracting $L_t^{(i)}$ from both sides, dividing by δt and letting $\delta t \to 0$, we get the following set of ODEs driving the evolution of L_t,

$$\begin{aligned} \forall i \in \mathbb{N}_0,\ \ & L_0^{(i)} = \rho^{k_0}(1-\rho)^i \\ & \dot{L}_t^{(i)} = -\gamma(k+i)L_t^{(i)} + \lambda(2k+i)L_t^{(i+1)} + \mu i L_t^{(i-1)} \quad . \end{aligned} \tag{3.10}$$

Last, we need to study how L_t changes at punctual events. We call *unsampled lineages* the lineages that do not appear on the reconstructed phylogenetic tree, i.e. have not been ρ- or ψ-sampled. Note that these unsampled lineages might still be subject to ω-sampling events.

There are 6 types of punctual events that we can come across at time t in the past, listed below and

illustrated in Figure 3. We denote L_{t^+} the probability just before (i.e. up) the punctual event and L_{t^-} the probability immediately after (i.e. down). One directly gets L_{t^+} by decomposing it into what must occur below t^-, multiplied by the rate of the specific event happening on the infinitesimal time window (t^-, t^+). We can either find,

1. a leaf of $\mathcal{T}_t^{\downarrow}$, labeled as removed. This is a ψ-sampling with removal event for which the number of unsampled lineages remains constant, and the number of sampled lineages increases by one (going backward in time). It thus gives,

$$L_{t^+}^{(i)} = \psi r L_{t^-}^{(i)} \quad . \tag{3.11}$$

2. a leaf of $\mathcal{T}_t^{\downarrow}$, labeled as non-removed. This is a ψ-sampling without removal event for which one of the unsampled lineage becomes a sampled one (going backward in time). It thus gives,

$$L_{t^+}^{(i)} = \psi(1-r) L_{t^-}^{(i+1)} \quad . \tag{3.12}$$

3. a sampled ancestor along a branch of $\mathcal{T}_t^{\downarrow}$, necessarily labeled as non-removed. This is a ψ-sampling without removal event, not impacting the number of sampled or unsampled lineages. It thus gives,

$$L_{t^+}^{(i)} = \psi(1-r) L_{t^-}^{(i)} \quad . \tag{3.13}$$

4. an occurrence in $\mathcal{O}_t^{\downarrow}$, labeled as removed. This is a ω-sampling with removal event, for which the number of unsampled lineages increases by one (going backward in time). It thus gives,

$$L_{t^+}^{(i)} = \omega r i L_{t^-}^{(i-1)} \quad . \tag{3.14}$$

   Note that here also, for $i = 0$, $L_t^{(-1)}$ is not defined but the term cancels out thanks to the factor i.

5. an occurrence in $\mathcal{O}_t^{\downarrow}$, labeled as non-removed. This is a ω-sampling without removal event, not impacting the number of sampled or unsampled lineages. It thus gives,

$$L_{t^+}^{(i)} = \omega(k+i)(1-r) L_{t^-}^{(i)} \quad . \tag{3.15}$$

6. a branching event between two branches of $\mathcal{T}_t^{\downarrow}$. The number of sampled lineages decreases by one (going backward in time). It thus gives,

$$L_{t^+}^{(i)} = \lambda L_{t^-}^{(i)} \quad . \tag{3.16}$$

Note that these updates can be adapted to the case when we don't observe the removal status of individuals. The update corresponding to a leaf of $\mathcal{T}$ is the sum of updates (3.11) and (3.12), the update corresponding to an occurrence event is the the sum of updates (3.14) and (3.15), while updates (3.13) and (3.16) are unchanged.

Figure 3: Six observable punctual events in the data.

This set of ODEs (3.10) together with update equations (3.11)-(3.16 can be numerically approximated. To do so, we fix a finite upper bound N on the number of hidden individuals and numerically integrate a truncated ODE system. We detail this in the following algorithm to compute an approximation of L_t at any time t.

Algorithm 1 Computes a numerical approximation of L_t for a specific set of times

Input:

Observed tree and occurrence data $(\mathcal{T}, \mathcal{O})$,
parameters $(t_{or}, \lambda, \mu, \psi, \omega, \rho, r)$,
set of time points $(\tau_j)_{j=1}^{S}$ for which we want to compute the density $L_{\tau_j}^{(i)}$,
and the truncation N setting the accuracy of the algorithm.

Output: A numerical approximation of L_t at times $(\tau_j)_{j=1}^{S}$, $(\widetilde{L}_{\tau_j}^{(i)})_{\substack{i\in\{0,1,\dots,N\}\\ j\in\{1,2,\dots,S\}}}$.

Pool all (τ_j) and all branching and sampling times of $(\mathcal{T}, \mathcal{O})$ in an ordered list $(t_h)_{h=1}^{n}$
Set $j = 1$ and initialize B as a $S \times (N+1)$ empty matrix
Set $\forall i \in \{0, 1, \dots, N\},\ \widetilde{L}_0^{(i)} = \rho^{k_0}(1-\rho)^i$
for $h = 1, 2, \dots, n$ **do**
    Numerically solve the ODE $\dot{\widetilde{L}}_t = A\widetilde{L}_t$ on (t_{h-1}, t_h), by computing $\widetilde{L}_{t_h} = e^{(t_h - t_{h-1})A}\widetilde{L}_{t_{h-1}}$,
where matrix A is a $(N+1) \times (N+1)$ tridiagonal matrix with entries given by,

$$\forall i \in \{0, 1, \dots, N\} \quad A^{(i,i)} = -\gamma(k+i)$$
$$\forall i \in \{0, 1, \dots, N-1\} \quad A^{(i,i+1)} = \lambda(2k+i)$$
$$\forall i \in \{1, 2, \dots, N\} \quad A^{(i,i-1)} = \mu i$$

    if $t_h = \tau_j$ **then**
        Record $\forall i,\ B^{(j,i)} = \widetilde{L}_{t_h}^{(i)}$
        Set $j = j + 1$
    end if
    if $t_h = t_n$ or $t_h = \tau_S$ **then**
        return B
    else if t_h is a removed leaf **then**
        Set $\widetilde{L}_{t_h^+} = \psi r \widetilde{L}_{t_h^-}$
    else if t_h is a non-removed leaf **then**
        Set $\forall i < N,\ \widetilde{L}_{t_h^+}^{(i)} = \psi(1-r)\widetilde{L}_{t_h^-}^{(i+1)}$ and $\widetilde{L}_{t_h^+}^{(N)} = 0$
    else if t_h is a sampled ancestor **then**
        Set $\widetilde{L}_{t_h^+} = \psi(1-r)\widetilde{L}_{t_h^-}$
    else if t_h is a removed occurrence **then**
        Set $\forall i > 0, \widetilde{L}_{t_h^+}^{(i)} = \omega r i \widetilde{L}_{t_h^-}^{(i-1)}$ and $\widetilde{L}_{t_h^-}^{(0)} = 0$
    else if t_h is a non-removed occurrence **then**
        Set $\widetilde{L}_{t_h^+}^{(i)} = \omega(1-r)(k+i)\widetilde{L}_{t_h^-}^{(i)}$
    else t_h is a branching event
        Set $\widetilde{L}_{t_h^+} = \lambda \widetilde{L}_{t_h^-}$
    end if
end for

We also define a slight variation of this algorithm, that we will refer to as Algorithm 1', where no set of time points (τ_j) is required, and the values of $\widetilde{L}_t$ are not recorded through time (i.e. matrix B disappears). Instead, when reaching $t_n = t_{or}$ we simply return $\widetilde{L}_t^{(0)}$, which by definition is an estimate of the probability density of $(\mathcal{T}, \mathcal{O})$. Note that this strategy is identical to what has been used to compute the probability density of a reconstructed phylogenetic tree under a logistic birth-death process (Leventhal et al., 2013).

These two algorithms will prove useful to deal with the general case. Furthermore, we may obtain analytical expressions for L_t when $\omega = 0$ as well as when $r = 1$ (Gupta et al., 2019). We reveal these in the next two subsections.

3.2. Special case $\omega = 0$

Suppose we can express $L_t^{(i)}$ as the product $L_t^{(i)} = u_t^i W_t$ where W_t is a function of time only, and u_t is defined as in equation 2.2. We first get, from the initialization in equation (3.10), that $W_0 = \rho^{k_0}$. Moreover, substituting $u_t^i W_t$ in the ODE leads to

$$\begin{aligned}\dot{L}_t^{(i)} &= i u_t^{i-1} \dot{u}_t W_t + u_t^i \dot{W}_t \\ &= \left(\lambda i u_t^{i+1} - \gamma i u_t^i + \mu i u_t^{i-1}\right) W_t + u_t^i \dot{W}_t \quad .\end{aligned}$$

Thus leading to the following ODE for W_t, on any epoch (t_h, t_{h+1}) where the number of sampled lineages remains fixed and equal to k,

$$\begin{aligned} & u_t^i \dot{W}_t = \left(-\gamma(i+k)u_t^i + \lambda(2k+i)u_t^{i+1} + \mu i u_t^{i-1} - \lambda i u_t^{i+1} + \gamma i u_t^i - \mu i u_t^{i-1}\right) W_t \\ \Rightarrow \quad & \dot{W}_t = (2\lambda u_t - \gamma) k W_t \quad .\end{aligned}$$

This is very close to the ODE (2.3) governing the evolution of p_t, and it leads to (see derivation in AppendixA),

$$\forall t \in (t_h, t_{h+1}), \quad W_t = W_{t_h} \left(\frac{p_t}{p_{t_h}}\right)^k \quad . \tag{3.17}$$

Last, because $\omega = 0$, updates (3.11) to (3.16) simplify to only the following ψ- and λ-events,

$$\text{if } t \text{ is a removed leaf}, \quad W_{t^+} = \psi r W_{t^-} \tag{3.18}$$

$$\text{if } t \text{ is a non-removed leaf}, \quad W_{t^+} = \psi(1-r) u_t W_{t^-} \tag{3.19}$$

$$\text{if } t \text{ is a sampled ancestor}, \quad W_{t^+} = \psi(1-r) W_{t^-} \tag{3.20}$$

$$\text{if } t \text{ is a branching time}, \quad W_{t^+} = \lambda W_{t^-} \quad . \tag{3.21}$$

Combining these updates with equation (3.17) leads to the following proposition.

Proposition 3.1. *When $\omega = 0$, at any time t across epoch (t_h, t_{h+1}), considering that we observed so far –i.e. on $(0, t_{h+1})$ – v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, x branching events at times $t_j \in \mathcal{X}$, y non-removed leaves at times $t_j \in \mathcal{Y}$, we get,*

$$L_t^{(i)} = u_t^i W_t$$

$$\textit{where} \quad W_t = \lambda^x \psi^{v+w+y} (1-r)^{v+y} r^w p_t^{k_t} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1}.$$

Proof We prove this proposition by induction across the epochs in AppendixE, using as the main arguments the equation updates (3.18) to (3.21), combined with equation (3.17). □

Note that this proposition is very similar to what is presented in Section 3 by Gupta et al. (2019). We nevertheless need to highlight two differences.

The first one is that we allow here for removal or not of the individual upon sampling, with a given probability r, whereas Gupta et al. (2019) considered that all individuals were removed upon sampling ($r = 1$), and Stadler (2010) considered that individuals were not removed upon sampling ($r = 0$).

The second difference concerns the underlying framework under which we derive our results. In Gupta et al. (2019), individuals where distinguishable (say, each one is assigned a number and they can be ordered), whereas in the present paper they are not. When individuals are ordered, the probability density $L_t^{(i)}$ is changed by a factor $\frac{(k+i)!}{i!}$, which is the number of ways we can arrange $k+i$ elements in a list of size k, i.e. the number of ordered configurations of hidden individuals.

Note that, when reaching the origin of the tree, the formula in Proposition 3.1 reduces to a very similar formula for the probability density of $\mathcal{T}$ because $i = 0$ and $k = 1$. We summarize this as the following corollary.

Corollary 3.2. *When $\omega = 0$, the probability density of a reconstructed tree $\mathcal{T}$ with v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, y non-removed leaves at times $t_j \in \mathcal{Y}$, and branching events at times $t_j \in \mathcal{X}$, is*

$$\mathbb{P}(\mathcal{T}) = \lambda^{w+y+k_0-1} \psi^{v+w+y} (1-r)^{v+y} r^w \prod_{t_j \in \mathcal{X} \cup \{t_{or}\}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \tag{3.22}$$

Proof It directly follows from Proposition 3.1, by noting that $\mathbb{P}(\mathcal{T}) = L_{t_{or}}^{(0)}$. Note also that a rooted binary tree with $w + y + k_0$ leaves shows necessarily $x = w + y + k_0 - 1$ branching times. □

Note that this formula is a straightforward generalization of formulas provided in Stadler (2010) (where $r = 0$) or Stadler et al. (2011) (where $\rho = 0$).

3.3. Special case $r = 1$

When $r = 1$, only three kinds of punctual events, corresponding to updates (3.11), (3.14) and (3.16) need to be taken into account. Because the number of unsampled individuals i goes into formula (3.14),

the simple expression $L_t^{(i)} = u_t^i W_t$ cannot be considered anymore, and one needs to find another expression. This has already been done in Gupta et al. (2019) and we only need to adapt here their result to our slightly different framework.

Proposition 3.3. *When $r = 1$, we can compute the $L_t^{(i)}$ values at any time t as*

$$L_t^{(i)} = \sum_{\ell=0}^{q} \frac{i!}{(i-\ell)!} u_t^{i-\ell} W_t^{(\ell)}.$$

where W_t is a q dimensional time-varying vector which can be computed following Algorithm 2 in Gupta et al. (2019).

Proof The proof relies on the definition of a *distinguishable version* of the probability $L_t^{(i)}$ as

$$\overline{L}_t^{(i)} = \frac{(k+i)!}{i!} L_t^{(i)} \tag{3.23}$$

which allows us to use results previously derived in Gupta et al. (2019). Details are provided in Appendix AppendixB. □

Note that when there is no ω-sampling, then $q = 0$ for all times and $W_t^{(0)}$ is the same as W_t defined in the previous section.

This ends our section on the computation of L_t. It thus remains to (i) present a way to compute M_t and (ii) combine L_t and M_t to get the target distribution K_t at any time t. We do this in turn in the next two sections.

4. Calculation of M_t – The joint density of observations above t and past population size

Recall that we are now interested in computing the joint density of observations above time t and past population size at time t, i.e. $\forall i \in \mathbb{N}_0, \quad M_t^{(i)} := \mathbb{P}(\mathcal{T}_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)$. We start by presenting the ODEs satisfied by M_t, before turning to its resolution for specific parameter sets. The approach is very similar to the one presented in the previous section to compute L_t, with the slight difference that we will need to traverse the tree forward in time instead of backward in time.

4.1. Set of ODEs satisfied by M_t

At the time of origin of the process t_{or}, we only observe one starting lineage in $\mathcal{T}_{t_{or}}^{\uparrow}$. This provides us with the following initialization condition on M,

$$M_{t_{or}}^{(i)} = \mathbb{P}(I_{t_{or}} = 1 + i) = \mathbb{1}_{i=0} \quad .$$

We then derive the ODEs driving the evolution of M_t across an epoch on which the number of observed lineages is fixed and equal to k. Suppose we know M_t, and we observe no punctual event on the infinitesimal time interval $(t-\delta t, t)$. Unobservable events have already been illustrated in Figure 2. It allows us to get

$$M^{(i)}_{t-\delta t} = (1-\gamma(i+k)\delta t)\, M^{(i)}_t + \lambda(2k+i-1)\delta t \mathbb{1}_{i>0} M^{(i-1)}_t + \mu(i+1)\delta t M^{(i+1)}_t \quad .$$

Subtracting $M^{(i)}_t$ from both sides, multiplying by -1, dividing by δt and letting $\delta t \to 0$, we get the following set of ODEs driving the evolution of M_t,

$$\begin{aligned} \forall i \in \mathbb{N}_0, \;\; & M^{(i)}_{t_{or}} = \mathbb{1}_{i=0} \\ & \dot{M}^{(i)}_t = \gamma(i+k) M^{(i)}_t - \lambda(2k+i-1)\mathbb{1}_{i>0} M^{(i-1)}_t - \mu(i+1) M^{(i+1)}_t \quad . \end{aligned} \tag{4.24}$$

Last, we need to take into account the evolution of M_t at punctual events. Again, there are 6 types of punctual events that we can come across at time t in the past, listed below and illustrated in Figure 3. We denote M_{t^-} the probability just after (i.e. below) the punctual event and M_{t^+} the probability immediately before (i.e. up). Because we are here deriving M_t forward in time, one needs to carefully note differences with results derived in Section 3 relating to the number of lineages before and after the event. We can indeed find the same punctual events, namely,

1. a leaf of $\mathcal{T}^{\downarrow}_t$, labeled as removed. This is a ψ-sampling with removal event for which the number of sampled lineages decreases by one and the number of unsampled lineages remains unchanged. This gives,

$$M^{(i)}_{t^-} = \psi r M^{(i)}_{t^+} \quad . \tag{4.25}$$

2. a leaf of $\mathcal{T}^{\downarrow}_t$, labeled as non-removed. This is a ψ-sampling without removal event for which one sampled lineages becomes unsampled. This gives,

$$M^{(i)}_{t^-} = \psi(1-r)\mathbb{1}_{i>0} M^{(i-1)}_{t^+} \quad . \tag{4.26}$$

3. a sampled ancestor along a branch of $\mathcal{T}^{\downarrow}_t$, necessarily labeled as non-removed. This is a ψ-sampling without removal event which does not affect the number of lineages. It gives,

$$M^{(i)}_{t^-} = \psi(1-r) M^{(i)}_{t^+} \quad . \tag{4.27}$$

4. an occurrence in $\mathcal{O}^{\downarrow}_t$, labeled as removed. This is a ω-sampling with removal event, for which the number of unsampled lineages decreases by one. This gives,

$$M^{(i)}_{t^-} = \omega r (i+1) M^{(i+1)}_{t^+} \quad . \tag{4.28}$$

5. an occurrence in $\mathcal{O}_t^{\downarrow}$, labeled as non-removed. This is a ω-sampling without removal event which does not affect the number of lineages. It gives,

$$M_{t^-}^{(i)} = (k+i)\omega(1-r)M_{t^+}^{(i)} \quad . \tag{4.29}$$

6. a branching event between two branches of $\mathcal{T}_t^{\downarrow}$.
   This is a λ-event increasing the number of sampled lineages by one. This gives,

$$M_{t^-}^{(i)} = \lambda M_{t^+}^{(i)} \quad . \tag{4.30}$$

Finally, upon reaching present time 0, one needs to take into account the ρ-sampling, leading to the following update,

$$M_{0^-}^{(i)} = (1-\rho)^i \rho^{k_0} M_{0^+}^{(i)} \quad . \tag{4.31}$$

Note, as for L_t, that these updates can be adapted to the case when we don't observe the removal status of individuals. The update corresponding to a leaf of $\mathcal{T}$ is the sum of updates (4.25) and (4.26), the update corresponding to an occurrence event is the the sum of updates (4.28) and (4.29), while updates (4.27) and (4.30) are unchanged.

As already exhibited for L_t, we can build a similar algorithm to compute M_t in the general case, relying on a numerical ODE solver for approximating equation (4.24). As for Algorithm 1' previously introduced to compute the probability density of $(\mathcal{T}, \mathcal{O})$, a slight variation of this algorithm would allow one to compute an estimate of the probability density of $(\mathcal{T}, \mathcal{O})$ by summing the $M_0^{(i)}$'s over all i. Note that this strategy is identical to what has been used to compute the probability density of a reconstructed phylogenetic tree under a logistic birth-death process (Etienne et al., 2012; Laudanno et al., 2020).

While this approach is in theory a good approximation, it requires fixing arbitrarilly a truncation parameter N, and exponentiating matrices of dimension $N \times N$, leading to potential speed or accurracy issues. In the remainder of this section, we derive analytical results to avoid resorting to a numerical ODE solver in specific cases.

4.2. The corresponding generating function

We introduce now the generating function corresponding to the density M_t, which will prove useful to get analytical results,

$$\widehat{M}(t,z) := \sum_{i=0}^{\infty} z^i M_t^{(i)} \quad .$$

The initial condition on M translates into, $\forall z, \widehat{M}(t_{or}, z) = 1$. The ODE (4.24) furthermore translates into the following partial differential equation (PDE),

$$\begin{aligned}
\partial_t \widehat{M} &= \sum_{i=0}^{\infty} z^i \left(\gamma(i+k) M_t^{(i)} - \lambda(2k+i-1) \mathbb{1}_{i>0} M_t^{(i-1)} - \mu(i+1) M_t^{(i+1)} \right) \\
&= \gamma k \sum_{i=0}^{\infty} z^i M_t^{(i)} + \gamma \sum_{i=1}^{\infty} i z^i M_t^{(i)} - \lambda \sum_{i=0}^{\infty} z^{i+1}(2k+i) M_t^{(i)} - \mu \sum_{i=1}^{\infty} i z^{i-1} M_t^{(i)} \\
&= \gamma k \widehat{M} + \gamma z \partial_z \widehat{M} - 2k\lambda z \widehat{M} - \lambda z^2 \partial_z \widehat{M} - \mu \partial_z \widehat{M} \\
&= -k(2\lambda z - \gamma)\widehat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \widehat{M} \quad .
\end{aligned}$$

Our target generating function $\widehat{M}$ is thus the solution of the following PDE problem across a given epoch (t_{h-1}, t_h), on which the number of observed lineages remains constant and equal to k,

$$\begin{aligned}
&\widehat{M}(t_h, z) = F(z) \\
&\partial_t \widehat{M} + (\lambda z^2 - \gamma z + \mu)\partial_z \widehat{M} + k(2\lambda z - \gamma)\widehat{M} = 0 \quad . \qquad (4.32)
\end{aligned}$$

Solving this PDE problem allows us to obtain an analytical expression of $\widehat{M}$ for any time across an epoch, provided we know the expression of $\widehat{M}(t_h, z)$ at the end of the epoch.

Proposition 4.1. *The solution to the PDE problem (4.32) is given by*

$$\widehat{M}(t, z) = F\left(u(t_h - t, z)\right) R(t_h - t, z)^k$$

where we introduce $R(t, z) = p(t, z)/(1 - z)$ to ease the notation.

Proof We used the method of characteristics to solve this first order linear PDE, see derivations in AppendixC. □

Between epochs, one must also update $\widehat{M}$ according to punctual events taking place. Previously presented updates of M (equations (4.25) to (4.30)) translate into the following updates for $\widehat{M}$,

$$\text{if } t \text{ is a removed leaf,} \quad \widehat{M}(t^-, z) = \sum_{i=0}^{\infty} z^i \left(\psi r M_{t^+}^{(i)} \right) = \psi r \widehat{M}(t^+, z) \qquad (4.33)$$

$$\text{if } t \text{ is a non-removed leaf,} \quad \widehat{M}(t^-, z) = \sum_{i=0}^{\infty} z^i \left(\psi(1-r) \mathbb{1}_{i>0} M_{t^+}^{(i-1)} \right) = \psi(1-r) z \widehat{M}(t^+, z) \qquad (4.34)$$

$$\text{if } t \text{ is a sampled ancestor,} \quad \widehat{M}(t^-, z) = \sum_{i=0}^{\infty} z^i \left(\psi(1-r) M_{t^+}^{(i)} \right) = \psi(1-r) \widehat{M}(t^+, z) \qquad (4.35)$$

$$\text{if } t \text{ is a removed occurrence,} \quad \widehat{M}(t^-, z) = \sum_{i=0}^{\infty} z^i \left(\omega r (i+1) M_{t^+}^{(i+1)} \right) = \omega r \partial_z \widehat{M}(t^+, z) \qquad (4.36)$$

$$\text{if } t \text{ is a non-removed occurrence,} \quad \widehat{M}(t^-,z) = \sum_{i=0}^{\infty} z^i \left(\omega(1-r)(k+i)M_{t^+}^{(i)}\right)$$
$$= \omega(1-r)\left(k\widehat{M}(t^+,z) + z\partial_z\widehat{M}(t^+,z)\right) \tag{4.37}$$
$$\text{if } t \text{ is a branching event,} \quad \widehat{M}(t^-,z) = \sum_{i=0}^{\infty} z^i \left(\lambda M_{t^+}^{(i)}\right) = \lambda\widehat{M}(t^+,z) \quad . \tag{4.38}$$

If we are interested in the distribution at some point, we can thus start the formula at t_{or} with $F(z) = 1$, and then iteratively alternate between the updates at punctual events and the use of Proposition 4.1 over each epoch. When reaching present time 0, the step of ρ-sampling expressed in equation (4.31) moreover translates into,

$$\widehat{M}(0^-,z) = \sum_{i=0}^{\infty} z^i(1-\rho)^i\rho^{k_0}M_{0^+}^{(i)} = \rho^{k_0}\widehat{M}(0^+,(1-\rho)z) \quad . \tag{4.39}$$

While this procedure in theory allows us to get the analytical formula of $\widehat{M}$ at any time, updates (4.36) and (4.37) require differentiating the generating function, greatly complicating the expression of the function after a few occurrences. When $\omega = 0$, these two updates disappear and a nice recursion leads to a closed-form formula that we will detail in Proposition 4.3.

We implemented this procedure in the *SageMath* programming language able to deal with symbolic calculus. We were however not able to make it find concise expressions, and computing these successive derivatives was too time-consuming to be applicable to standard datasets in the field. Instead, when $\omega \neq 0$, we suggest another strategy for computing the $M_t^{(i)}$'s, namely approximating $\widehat{M}$ across punctual events by a polynomial of order N , $\sum_{l=0}^{N}\widetilde{M}_t^{(l)}z^l$, while still relying on Proposition 4.1 to drive the evolution of the probability generating function between events. This is a more efficient alternative to numerically solving the ODE system. We only need to derive the expression of the generating function at punctual events as given in the following proposition 4.2.

Proposition 4.2. *The derivatives in $z = 0$ of a generative function which can be expressed as*

$$\widehat{M}(t_h - t, z) := R(t_h - t, z)^k \sum_{l=0}^{N} \widetilde{M}_{t_h}^{(l)} u(t_h - t, z)^l$$

can be numerically computed using the formula

$$\left(\partial_z^i\widehat{M}(t_h - t, z)\right)_{z=0} = \left(\frac{\Delta}{\lambda^2}e^{-\sqrt{\Delta}(t_h-t)}\right)^k \sum_{\alpha=0}^{i}\sum_{l=\alpha}^{N}\widetilde{M}_{t_h}^{(l)}\binom{i}{\alpha}\frac{l!}{(l-\alpha)!}\left(\prod_{m=0}^{i-\alpha-1}(2k+l+m)\right)(x_1x_2)^{l-\alpha}$$
$$\left(-x_1 + x_2e^{-\sqrt{\Delta}(t_h-t)}\right)^{\alpha}\left(1-e^{-\sqrt{\Delta}(t_h-t)}\right)^{l+i-2\alpha}\left(x_2 - x_1e^{-\sqrt{\Delta}(t_h-t)}\right)^{-(2k+l+i-\alpha)} \quad .$$

Proof The derivation is detailed in AppendixD.1. □

This derivation is at the heart of Algorithm 2, allowing to follow the evolution of the $\widetilde{M}_t^{(i)}$'s through each epoch, as well as at times when we want to record them.

We will refer to Algorithm 2' as the slight variation of this algorithm aimed at computing the density of $(\mathcal{T}, \mathcal{O})$. No set of time points (τ_j) is required, and the values of $\widetilde{M}_t$ are not recorded through time (i.e. matrix B' disappears). Instead, when reaching $t_h = t_0$ we simply return $\sum_{i=0}^{N} \rho^{k_0}(1-\rho)^i \widetilde{M}^{(i)}$.

Algorithm 2 Computes a numerical approximation of M_t for a specific set of times

Input:

Observed tree and occurrence data $(\mathcal{T}, \mathcal{O})$,
parameters $(t_{or}, \lambda, \mu, \psi, \omega, \rho)$,
set of time points $(\tau_j)_{j=1}^{S}$ for which we want to compute the density,
and the truncation N setting the accuracy of the algorithm.

Output: A numerical approximation of M_t at times $(\tau_j)_{j=1}^{S}$, $(\widetilde{M}_{\tau_j}^{(i)})_{\substack{i\in\{0,1,\dots,N\}\\ j\in\{1,2,\dots,S\}}}$.

Pool all (τ_j) and all branching and sampling times of $(\mathcal{T}, \mathcal{O})$ in an ordered list $(t_h)_{h=1}^{n}$
Set $j = S$ and B' as a $S \times (N+1)$ empty matrix
Set $\forall i \in \{0, 1, \dots, N\},\ \widetilde{M}^{(i)} = \mathbb{1}_{i=0}$
Set $k = 1$
for $h = n-1, n-2, \dots, 0$ **do**
    Compute the values right before the punctual event,

$$\widetilde{\widetilde{M}}^{(i)} = \left(\frac{\Delta}{\lambda^2} e^{-\sqrt{\Delta}(t_h - t)}\right)^k \sum_{\alpha=0}^{i} \sum_{l=\alpha}^{N} \widetilde{M}_{t_h}^{(l)} \binom{l}{\alpha} \frac{1}{(i-\alpha)!} \left(\prod_{m=0}^{i-\alpha-1} (2k + l + m)\right)$$
$$\left(-x_1 + x_2 e^{-\sqrt{\Delta}(t_h - t)}\right)^{\alpha} (x_1 x_2)^{l-\alpha} \left(1 - e^{-\sqrt{\Delta}(t_h - t)}\right)^{l+i-2\alpha} \left(x_2 - x_1 e^{-\sqrt{\Delta}(t_h - t)}\right)^{-(2k+l+i-\alpha)}$$

    if $t_h = \tau_j$ **then**
        Record the result in B' : $\forall i, B'^{(j,i)} = \widetilde{\widetilde{M}}^{(i)}$
        Set $j = j - 1$.
    end if
    if $t_h = 0$ or $t_h = \tau_S$ **then**
        return B'
    else if t_h is a removed leaf **then**
        Update $\forall i,\ \widetilde{M}^{(i)} = \psi r \widetilde{\widetilde{M}}^{(i)}$
        Set $k = k - 1$
    else if t_h is a non-removed leaf **then**
        Update $\widetilde{M}^{(0)} = 0$ and $\forall i > 0,\ \widetilde{M}^{(i)} = \psi(1-r)\widetilde{\widetilde{M}}^{(i-1)}$
        Set $k = k - 1$
    else if t_h is a sampled ancestor **then**
        Update $\forall i,\ \widetilde{M}^{(i)} = \psi(1-r)\widetilde{\widetilde{M}}^{(i)}$
    else if t_h is a removed occurrence **then**
        Update $\forall i < N,\ \widetilde{M}^{(i)} = \omega r (i+1)\widetilde{\widetilde{M}}^{(i+1)}$ and $\widetilde{M}^{(N)} = 0$
    else if t_h is a non-removed occurrence **then**
        Update $\forall i,\ \widetilde{M}^{(i)} = \omega(1-r)(k+i)\widetilde{\widetilde{M}}^{(i)}$
    else t_h is a branching event
        Update $\forall i,\ \widetilde{M}^{(i)} = \lambda \widetilde{\widetilde{M}}^{(i)}$
        Set $k = k + 1$
    end if
end for

Note that we tried to follow an analogous generating function approach as an alternative to Algorithm

1 to compute L_t as well. This leads to another PDE problem, described in AppendixF, that will require further work to be solved.

4.3. Special case $\omega = 0$

We were not able to come with any analytical simplification, as in the previous section, for the case $r = 1$. However, for the special case $\omega = 0$, corresponding to the special case leading to the observation of $\mathcal{O} = \emptyset$, a nice recursion leads to a closed-form formula for $\widehat{M}$.

Proposition 4.3. *When $\omega = 0$, at any time t, considering that we have observed so far –i.e. on (t, t_{or})– v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, x branching events at times $t_j \in \mathcal{X}$, y non-removed leaves at times $t_j \in \mathcal{Y}$, we get,*

$$\widehat{M}(t,z) = \lambda^x \psi^{v+w+y} r^w (1-r)^{v+y} \prod_{t_j \in \mathcal{X} \cup \{t_{or}\}} R(t_j - t, z) \prod_{t_j \in \mathcal{W}} R(t_j - t, z)^{-1} \prod_{t_j \in \mathcal{Y}} u(t_j - t, z) R(t_j - t, z)^{-1} \quad .$$

Proof We prove this result by induction across the epochs of $\mathcal{T}$ in AppendixE, using as the main arguments the update equations (4.33), (4.34), (4.35), (4.38), combined with Proposition 4.1 driving the evolution across an epoch. □

As a simple corollary of this result, when $t_h = 0$ is the present, we get back the same probability density formula of $\mathcal{T}$ as provided, e.g. in theorem 3.5 in Stadler (2010) (when $r = 0$), in Section 3 in Gupta et al. (2019) (when $r = 1$), or in our previous corollary 3.2.

Indeed, Proposition 4.3 offers yet another proof of corollary 3.2 by noting that

$$\mathbb{P}(\mathcal{T}) = \sum_{i=0}^{\infty} M_{0^-}^{(i)} = \widehat{M}(0^-, 1) = \rho^{k_0} \widehat{M}(0^+, 1-\rho)$$

where the last equality follows from equation (4.39) taking into account the ρ-sampling at present. Note that this alternative proof is also presented in (Laudanno et al., 2020).

When $\omega = 0$, Proposition 4.3 also offers an alternative to Algorithm 2 for deriving M_t. Indeed, resorting to the generating function to get back the probability density, one can get the following corollary.

Corollary 4.4. *When $\omega = 0$, at any time t, considering that we have observed so far –i.e. on (t, t_{or})– v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, x branching events at times $t_j \in \mathcal{X}$, y non-removed leaves at times $t_j \in \mathcal{Y}$, we can compute $M_t^{(i)}$ using the following recursion,*

$$M_t^{(0)} = \lambda^x \psi^{v+w+y} r^w (1-r)^{v+y} \prod_{t_j \in \mathcal{X} \cup \{t_{or}\}} R(t_j - t, 0) \prod_{t_j \in \mathcal{W}} R(t_j - t, 0)^{-1} \prod_{t_j \in \mathcal{Y}} u(t_j - t, 0) R(t_j - t, 0)^{-1}$$

$$M_t^{(i)} = \frac{1}{i} \sum_{\alpha=1}^{i} M_t^{(i-\alpha)} C^{(\alpha)}$$

where we define

$$\begin{aligned}
C^{(\alpha)} &= 2 \sum_{t_j \in \mathcal{X} \cup \{t_{or}\}} a_{t_j - t}^{\alpha} - 2 \sum_{t_j \in \mathcal{W}} a_{t_j - t}^{\alpha} - \sum_{t_j \in \mathcal{Y}} (a_{t_j - t}^{\alpha} + b_{t_j - t}^{\alpha}) \\
a_t &= \left(1 - e^{-\sqrt{\Delta} t}\right) \left(x_2 - x_1 e^{-\sqrt{\Delta} t}\right)^{-1} \\
b_t &= \left(x_1 - x_2 e^{-\sqrt{\Delta} t}\right) \left(x_1 x_2 - x_2 x_1 e^{-\sqrt{\Delta} t}\right)^{-1} \quad .
\end{aligned}$$

Proof The probability density $M_t^{(i)}$ can be found back by taking

$$M_t^{(i)} = \frac{1}{i!} \left(\partial_z^i \widehat{M}(t, z)\right)_{z=0} \quad .$$

The result follows from the derivation of these derivatives in AppendixD.2. □

This special case ends the section. In the next section, we will combine results from Sections 3 and 4 and use our ability to compute L_t and M_t to compute K_t, the probability distribution of the population size given $(\mathcal{T}, \mathcal{O})$.

5. The distribution of past population size conditioned on observations

5.1. The distribution at fixed times

In Section 3, we explained how to compute L_t, the probability density of the observations below time t conditioned on the population size at time t. This relies either on Algorithm 1 in the general case, or on the more optimized Proposition 3.1 in case $\omega = 0$, or Proposition 3.3 in the case $r = 1$.

In Section 4, we explained how to compute M_t, the probability density of the observations above time t and the population size at time t. This relies either on Algorithm 2 in the general case, or on the more optimized Corollary 4.4 when $\omega = 0$.

We now combine L_t and M_t to derive the probability distribution of the population size given $(\mathcal{T}, \mathcal{O})$. Provided we have stored numerical values $(\widetilde{L}_{\tau_j}^{(i)})_{\substack{i \in \{0,1,\ldots,N\} \\ j \in \{1,2,\ldots,S\}}}$ and $(\widetilde{M}_{\tau_j}^{(i)})_{\substack{i \in \{0,1,\ldots,N\} \\ j \in \{1,2,\ldots,S\}}}$ for a set of time points $(\tau_j)_{j=1}^{S}$, recall from the first section that we obtain

$$\begin{aligned}
K_{\tau_j}^{(i)} &= \mathbb{P}\left(I_{\tau_j} = k_{\tau_j} + i \mid \mathcal{T}, \mathcal{O}\right) \\
&= \frac{L_{\tau_j}^{(i)} M_{\tau_j}^{(i)}}{\mathbb{P}(\mathcal{T}, \mathcal{O})} \\
&\approx \frac{\widetilde{L}_{\tau_j}^{(i)} \widetilde{M}_{\tau_j}^{(i)}}{\mathbb{P}(\mathcal{T}, \mathcal{O})} \quad \text{if } i \leq N, \quad \text{and 0 otherwise.}
\end{aligned}$$

Note that the denominator needs only be computed once, by evaluating $\sum_{i=0}^{N} \widetilde{L}_{\tau_j}^{(i)} \widetilde{M}_{\tau_j}^{(i)}$ for example at time $\tau_j = t_{or}$ or $\tau_j = 0$ as described in previous sections.

Depending on the parameter space that one wants to consider, it thus remains to arrange pieces stemming from the previous sections. We provide a flowchart in Figure 4 to guide the reader to chose the most efficient path.

Figure 4: The most efficient results depending on the parameter space considered. In red, results already described in Stadler (2010) and Gupta et al. (2019). In blue, the new contribution of this manuscript.

5.2. Generator of trajectories

The previous result gives us the distribution of the population size at any time in the past, but does not state anything about population size trajectories. We provide now an approximate way of simulating population size trajectories conditioned on $(\mathcal{T}, \mathcal{O})$.

Indeed, recall we have,

$$\begin{aligned}
K_t^{(i)} &:= \mathbb{P}(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}) \propto L_t^{(i)} M_t^{(i)} \\
\dot{L}_t^{(i)} &= -\gamma(k_t + i) L_t^{(i)} + \lambda(2k_t + i) L_t^{(i+1)} + \mu i L_t^{(i-1)} \\
\dot{M}_t^{(i)} &= \gamma(k_t + i) M_t^{(i)} - \mu(i+1) M_t^{(i+1)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} M_t^{(i-1)} \quad .
\end{aligned}$$

We thus get,

$$\begin{aligned}
\dot{K}_t^{(i)} &\propto \dot{L}_t^{(i)} M_t^{(i)} + L_t^{(i)} \dot{M}_t^{(i)} \\
&\propto -\gamma(k_t + i) L_t^{(i)} M_t^{(i)} + \lambda(2k_t + i) L_t^{(i+1)} M_t^{(i)} + \mu i L_t^{(i-1)} M_t^{(i)} \\
&\quad + \gamma(k_t + i) M_t^{(i)} L_t^{(i)} - \mu(i+1) M_t^{(i+1)} L_t^{(i)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} M_t^{(i-1)} L_t^{(i)} \\
&\propto \lambda(2k_t + i) \frac{L_t^{(i+1)}}{L_t^{(i)}} K_t^{(i)} + \mu i \frac{L_t^{(i-1)}}{L_t^{(i)}} K_t^{(i)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} \frac{L_t^{(i)}}{L_t^{(i-1)}} K_t^{(i-1)} - \mu(i+1) \frac{L_t^{(i)}}{L_t^{(i+1)}} K_t^{(i+1)} \\
&\propto Q_t^{(i,i)} K_t^{(i)} + Q_t^{(i-1,i)} K_t^{(i-1)} + Q_t^{(i+1,i)} K_t^{(i+1)} \quad . \qquad (5.40)
\end{aligned}$$

We introduced in the last line the following notation,

$$\begin{aligned}
Q_t^{(i+1,i)} &= -\mu(i+1) \frac{L_t^{(i)}}{L_t^{(i+1)}} \\
Q_t^{(i-1,i)} &= -\lambda(2k_t + i - 1)\mathbb{1}_{i>0} \frac{L_t^{(i)}}{L_t^{(i-1)}} \\
Q_t^{(i,i)} &= \lambda(2k_t + i) \frac{L_t^{(i+1)}}{L_t^{(i)}} + \mu i \frac{L_t^{(i-1)}}{L_t^{(i)}} \quad .
\end{aligned}$$

Using these, we see that $Q_t^{(i,i)} = -\left(Q_t^{(i,i+1)} + Q_t^{(i,i-1)}\right)$. This allows us to draw trajectories of the number of ancestors in the past as a time-continuous Markov process with the (inhomogeneous) rates Q_t written above.

Observe that we could equally write these ODE coefficients using the $M_t^{(i)}$'s. This gives,

$$\begin{aligned}\dot{K}_t^{(i)} &\propto \lambda(2k_t+i)\frac{M_t^{(i)}}{M_t^{(i+1)}}K_t^{(i+1)} + \mu i\frac{M_t^{(i)}}{M_t^{(i-1)}}K_t^{(i-1)} - \mu(i+1)\frac{M_t^{(i+1)}}{M_t^{(i)}}K_t^{(i)} - \lambda(2k_t+i-1)\mathbb{1}_{i>0}\frac{M_t^{(i-1)}}{M_t^{(i)}}K_t^{(i)} \\ &\propto R_t^{(i+1,i)}K_t^{(i+1)} + R_t^{(i-1,i)}K_t^{(i-1)} + R_t^{(i,i)}K_t^{(i)} \end{aligned} \tag{5.41}$$

where we introduced in the last line the following notation,

$$\begin{aligned} R_t^{(i+1,i)} &= \lambda(2k_t+i)\frac{M_t^{(i)}}{M_t^{(i+1)}} \\ R_t^{(i-1,i)} &= \mu i\frac{M_t^{(i)}}{M_t^{(i-1)}} \\ R_t^{(i,i)} &= -\lambda(2k_t+i-1)\mathbb{1}_{i>0}\frac{M_t^{(i-1)}}{M_t^{(i)}} - \mu(i+1)\frac{M_t^{(i+1)}}{M_t^{(i)}} \quad . \end{aligned}$$

This is a standard result for Markov chains that are conditioned on a final state, and the shape of the newly derived transition kernel is called a Doob's transform (Levin and Peres, 2017). Note that these transitions symplify for special cases when we have an analytical expression of either $L_t^{(i)}$ or $M_t^{(i)}$.

5.3. Numerical implementation

Results of this paper have been implemented numerically and the code is freely available on GitLab: *https://gitlab.com/MMarc/popsize-distribution/.*

We used the numerical implementation to verify the correctness of the results in several ways:

1. We verified that the values of the probability density of $(\mathcal{T}, \mathcal{O})$ computed using L_t and M_t (i.e. respectively using Algorithms 1' and 2') were equivalent to values computed using already known formulas when $(\omega = 0, r = 0)$ (Stadler, 2010) or when $r = 1$ (Gupta et al., 2019). See result in Figure 5AB.
2. We verified that the values of the probability density of $(\mathcal{T}, \mathcal{O})$ computed using L_t or M_t (Algorithms 1' and 2') were identical on examples for which no previous formula was known. See result in Figure 5C.
3. We assessed the distribution of the population size against the only numerical method performing the same goal, the particle filtering developed in Vaughan et al. (2019). We compared values of a few quantiles computed using the two methods, see result in Figure 5DEF). Note that Vaughan

et al. considered that we never have data on the removal status of individuals. We thus adapted our developments to this scenario in this specific comparison, by summing updates corresponding to the removal or not of the sampled individuals.

On each of these sanity checks, we verified that different quantities match across different λ values. Note that we could equivalently have chosen any other parameter to be varied.

Figure 5: Assessment of the accuracy of the methods presented in this paper, on toy datasets. First row, probability density of data, A) against known analytical formula when $\omega = 0$ and $(\mu, \rho, \psi, r) = (1, 0.5, 0.3, 0.2)$; B) against known analytical formula when $r = 1$ and $(\mu, \rho, \psi, \omega) = (1, 0.5, 0.3, 0.6)$; C) obtained using Algorithms 1' or 2' otherwise, with $(\mu, \rho, \psi, r, \omega) = (1, 0.5, 0.3, 0.2, 0.6)$. Second row, quantiles of the population size distribution, against the particle filter in Vaughan et al. (2019), with parameters $(\mu, \rho, \psi, r, \omega) = (1, 0.1, 0.001, 0.5, 0.001)$. D) quantile of level 0.2; E) median; F) quantile of level 0.8.

We also illustrate in Figure 6 our target distribution K_t of the past population size conditioned on $(\mathcal{T}, \mathcal{O})$, on a few simulated examples.

Figure 6: Inferred population size distribution K_t using $(\mathcal{T}, \mathcal{O})$ matches the simulated population size trajectory I_t under three different processes: A) A homogeneous birth-death with ρ-sampling at present; B) A homogeneous birth-death with ρ-sampling at present and ψ-sampling through time; C) A homogeneous birth-death process with ρ-, ψ- and ω-sampling. Note that we plot on the same graph k_t, the number of observed lineages in the tree, as this is an obvious lower bound in our population size inference.

6. Discussion

The results we have derived in this paper fit into two main categories. The first category concerns results allowing one to compute the probability density of a tree and occurrences, while the second category concerns results allowing one to compute the probability distribution of the population size in the past. We discuss these two categories below, before presenting ideas for future extensions of the model.

6.1. Using the probability density of the data

We present in this article new ways to compute the probability density of the data, $\mathbb{P}(\mathcal{T}, \mathcal{O})$. For the special cases ($\omega = 0, r = 0$) or ($r = 1$), efficient calculations are available in Stadler (2010); Gupta et al. (2019). Our two Algorithms 1' and 2' have the potential to improve the computation time of $\mathbb{P}(\mathcal{T}, \mathcal{O})$ also when $\omega \neq 0$ and $r \neq 1$. When analysing data, as described below, often this probability density is conditioned on sampling at least one individual, using $u_{t_{or}}$ (Stadler, 2012).

In the case that the tree is known, we can use $\mathbb{P}(\mathcal{T}, \mathcal{O}|\lambda, \mu, \rho, \psi, r, \omega, t_{or})$ (with conditioning on sampling at least one individual) to obtain maximum likelihood parameter estimates for the birth-death parameters λ, μ as well as the sampling parameters ρ, ψ, r, ω. For special cases of this model, it has been shown that not all sampling parameters are identifiable (see e.g. Stadler and Steel (2019)). Future work will involve investigating which of the sampling parameters in the general model can be estimated.

On the other hand, data may consist of sequencing data $\mathcal{A}$ and occurrence data $\mathcal{O}$. Bayesian tools are then typically employed to obtain a sample from the posterior distribution of the parameters using Markov chain Monte Carlo methods. The posterior distribution is,

$$f(\mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or} \mid \mathcal{O}, \mathcal{A}) \propto f(\mathcal{A} \mid \mathcal{T}, \theta) f(\mathcal{O}, \mathcal{T} \mid \lambda, \mu, \rho, \psi, r, \omega, t_{or}) f(\lambda, \mu, \rho, \psi, r, \omega, \theta, t_{or}),$$

with θ summarizing the parameters of the model of molecular evolution and $f(\lambda, \mu, \rho, \psi, r, \omega, \theta, t_{or})$ being the prior distribution on the model parameters.

6.2. Probability distribution of past population sizes

The main results of this paper allow one to compute the probability distribution of the population size in the past and to generate population size trajectories conditioned on $(\mathcal{T}, \mathcal{O})$ (Section 5).

Given a tree and occurrences together with birth-death parameters (which may be the maximum likelihood parameters obtained based on the tree and record of occurrences), we can simulate the distribution of past population sizes as described in Section 5.2. Furthermore, we can calculate the probability of a population size at any time in the past as described in Section 5.1.

If we are instead provided with sequencing data $\mathcal{A}$ and occurrence data $\mathcal{O}$, and want to generate a simulated ensemble characterizing the posterior distribution of past population size trajectories $\mathcal{I}$, we can use the following strategy. The posterior distribution is,

$$f(\mathcal{T}, \mathcal{I}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or} \mid \mathcal{O}, \mathcal{A}) = f(\mathcal{I} \mid \mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or}, \mathcal{O}, \mathcal{A}) f(\mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or} \mid \mathcal{O}, \mathcal{A})$$

We have described above how to obtain a sample from the posterior distribution $f(\mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or} \mid \mathcal{O}, \mathcal{A})$ using Markov chain Monte Carlo. For each sample of $\mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or}$ thus obtained, we can simulate an appropriately conditioned population size trajectory $\mathcal{I}$ as described in Section 5.2. The ensemble of trajectories thus generated has the required distribution. We can employ an analogous procedure if we are interested in the posterior probability distribution of the population size at a particular time t. For each posterior sample of $\mathcal{T}, \theta, \lambda, \mu, \rho, \psi, \omega, t_{or}$, we can calculate the population size distribution at time t using Section 5.1. The posterior population size at time t is then the average over all these conditional distributions.

6.3. Increased efficiency opens new research avenues

Both the density $\mathbb{P}(\mathcal{T}, \mathcal{O})$ and the probability distribution of the population size in the past (K_t) can be obtained using the Monte-Carlo particle filtering algorithm developed in Vaughan et al. (2019). The new approach presented in this paper is nevertheless appealing for two reasons. First, it provides a direct link with previous analytical formulas developed in Stadler (2010); Gupta et al. (2019), thus improving our understanding of these processes and leading to very efficient results in the specific case where $\omega = 0$. Second, Algorithms 1 and 2 have the potential to be more efficient alternatives to the Monte-Carlo particle filtering algorithm. Computing quantiles shown in Figure 5DEF using the particle filtering took a few days, as compared to a few minutes with our method, mainly because it can be applied directly on a fixed tree and does not need to be part of a MCMC. A more thorough quantitative comparison of both approaches would require to implement this work in a MCMC framework, which is beyond the scope of this paper.

This increased efficiency could open up the possibility to analyse much bigger datasets in the near future. In macroevolution, the study of clades with a huge fossil record like *cetaceans* could benefit from our approach. This dataset is characterized by a rather small number of extant species and fossils with morphological data available (respectively ρ-sampled and ψ-sampled species), but includes a huge number of fossils without morphological data (ω-sampled species) (Morlon et al., 2011; Barido-Sottani et al., 2019). For the cetaceans as well as many other clades, it will be of great interest to compute diversity estimates under the same model, our modelling framework presented here (assuming $\rho \neq 0, \omega \neq 0, r = 0$)." Ultimately, all ω-samples could be taken into account to inform the tree and diversity estimates.

In the context of epidemiology, typically, the genetic sequences of the pathogen are only available for a fraction of the infected individuals. These correspond to ψ-samples, while other sampled infected individuals

correspond to ω-samples. Further developing our approach in a Bayesian framework, both the genetic sequences and the record of occurrence could be jointly used to estimate the underlying transmission tree and prevalence of the disease through time. Depending on the cost of sequencing and the ability of numerical methods to handle some critical amount of both genetic sequences and number of occurrences, optimal sampling procedure could be investigated, to make the most of both types of data.

Finally, while improving on current methods, these two Algorithms 1 and 2 still only provide approximations of, respectively, L_t and M_t, that critically rely on the truncation parameter of the state space N. Increasing N leads to a more accurate approximation, while increasing the runtime of the method. If the probability mass of the number of hidden individuals is non-negligible above N, both algorithms will lead to very poor approximations of L_t and M_t. This value should thus be carefully chosen in empirical applications, depending on what is expected with the data at hand. We point out that the behaviour of these algorithms strongly relies on the runtime and accurracy of the matrix exponentiation steps. Numerous matrix exponentiation methods have been proposed in the literature (Moler and Van Loan, 2003). In our current implementation, we rely on a recent matrix exponentiation method already implemented in *scipy* (Al-Mohy and Higham, 2010). Future avenues towards improving this specific step could focus on new theoretical results adapted to tridiagonal matrices (Smith and Shahrezaei, 2015) or alternatively try to adapt Laplace transform approximations derived in Crawford et al. (2014), who present theoretical results bounding the errors made in their approximation.

6.4. Future extensions

Our proposed modelling framework lends itself well for various biologically realistic extensions to allow closer fit to empirical data in a variety of situations.

The first extension that we envision is to relax the assumption of rate homogeneity and instead work with time-varying rates. This has already been considered in different studies relying on birth-death processes, either with exponentially varying functions (Morlon et al., 2011) or with piecewise constant rates (a model dubbed as *skyline birth-death process*, see Stadler et al., 2013; Gavryushkina et al., 2016). As all our results can be straightforwardly adapted to such a framework, this would not require much theoretical work. However, the challenge would be to do so without overfitting the data.

Another popular extension that has been described in the literature on birth-death processes for phylodynamics is to consider multi-type birth-death processes (Maddison et al., 2007). Each individual is assigned a type, which impacts its propensity to give birth to other types. All sampling-related parameters can also be considered type-dependent. The main challenge here boils down to dealing with an increase of dimensionality, because we would be interested in the joint distribution of all subpopulation sizes. This extension is

particularly interesting for epidemiological applications, when different populations of infected individuals, clustered according to some characteristic (e.g. patient behaviour or geography) might have very different dynamics (Stadler and Bonhoeffer, 2013).

Finally, we are very hopeful that this piece of work could be applied as well to density-dependent birth-death processes, also known as *logistic birth-death models*. Indeed, very similar ideas to the breadth-first forward and backward traversals as applied in Algorithms 1' and 2' appear in the context of logistic birth-death models (Etienne et al., 2012; Leventhal et al., 2013; Laudanno et al., 2020). Preliminary results obtained by adapting our numerical algorithms to this framework are very encouraging, and we are currently in the process of deriving as much analytical results as we can to speed up the method. We are hoping to present this in a subsequent paper.

6.5. Conclusion

This manuscript presents a way to efficiently compute the distribution of the past population size in a linear birth-death process, conditioned on the observation of a reconstructed phylogenetic tree and a record of occurrences through time. Such data are very common in macroevolution where the reconstructed phylogenetic tree of extant species is available together with occurrences from the fossil record. In epidemiology, pathogen genetic sequencing data and case count data are a common data source. Our method thus promises to allow efficient quantification of past population sizes, representing past biodiversity or past prevalence, from these rich datasets.

We believe that this method also paves the way for the consideration of more complex and more realistic demographic scenarios, assuming either time-dependent (Morlon et al., 2011; Stadler et al., 2013; Gavryushkina et al., 2016) or density-dependent parameters (Etienne et al., 2012; Leventhal et al., 2013), potentially catering for populations with multiple demographic categories/types (Maddison et al., 2007; Stadler and Bonhoeffer, 2013; Freyman and Höhna, 2018). It is our hope that this manuscript will foster important research advances for unravelling demographic histories in epidemiology, macroevolution, and any other fields where birth-death processes form a relevant model framework.

Acknowledgements

The authors are grateful to Rachel Warnock for helpful discussion on potential applications of the model, and Alex Zarebski for his thorough examination of this manuscript. A.G. is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 743269 (CyberGenetics project)

Al-Mohy, A. H., Higham, N. J., 2010. A new scaling and squaring algorithm for the matrix exponential. SIAM Journal on Matrix Analysis and Applications 31 (3), 970–989.
Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (3), 269–342.
Barido-Sottani, J., Aguirre-Fernández, G., Hopkins, M. J., Stadler, T., Warnock, R., 2019. Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process. Proceedings of the Royal Society B 286 (1902), 20190685.
Billaud, O., Moen, D. S., Parsons, T. L., Morion, H., 2019. Estimating diversity through time using molecular phylogenies: Old and species-poor frog families are the remnants of a diverse past. bioRxiv.
Crawford, F. W., Minin, V. N., Suchard, M. A., 2014. Estimation for general birth-death processes. Journal of the American Statistical Association 109 (506), 730–747.
Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A., Phillimore, A. B., 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proceedings of the Royal Society B: Biological Sciences 279 (1732), 1300–1309.
Foote, M., 2000. Origination and extinction components of taxonomic diversity: general problems. Paleobiology 26 (S4), 74–102.
Freyman, W. A., Höhna, S., 11 2018. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible onagraceae lineages.
Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., Drummond, A. J., 08 2016. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Systematic Biology 66 (1), 57–73.
Gray, R. D., Drummond, A. J., Greenhill, S. J., 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. Science 323 (5913), 479–483.
Gupta, A., Manceau, M., Vaughan, T., Khammash, M., Stadler, T., 2019. The probability distribution of the reconstructed phylogenetic tree with occurrence data. bioRxiv, 679365.
Heath, T. A., Huelsenbeck, J. P., Stadler, T., 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111 (29), 2957–2966.
Kendall, D. G., 1948. On the generalized 'birth-and-death' process. Ann. Math. Stat. 19, 1–15.
Laudanno, G., Haegeman, B., Etienne, R. S., 2020. Additional analytical support for a new method to compute the likelihood of diversification models. Bulletin of mathematical biology 82 (2), 22.
Leventhal, G. E., Günthard, H. F., Bonhoeffer, S., Stadler, T., 2013. Using an epidemiological model for phylogenetic inference reveals density dependence in hiv transmission. Molecular biology and evolution 31 (1), 6–17.
Levin, D. A., Peres, Y., 2017. Markov chains and mixing times. Vol. 107. American Mathematical Soc.
Maddison, W. P., Midford, P. E., Otto, S. P., 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56 (5), 701–710.
Moler, C., Van Loan, C., 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM review 45 (1), 3–49.
Morlon, H., Parsons, T. L., Plotkin, J. B., 2011. Reconciling molecular phylogenies with the fossil record. P. Natl. Acad. Sci. USA 108 (39), 16327–16332.
Morlon, H., Potts, M. D., Plotkin, J. B., 2010. Inferring the dynamics of diversification: A coalescent approach. PLoS Biol. 8 (9), 1–13.
Nee, S., May, R. M., Harvey, P. H., 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London B: Biological Sciences 344 (1309), 305–311.
Quental, T. B., Marshall, C. R., 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends in Ecology & Evolution 25 (8), 434–441.
Ratmann, O., Hodcroft, E. B., Pickles, M., Cori, A., Hall, M., Lycett, S., Colijn, C., Dearlove, B., Didelot, X., Frost, S., et al., 2016. Phylogenetic tools for generalized hiv-1 epidemics: findings from the pangea-hiv methods comparison. Molecular biology and evolution 34 (1), 185–203.
Smith, S., Shahrezaei, V., 2015. General transient solution of the one-step master equation in one dimension. Physical Review E 91 (6), 062119.
Stadler, T., 2010. Sampling-through-time in birth–death trees. Journal of theoretical biology 267 (3), 396–404.
Stadler, T., 2011. Inferring speciation and extinction processes from extant species data. P. Natl. Acad. Sci. USA.
Stadler, T., 2012. How can we improve accuracy of macroevolutionary rate estimates? Systematic biology 62 (2), 321–329.
Stadler, T., Bonhoeffer, S., 2013. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. Philosophical Transactions of the Royal Society of London B: Biological Sciences 368 (1614), 20120198.
Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., et al., 2011. Estimating the basic reproductive number from viral sequence data. Molecular biology and evolution 29 (1), 347–357.
Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A. J., 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences 110 (1), 228–233.
Stadler, T., Steel, M., 05 2019. Swapping Birth and Death: Symmetries and Transformations in Phylodynamic Models. Systematic Biology 68 (5), 852–858.
Starrfelt, J., Liow, L. H., 2016. How many dinosaur species were there? fossil bias and true richness estimated using a poisson sampling model. Philosophical Transactions Of The Royal Society B-Biological Sciences 371 (1691), 20150219.
Vaughan, T. G., Leventhal, G. E., Rasmussen, D. A., Drummond, A. J., Welch, D., Stadler, T., 05 2019. Estimating epidemic incidence and prevalence from genomic data. Molecular Biology and Evolution.
Yule, G. U., 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Phil. Trans. R. Soc. Lond. B.

AppendixA. Solving well-known ODEs

AppendixA.1. The extinction probability

We first deal with equation (2.1) governing u_t, and start by studying the polynomial $\lambda x^2 - \gamma x + \mu$. This polynomial has discriminant $\Delta = \gamma^2 - 4\lambda\mu > (\lambda+\mu)^2 - 4\lambda\mu \geq (\lambda-\mu)^2 \geq 0$. Note that the first inequality holds in the case we are interested in because we can assume that $\psi + \omega > 0$. When this is not the case, one needs to consider that $\lambda \neq \mu$. Roots are

$$x_1 = \frac{\gamma - \sqrt{\Delta}}{2\lambda} \quad \text{and} \quad x_2 = \frac{\gamma + \sqrt{\Delta}}{2\lambda} \quad .$$

Moreover, we know that both roots are positive because $\Delta < \gamma^2 \quad \Rightarrow \quad \sqrt{\Delta} < \gamma \quad \Rightarrow \quad x_1 > 0$.
On an interval including zero and where the polynomial remains positive (as $(-\infty, x_1)$ for example), we can write,

$$\begin{aligned}
& \frac{du}{\lambda u^2 - \gamma u + \mu} = dt \\
\iff \quad & \frac{du}{(x_1 - u)(x_2 - u)} = \lambda dt \\
\iff \quad & \frac{1}{x_2 - x_1}\left(\frac{1}{x_1 - u} - \frac{1}{x_2 - u}\right) du = \lambda dt \\
\iff \quad & \left(\frac{1}{x_1 - u} - \frac{1}{x_2 - u}\right) du = \sqrt{\Delta} dt \quad .
\end{aligned}$$

Integrating both sides between time 0 and t, we get

$$\begin{aligned}
& \frac{x_2 - u_t}{x_1 - u_t} = \frac{x_2 - z}{x_1 - z} e^{\sqrt{\Delta} t} \\
\iff \quad & x_2(x_1 - z)e^{-\sqrt{\Delta}t} - u_t(x_1 - z)e^{-\sqrt{\Delta}t} = x_1(x_2 - z) - u_t(x_2 - z) \\
\iff \quad & u_t\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right) = x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t} \\
\iff \quad & u_t = \frac{x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}}{(x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}} \quad .
\end{aligned}$$

This is the result stated in equation (2.2). Note that this quantity is called $p_0(t)$ in Stadler (2010), or $E(t)$ in Maddison et al. (2007).

AppendixA.2. Probability to leave only one sampled descendent

We aim here to integrate a slight variation of equation (2.3) governing p_t when $k = 1$. The equation we are interested in is,

$$\frac{dW_s}{ds} = (2\lambda u(s,z) - \gamma) k W_s \tag{A.1}$$

$$\frac{dW_s}{W_s} = \left(2\lambda k \frac{x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}s}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}s}} - \gamma k\right) ds$$
$$= \left(\frac{2\lambda k x_1}{\sqrt{\Delta}} \frac{\sqrt{\Delta}(x_2-z)e^{\sqrt{\Delta}s}}{(x_2-z)e^{\sqrt{\Delta}s}-(x_1-z)} - \frac{2\lambda k x_2}{\sqrt{\Delta}} \frac{(x_1-z)\sqrt{\Delta}e^{-\sqrt{\Delta}s}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}s}} - \gamma k\right) ds \quad .$$

All these three terms can be integrated visually between some time t_h and t, leading to,

$$\ln \frac{W_t}{W_{t_h}} = \frac{2\lambda k x_1}{\sqrt{\Delta}} \left[\ln\left((x_2-z)e^{\sqrt{\Delta}s} - (x_1-z)\right)\right]_{t_h}^{t} - \frac{2\lambda k x_2}{\sqrt{\Delta}} \left[\ln\left((x_2-z) - (x_1-z)e^{-\sqrt{\Delta}s}\right)\right]_{t_h}^{t} - \gamma k(t-t_h)$$
$$= \frac{2\lambda k x_1}{\sqrt{\Delta}} \ln \frac{(x_2-z)e^{\sqrt{\Delta}t} - (x_1-z)}{(x_2-z)e^{\sqrt{\Delta}t_h} - (x_1-z)} - \frac{2\lambda k x_2}{\sqrt{\Delta}} \ln \frac{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t_h}} - \gamma k(t-t_h)$$
$$= -\frac{2\lambda k(x_2-x_1)}{\sqrt{\Delta}} \ln \frac{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t_h}} - \gamma k(t-t_h) + 2\lambda k x_1(t-t_h)$$
$$= -2k \ln \frac{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t_h}} - k\sqrt{\Delta}(t-t_h) \quad .$$

Leading to the final expression below

$$W_t = W_{t_h} \left(\frac{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}}{(x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t_h}}\right)^{-2k} e^{-k\sqrt{\Delta}(t-t_h)} \quad . \tag{A.2}$$

Note that the case $k = 1$, $t_h = 0$ and $W_0 = 1 - z$ corresponds to the probability p_t given as equation (2.4),

$$p(t,z) = (1-z)\frac{\Delta}{\lambda^2}\left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-2} e^{-\sqrt{\Delta}t} \quad .$$

while the general case can be expressed using function p as

$$W_t = W_{t_h}\left(\frac{p(t,z)}{p(t_h,z)}\right)^k \quad .$$

AppendixA.3. A few useful properties

Solutions $u(t,z)$ and $p(t,z)$ to ODEs (2.1) and (2.3) satisfy two properties relying on the semi-group property of solutions of ODEs, namely,

$$u(t_2-t_1, u(t_1,z)) = u(t_2,z) \tag{A.3}$$
$$p(t_2-t_1, u(t_1,z)) = \frac{p(t_2,z)p(0,u(t_1,z))}{p(t_1,z)} = \frac{p(t_2,z)(1-u(t_1,z))}{p(t_1,z)} \quad . \tag{A.4}$$

These two properties are useful in many calculations throughout this document, e.g.

- Solving the main PDE in AppendixC requires inverting u, using the first property with,

$$\begin{aligned} & z = u(t - t_h, z_0) \\ \iff & u(t_h - t, z) = u(t_h - t, u(t - t_h, z_0)) \\ \iff & u(t_h - t, z) = u(0, z_0) \\ \iff & z_0 = u(t_h - t, z) \quad . \end{aligned}$$

- The same Appendix section requires also composing function p and u, using

$$\frac{p(t - t_h, u(t_h - t, z))}{p(0, u(t_h - t, z)} = \frac{p(0, z)}{p(t_h - t, z)} \quad .$$

- In the proof of Proposition 4.3, we switch to the notation $R(t, z) = p(t, z)/(1 - z)$ and again compose R and u in the same way,

$$\begin{aligned} & \frac{p(t_j - t, u(t_h - t, z))}{p(0, u(t_h - t, z))} = \frac{p(t_j - t, z)}{p(t_h - t, z)} \\ \iff & \frac{p(t_j - t, u(t_h - t, z))}{1 - u(t_h - t, z)} = \frac{p(t_j - t, z)}{p(t_h - t, z)} \\ \iff & R(t_j - t_h, u(t_h - t, z)) = \frac{R(t_j - t, z)}{R(t_h - t, z)} \quad . \end{aligned}$$

AppendixB. Link with previous work by Gupta et al. (2019)

We aim here at providing details to link this work with results previously derived by Gupta et al. (2019), allowing efficient computation of $L_t^{(i)}$ in the special case $r = 1$.

To do so, we define the distinguishable version of the probability $L_t^{(i)}$ as

$$\overline{L}_t^{(i)} = \frac{(k+i)!}{i!} L_t^{(i)}. \tag{B.1}$$

We now derive the ODE for $\overline{L}_t^{(i)}$. Multiplying both sides of (3.10) by $\frac{(k+i)!}{i!}$ we obtain

$$\begin{aligned} \dot{\overline{L}}_t^{(i)} &= -\gamma(k+i)\frac{(k+i)!}{i!}L_t^{(i)} + \lambda(2k+i)\frac{(k+i)!}{i!}L_t^{(i+1)} + \mu i\frac{(k+i)!}{i!}L_t^{(i-1)} \\ &= -\gamma(k+i)\frac{(k+i)!}{i!}L_t^{(i)} + \lambda(k+i)\left[\frac{(2k+i)(i+1)}{(k+i+1)(k+i)}\right]\frac{(k+i+1)!}{(i+1)!}L_t^{(i+1)} + \mu(k+i)\frac{(k+i-1)!}{(i-1)!}L_t^{(i-1)} \\ &= -\gamma(k+i)\overline{L}_t^{(i)} + \lambda(k+i)\phi_{i,k}\overline{L}_t^{(i+1)} + \mu(k+i)\overline{L}_t^{(i-1)}, \end{aligned} \tag{B.2}$$

where

$$\phi_{i,k} = \frac{(2k+i)(i+1)}{(k+i+1)(k+i)} = 1 - \frac{k(k-1)}{(k+i+1)(k+i)}$$

is the probability that a coalescing pair of randomly chosen lineages (from $(k+i+1)$ total lineages) does not consist of two sampled lineages. This shows that $\overline{L}_t^{(i)}$ satisfies the ODE (B.2) across any epoch. One can see that at punctual events the transition conditions (3.11) and (3.16) hold for $\overline{L}_t^{(i)}$ for ψ-sampling and branching events respectively. Moreover at ω-sampling events the transition condition (3.14) transforms to

$$\overline{L}_{t^+}^{(i)} = \omega(k+i)\overline{L}_{t^-}^{(i-1)}.$$

With these transition conditions and initial condition $\overline{L}_0^{(i)} = \frac{(k_0+i)!}{i!}L_0^{(i)} = \frac{(k_0+i)!}{i!}(1-\rho)^{k_0}\rho^i$, the ODE (B.2) was solved explicitly in Gupta et al. (2019) and the solution is of the form

$$\overline{L}_{t^+}^{(i)} = \sum_{\ell=0}^{q} \frac{(k+i)!}{(i-\ell)!} u_t^{i-\ell} W_t^{(\ell)}$$

where q is the number of ω-sampling events in the time-interval $[0,t)$ and the $(q+1)$-dimensional time-varying vector $W_t = (W_t^{(0)}, \dots, W_t^{(q)})$ can be analytically computed following the approach in Gupta et al. (2019). Therefore from (B.1) we state Proposition 3.3.

AppendixC. Solving the main PDE

We aim now at finding an analytical solution for the following PDE, driving the evolution of $\widehat{M}$ across a given epoch (t_{h-1}, t_h), on which the number of observed lineages remains constant and equal to k,

$$\begin{aligned}
&\widehat{M}(t_h, z) = F(z) \\
&\partial_t \widehat{M} + (\lambda z^2 - \gamma z + \mu)\partial_z \widehat{M} + k(2\lambda z - \gamma)\widehat{M} = 0 \quad .
\end{aligned}$$

AppendixC.1. Principle of the method of characteristics

This problem can be solved by the method of characteristics. We suppose that we can write $\widehat{M}(t,z) = \widehat{M}(t(s), z(s))$ where functions t and z satisfy the ODEs,

$$\begin{aligned}
\frac{dz}{ds} &= \lambda z^2 - \gamma z + \mu \\
\frac{dt}{ds} &= 1 \quad .
\end{aligned}$$

This way, the function $g(s) = \widehat{M}(t(s), z(s))$ satisfies another ODE, that we will have to solve,

$$\begin{aligned}
&\frac{dg}{ds} = \frac{dz}{ds}\partial_z \widehat{M} + \frac{dt}{ds}\partial_t \widehat{M} = (\lambda z^2 - \gamma z + \mu)\partial_z \widehat{M} + \partial_t \widehat{M} = -k(2\lambda z - \gamma)\widehat{M} \\
\iff\ &\frac{dg}{ds} + k(2\lambda z - \gamma)g = 0 \quad .
\end{aligned}$$

AppendixC.2. Step 1, solve for $t(s), z(s)$ and $g(s)$

We start by integrating $t(s)$. We moreover fix that $t(0) = t_h$, thus leading to $t(s) = t_h + s$.

We now turn to z, and notice that it satisfies previously studied ODE (2.1). Integrating between 0 and s leads to,

$$z(s) = u(s, z_0) = \frac{x_1(x_2 - z_0) - x_2(x_1 - z_0)e^{-\sqrt{\Delta}t}}{(x_2 - z_0) - (x_1 - z_0)e^{-\sqrt{\Delta}t}} \quad .$$

Last, g satisfies an ODE very similar to (A.1). Taking care of the minus sign, it leads to the following result,

$$g_s = g_0 \left(\frac{1}{R(s, z_0)} \right)^k \quad . \tag{C.1}$$

AppendixC.3. Step 2, express $\widehat{M}$ back as a function of t, z

We want to express our two unknown quantities s and z_0 as functions of t and z.

On a first hand, we get easily $s = t - t_h$. We moreover can solve for z_0 in the following equation, remembering the semi-group property of u,

$$z = u(t - t_h, z_0) \iff z_0 = u(t_h - t, z) \quad .$$

Substituting these into the previous expression (C.1) of g_s then leads to,

$$\begin{aligned} \widehat{M}(t, z) &= F\left(u(t_h - t, z)\right) R(t - t_h, u(t_h - t, z))^{-k} \\ &= F\left(u(t_h - t, z)\right) R(t_h - t, z)^k \quad . \end{aligned}$$

where the first to second equality relies on a property exposed in AppendixA.3. This gives us the final formula which is stated in Proposition 4.1.

AppendixD. Some useful algebra

This section of the Appendix pools together all bits of algebra that are not really digestible, but are used in the main text.

AppendixD.1. Deivative of $\widehat{M}$

We first modify a bit the expression of the generating function,

$$\widehat{M}(t_h - t, z) = R(t_h - t, z)^k \sum_{l=0}^{N} \widetilde{M}_{t_h}^{(l)} u(t_h - t, z)^l$$

$$= \left(\frac{\Delta}{\lambda^2}e^{-\sqrt{\Delta}t}\right)^k \sum_{l=0}^{N} \widetilde{M}_{t_h}^{(l)} \left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^l \left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l)} \quad .$$

Applying Leibniz's derivation rule to the product, we get,

$$\left(\partial_z^i \widehat{M}(t_h-t,z)\right)_{z=0} = \left(\frac{\Delta}{\lambda^2}e^{-\sqrt{\Delta}t}\right)^k \sum_{l=0}^{N} \widetilde{M}_{t_h}^{(l)} \sum_{\alpha=0}^{i} \binom{i}{\alpha} \left(\partial_z^\alpha \left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^l\right)_{z=0} \left(\partial_z^{i-\alpha}\left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l)}\right)_{z=0} . \tag{D.1}$$

The first of the two derivatives in the sum can be computed as,

$$\begin{aligned}
\partial_z \left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^l &= l\left(-x_1+x_2e^{-\sqrt{\Delta}t}\right)\left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^{l-1} \mathbb{1}_{l\geq 1} \\
\partial_z^2 \left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^l &= l(l-1)\left(-x_1+x_2e^{-\sqrt{\Delta}t}\right)^2\left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^{l-2} \mathbb{1}_{l\geq 2} \\
&\vdots \\
\partial_z^\alpha \left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^l &= \frac{l!}{(l-\alpha)!}\left(-x_1+x_2e^{-\sqrt{\Delta}t}\right)^\alpha\left(x_1(x_2-z) - x_2(x_1-z)e^{-\sqrt{\Delta}t}\right)^{l-\alpha} \mathbb{1}_{l\geq \alpha} \quad .
\end{aligned}$$

While the second gives us,

$$\begin{aligned}
\partial_z \left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l)} &= (2k+l)\left(1-e^{-\sqrt{\Delta}t}\right)\left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l+1)} \\
\partial_z^2 \left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l)} &= (2k+l)(2k+l+1)\left(1-e^{-\sqrt{\Delta}t}\right)^2\left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l+2)} \\
&\vdots \\
\partial_z^{i-\alpha} \left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l)} &= \left(\prod_{m=0}^{i-\alpha-1}(2k+l+m)\right)\left(1-e^{-\sqrt{\Delta}t}\right)^{i-\alpha} \\
&\quad \left((x_2-z)-(x_1-z)e^{-\sqrt{\Delta}t}\right)^{-(2k+l+i-\alpha)} \quad .
\end{aligned}$$

Applying these derivatives in $z=0$ in equation (D.1) yields,

$$\begin{aligned}
\left(\partial_z^i \widehat{M}(t_h-t,z)\right)_{z=0} = \left(\frac{\Delta}{\lambda^2}e^{-\sqrt{\Delta}t}\right)^k &\sum_{\alpha=0}^{i}\sum_{l=\alpha}^{N} \widetilde{M}_{t_h}^{(l)}\binom{i}{\alpha}\frac{l!}{(l-\alpha)!}\left(\prod_{m=0}^{i-\alpha-1}(2k+l+m)\right) \\
&\left(-x_1+x_2e^{-\sqrt{\Delta}t}\right)^\alpha (x_1x_2)^{l-\alpha}\left(1-e^{-\sqrt{\Delta}t}\right)^{l+i-2\alpha}\left(x_2-x_1e^{-\sqrt{\Delta}t}\right)^{-(2k+l+i-\alpha)}
\end{aligned}$$

which is the expression provided in Proposition (4.2).

AppendixD.2. Derivatives of $\widehat{M}$ when $\omega = 0$

We wish here to derive the $\partial_z^i \widehat{M}(t,z)$ where function $\widehat{M}$ is as given in Proposition 4.3, i.e.

$$\widehat{M}(t,z) = \lambda^x \psi^{v+w+y} r^w (1-r)^{v+y} R(t_{or}-t,z) \prod_{t_j\in\mathcal{X}} R(t_j-t,z) \prod_{t_j\in\mathcal{W}} R(t_j-t,z)^{-1} \prod_{t_j\in\mathcal{Y}} u(t_j-t,z)R(t_j-t,z)^{-1} \quad .$$

We take for simplicity the derivative of the logarithm of $\widehat{M}$ and express the derivatives of $\widehat{M}$ using these and Leibniz's formula,

$$\begin{aligned}
\partial_z \widehat{M} &= \widehat{M}\partial_z(\ln \widehat{M}) \\
\partial_z^2 \widehat{M} &= \partial_z \widehat{M}\partial_z(\ln \widehat{M}) + \widehat{M}\partial_z^2(\ln \widehat{M}) \\
\partial_z^3 \widehat{M} &= \partial_z^2 \widehat{M}\partial_z(\ln \widehat{M}) + 2\partial_z \widehat{M}\partial_z^2(\ln \widehat{M}) + \widehat{M}\partial_z^3(\ln \widehat{M}) \\
&\vdots \\
\partial_z^i \widehat{M} &= \sum_{\alpha=1}^{i} \binom{i-1}{\alpha-1} \left(\partial_z^{i-\alpha} \widehat{M}\right) \left(\partial_z^{\alpha}(\ln \widehat{M})\right) \quad .
\end{aligned} \tag{D.2}$$

In order to compute the derivatives of $\ln \widehat{M}$, one needs to get the derivatives of $\ln R(t,z)$ and $\ln u(t,z)$. We have

$$\begin{aligned}
\ln R(t,z) &= -2\ln\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right) + \ln\frac{\Delta}{\lambda^2} - \sqrt{\Delta}t \\
\partial_z \ln R(t,z) &= 2\left(1 - e^{-\sqrt{\Delta}t}\right)\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-1} \\
\partial_z^2 \ln R(t,z) &= 2\left(1 - e^{-\sqrt{\Delta}t}\right)^2\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-2} \\
\partial_z^3 \ln R(t,z) &= 4\left(1 - e^{-\sqrt{\Delta}t}\right)^3\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-3} \\
&\vdots \\
\partial_z^{\alpha} \ln R(t,z) &= 2(\alpha-1)!\left(1 - e^{-\sqrt{\Delta}t}\right)^{\alpha}\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-\alpha} \quad .
\end{aligned}$$

Finally taking the function in $z = 0$ leads to

$$\partial_z^{\alpha} \ln R(t,0) = 2(\alpha-1)!a_t^{\alpha} \tag{D.3}$$

$$\text{where we defined} \quad a_t := \left(1 - e^{-\sqrt{\Delta}t}\right)\left(x_2 - x_1 e^{-\sqrt{\Delta}t}\right)^{-1} \quad . \tag{D.4}$$

In the same way we get,

$$\begin{aligned}
\ln u(t,z) =& \ln\left(x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}\right) - \ln\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right) \\
\partial_z \ln u(t,z) =& -\left(x_1 - x_2 e^{-\sqrt{\Delta}t}\right)\left(x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-1} \\
&+ \left(1 - e^{-\sqrt{\Delta}t}\right)\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-1} \\
\partial_z^2 \ln u(t,z) =& -\left(x_1 - x_2 e^{-\sqrt{\Delta}t}\right)^2\left(x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-2} \\
&+ \left(1 - e^{-\sqrt{\Delta}t}\right)^2\left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-2} \\
\partial_z^3 \ln u(t,z) =& -2\left(x_1 - x_2 e^{-\sqrt{\Delta}t}\right)^3\left(x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta}t}\right)^{-3}
\end{aligned}$$

$$
\begin{aligned}
&+ 2\left(1 - e^{-\sqrt{\Delta t}}\right)^3 \left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta t}}\right)^{-3} \\
&\vdots \\
\partial_z^\alpha \ln u(t,z) =& (\alpha-1)! \left[-\left(x_1 - x_2 e^{-\sqrt{\Delta t}}\right)^\alpha \left(x_1(x_2 - z) - x_2(x_1 - z)e^{-\sqrt{\Delta t}}\right)^{-\alpha} \right. \\
&\left. + \left(1 - e^{-\sqrt{\Delta t}}\right)^\alpha \left((x_2 - z) - (x_1 - z)e^{-\sqrt{\Delta t}}\right)^{-\alpha} \right] \quad .
\end{aligned}
$$

Here also, we are interested in the function in $z = 0$,

$$\partial_z^\alpha \ln u(t,0) = (\alpha-1)!(a_t^\alpha - b_t^\alpha) \tag{D.5}$$

$$\text{where we defined} \quad b_t := \left(x_1 - x_2 e^{-\sqrt{\Delta t}}\right)\left(x_1 x_2 - x_2 x_1 e^{-\sqrt{\Delta t}}\right)^{-1} . \tag{D.6}$$

Last ingredient needed to write the derivative of $\ln \widehat{M}$, we get,

$$
\begin{aligned}
\left(\partial_z^\alpha \ln\left(u(t,z)R(t,z)^{-1}\right)\right)_{z=0} &= \left(\partial_z^\alpha \ln u(t,z)\right)_{z=0} - \left(\partial_z^\alpha \ln R(t,z)\right)_{z=0} \\
&= -(\alpha-1)!(a_t^\alpha + b_t^\alpha) \quad . \qquad (D.7)
\end{aligned}
$$

Finally, using equations D.4 and D.7, one can compute

$$
\begin{aligned}
&\left(\partial_z^\alpha (\ln \widehat{M}(t,z))\right)_{z=0} = (\alpha-1)! C^{(\alpha)} \\
&\text{where we defined} \quad C^{(\alpha)} := 2a_{t_{or}-t}^\alpha + 2\sum_{t_j \in \mathcal{X}} a_{t_j - t}^\alpha - 2\sum_{t_j \in \mathcal{W}} a_{t_j - t}^\alpha - \sum_{t_j \in \mathcal{Y}} (a_{t_j-t}^\alpha + b_{t_j-t}^\alpha) \quad .
\end{aligned}
$$

Plugging this into equation D.2 and noting that $\left(\partial_z^i \widehat{M}(t,z)\right)_{z=0} = i! M_t^{(i)}$, we get

$$M_t^{(i)} = \sum_{\alpha=1}^{i} \binom{i-1}{\alpha-1} \frac{(i-\alpha)!(\alpha-1)!}{i!} M_t^{(i-\alpha)} C^{(\alpha)} = \frac{1}{i} \sum_{\alpha=1}^{i} M_t^{(i-\alpha)} C^{(\alpha)}$$

which is the result stated in corollary 4.4.

AppendixE. Inductions across the epochs

AppendixE.1. Proof of Proposition 3.1

We prove the proposition by induction across the epochs.

If we observe only the first epoch and the k_0 leaves at present, then we get at any time t across the first epoch $(0, t_1)$, $L_t^{(i)} = \rho^{k_0}(1-\rho)^i = u_t^i p_t^{k_0}$, which satisfies Proposition 3.1.

Suppose we observed so far – i.e. on $(0, t_{h+1})$ – v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, x branching events at times $t_j \in \mathcal{X}$, y non-removed leaves at times $t_j \in \mathcal{Y}$. And suppose that Proposition

3.1 is verified across epoch (t_h, t_{h+1}). Let us have a look at what happen across epoch (t_{h+1}, t_{h+2}). The observed punctual event t_{h+1} can either be,

1. a removed ancestral leaf. Update (3.18) then applies. Subsequently, the number of sampled lineages increases by one and formula (3.17) applies on the next epoch, leading to

$$\begin{aligned} L_t^{(i)} &= u_t^i W_t \\ \text{where} \quad W_t &= \lambda^x \psi^{v+(w+1)+y} (1-r)^{v+y} r^{w+1} p_{t_{h+1}}^k \left(\frac{p_t}{p_{t_{h+1}}}\right)^{k+1} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \\ &= \lambda^x \psi^{v+(w+1)+y} (1-r)^{v+y} r^{w+1} p_t^k \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W} \cup \{t_{h+1}\}} p_{t_j}^{-1} \quad . \end{aligned}$$

2. a non-removed ancestral leaf. Update (3.19) then applies. Subsequently, the number of sampled lineages increases by one and formula (3.17) applies on the next epoch, leading to

$$\begin{aligned} L_t^{(i)} &= u_t^i W_t \\ \text{where} \quad W_t &= \lambda^x \psi^{v+w+(y+1)} (1-r)^{v+(y+1)} r^w p_{t_{h+1}}^k u_{t_{h+1}} \left(\frac{p_t}{p_{t_{h+1}}}\right)^{k+1} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \\ &= \lambda^x \psi^{v+w+(y+1)} (1-r)^{v+(y+1)} r^w p_t^{k+1} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y} \cup \{t_{h+1}\}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \quad . \end{aligned}$$

3. a non-removed sampled ancestor along a branch. Update (3.20) then applies. The number of sampled lineages does not changes, and formula (3.17) applies on the next epoch, leading to

$$\begin{aligned} L_t^{(i)} &= u_t^i W_t \\ \text{where} \quad W_t &= \lambda^x \psi^{(v+1)+w+y} (1-r)^{(v+1)+y} r^w p_{t_{h+1}}^k \left(\frac{p_t}{p_{t_{h+1}}}\right)^{k} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \\ &= \lambda^x \psi^{v+w+y+1} (1-r)^{v+y+1} r^w p_t^k \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \quad . \end{aligned}$$

4. a branching event between two sampled lineages. Update (3.21) then applies. The number of sampled lineages decreases by one, and formula (3.17) applies on the next epoch, leading to

$$\begin{aligned} L_t^{(i)} &= u_t^i W_t \\ \text{where} \quad W_t &= \lambda^{x+1} \psi^{v+w+y} (1-r)^{v+y} r^w p_{t_{h+1}}^k \left(\frac{p_t}{p_{t_{h+1}}}\right)^{k-1} \prod_{t_j \in \mathcal{X}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \\ &= \lambda^{x+1} \psi^{v+w+y} (1-r)^{v+y} r^w p_t^{k-1} \prod_{t_j \in \mathcal{X} \cup \{t_{h+1}\}} p_{t_j} \prod_{t_j \in \mathcal{Y}} u_{t_j} p_{t_j}^{-1} \prod_{t_j \in \mathcal{W}} p_{t_j}^{-1} \quad . \end{aligned}$$

In all four cases, Proposition 3.1 is satisfied across epoch (t_{h+1}, t_{h+2}).

AppendixE.2. Proof of Proposition 4.3

This Proposition is also proven by induction across the epochs.

We start at $t_{or} = t_n$ with $k = 1$ lineage. Across epoch (t_{n-1}, t_n), applying Proposition 4.1 with $F(z) = 1$ and $k = 1$, we get $\widehat{M}(t, z) = R(t_{or} - t, z)$, which verifies Proposition 4.3.

Suppose now that Proposition 4.3 is verified across epoch (t_h, t_{h+1}) and that we observed, on (t_h, t_{or}), v sampled ancestors, w removed leaves at times $t_j \in \mathcal{W}$, x branching events at times $t_j \in \mathcal{X}$, y non-removed leaves at times $t_j \in \mathcal{Y}$. Let us have a look at what happens on (t_{h-1}, t_h).
Punctual event t_h can either be,

1. a removed leaf. The number of sampled lineages then goes from $1 + x - y - w$ to $x - y - w$, and applying update (4.33) followed by Proposition 4.1 leads to

$$\begin{aligned}
\widehat{M}(t,z) &= \lambda^x \psi^{v+(w+1)+y} r^{w+1} (1-r)^{v+y} R(t_{or} - t_h, u(t_h - t, z)) R(t_h - t, z)^{x-y-w} \prod_{t_j \in \mathcal{X}} R(t_j - t_h, u(t_h - t, z)) \\
&\quad \prod_{t_j \in \mathcal{W}} R(t_j - t_h, u(t_h - t, z))^{-1} \prod_{t_j \in \mathcal{Y}} u(t_j - t_h, u(t_h - t, z)) R(t_j - t_h, u(t_h - t, z))^{-1} \\
&= \lambda^x \psi^{v+(w+1)+y} r^{w+1} (1-r)^{v+y} \frac{R(t_{or} - t, z)}{R(t_h - t, z)} R(t_h - t, z)^{x-y-w} \\
&\quad \prod_{t_j \in \mathcal{X}} \frac{R(t_j - t, z)}{R(t_h - t, z)} \prod_{t_j \in \mathcal{W}} \frac{R(t_h - t, z)}{R(t_j - t, z)} \prod_{t_j \in \mathcal{Y}} u(t_j - t, z) \frac{R(t_h - t, z)}{R(t_j - t, z)} \\
&= \lambda^x \psi^{v+(w+1)+y} r^{w+1} (1-r)^{v+y} R(t_{or} - t, z) \\
&\quad \prod_{t_j \in \mathcal{X}} R(t_j - t, z) \prod_{t_j \in \mathcal{W} \cup \{t_h\}} R(t_j - t, z)^{-1} \prod_{t_j \in \mathcal{Y}} u(t_j - t, z) R(t_j - t, z)^{-1} \quad .
\end{aligned}$$

   where the first to second equality is detailed in AppendixA, and the second to third comes after canceling out the $R(t_h - t, z)$.

2. a non-removed leaf. The number of sampled lineages then goes from $1 + x - y - w$ to $x - y - w$, and applying update (4.34) followed by Proposition 4.1 leads to

$$\begin{aligned}
\widehat{M}(t,z) &= \lambda^x \psi^{v+w+(y+1)} r^w (1-r)^{v+(y+1)} \frac{R(t_{or} - t, z)}{R(t_h - t, z)} R(t_h - t, z)^{x-y-w} u(t_h - t, z) \\
&\quad \prod_{t_j \in \mathcal{X}} \frac{R(t_j - t, z)}{R(t_h - t, z)} \prod_{t_j \in \mathcal{W}} \frac{R(t_h - t, z)}{R(t_j - t, z)} \prod_{t_j \in \mathcal{Y}} u(t_j - t, z) \frac{R(t_h - t, z)}{R(t_j - t, z)} \\
&= \lambda^x \psi^{v+w+(y+1)} r^w (1-r)^{v+(y+1)} R(t_{or} - t, z) \\
&\quad \prod_{t_j \in \mathcal{X}} R(t_j - t, z) \prod_{t_j \in \mathcal{W}} R(t_j - t, z)^{-1} \prod_{t_j \in \mathcal{Y} \cup \{t_h\}} u(t_j - t, z) R(t_j - t, z)^{-1} \quad .
\end{aligned}$$

3. a sampled ancestor. The number of sampled lineages then remains unchanged and equal to $1+x-y-w$. Applying update (4.35) followed by Proposition 4.1 leads to

$$\widehat{M}(t,z) = \lambda^x \psi^{(v+1)+w+y} r^w (1-r)^{(v+1)+y} \frac{R(t_{or} - t, z)}{R(t_h - t, z)} R(t_h - t, z)^{1+x-y-w}$$

$$\prod_{t_j\in\mathcal{X}}\frac{R(t_j-t,z)}{R(t_h-t,z)}\prod_{t_j\in\mathcal{W}}\frac{R(t_h-t,z)}{R(t_j-t,z)}\prod_{t_j\in\mathcal{Y}}u(t_j-t,z)\frac{R(t_h-t,z)}{R(t_j-t,z)}$$
$$=\lambda^x\psi^{(v+1)+w+y}r^w(1-r)^{(v+1)+y}R(t_{or}-t,z)$$
$$\prod_{t_j\in\mathcal{X}}R(t_j-t,z)\prod_{t_j\in\mathcal{W}}R(t_j-t,z)^{-1}\prod_{t_j\in\mathcal{Y}}u(t_j-t,z)R(t_j-t,z)^{-1}\quad .$$

4. a branching time. The number of sampled lineages then goes from $1+x-y-w$ to $2+x-y-w$, and applying update (4.38) followed by Proposition 4.1 leads to

$$\begin{aligned}\widehat{M}(t,z) &= \lambda^{x+1}\psi^{v+w+y}r^w(1-r)^{v+y}\frac{R(t_{or}-t,z)}{R(t_h-t,z)}R(t_h-t,z)^{2+x-y-w}\\ &\prod_{t_j\in\mathcal{X}}\frac{R(t_j-t,z)}{R(t_h-t,z)}\prod_{t_j\in\mathcal{W}}\frac{R(t_h-t,z)}{R(t_j-t,z)}\prod_{t_j\in\mathcal{Y}}u(t_j-t,z)\frac{R(t_h-t,z)}{R(t_j-t,z)}\\ &=\lambda^{x+1}\psi^{v+w+y}r^w(1-r)^{v+y}R(t_{or}-t,z)\\ &\prod_{t_j\in\mathcal{X}\cup\{t_h\}}R(t_j-t,z)\prod_{t_j\in\mathcal{W}}R(t_j-t,z)^{-1}\prod_{t_j\in\mathcal{Y}}u(t_j-t,z)R(t_j-t,z)^{-1}\quad .\end{aligned}$$

In all these cases, Proposition 4.3 is verified across epoch (t_{h-1},t_h), which ends the proof.

AppendixF. Using a generating function to solve for L_t

AppendixF.1. A slightly different strategy

Recall that L_t verifies the following ODEs,

$$\begin{aligned}\dot{L}_t^{(i)} &= -\gamma(i+k)L_t^{(i)}+\lambda(2k+i)L_t^{(i+1)}+\mu i L_t^{(i-1)}\\ L_0^{(i)} &= \rho_0^{k_0}(1-\rho_0)^i\quad .\end{aligned}$$

If we introduce the corresponding generating function,

$$\widehat{L}(t,z)=\sum_{i=0}^{\infty}z^iL_t^{(i)}$$

then the initial condition on L translates into,

$$\widehat{L}(0,z)=\sum_{i=0}^{\infty}(z(1-\rho))^i\rho^{k_0}=\rho^{k_0}\frac{1}{1-z(1-\rho)}\quad ,\forall z\in\left(\pm\frac{1}{1-\rho}\right)\quad .$$

The ODE translates into a PDE, but not as nicely as for M_t, see below,

$$\partial_t\widehat{L}=\sum_{i=0}^{\infty}z^i\left(-\gamma(i+k)L_t^{(i)}+\lambda(2k+i)L_t^{(i+1)}+\mu i L_t^{(i-1)}\right)$$

$$
\begin{aligned}
&= -\gamma k \sum_{i=0}^{\infty} z^i L_t^{(i)} - \gamma \sum_{i=1}^{\infty} i z^i L_t^{(i)} + \lambda \sum_{i=1}^{\infty} z^{i-1}(2k+i-1) L_t^{(i)} + \mu \sum_{i=0}^{\infty} (i+1) z^{i+1} L_t^{(i)} \\
&= -\gamma k \widehat{L} - \gamma z \partial_z \widehat{L} + (2k-1)\lambda \frac{1}{z}(\widehat{L} - L_t^{(0)}) + \lambda \partial_z \widehat{L} + \mu z^2 \partial_z \widehat{L} + \mu z \widehat{L} \\
&= \left(-\gamma k + (2k-1)\lambda \frac{1}{z} + \mu z \right) \widehat{L} + (\mu z^2 - \gamma z + \lambda) \partial_z \widehat{L} - (2k-1)\lambda \frac{1}{z} \widehat{L}(t,0) \quad .
\end{aligned}
$$

We are thus left with the following PDE problem,

$$
\begin{aligned}
&\widehat{L}(0,z) = \frac{\rho^{k_0}}{1 - z(1-\rho)} \\
&- z \partial_t \widehat{L} + (\mu z^3 - \gamma z^2 + \lambda z) \partial_z \widehat{L} + (\mu z^2 - \gamma k z + (2k-1)\lambda) \widehat{L} - (2k-1)\lambda \widehat{L}(t,0) = 0 \quad . \qquad \text{(F.1)}
\end{aligned}
$$

This remaining term with $\widehat{L}(t,0)$ complicates things a little bit. However, the initial condition on $\widehat{L}$ provides us with a first candidate function to satisfy this PDE.

AppendixF.2. Solution

We introduce below function f, and show that it satisfies the PDE problem (F.1).

$$
f(t,z) := \frac{p_t^k}{1 - z u_t} \quad .
$$

First, we observe that it satisfies the initial condition. We then need to check that it satisfies the PDE, and to do so we expand each of the four components of equation (F.1).

The first one gives us,

$$
\begin{aligned}
-z \partial_t f &= -z \frac{k \dot{p}_t p_t^{k-1}(1 - z u_t) + z \dot{u}_t p_t^k}{(1 - z u_t)^2} \\
&= -z \frac{k(2\lambda u_t - \gamma)(1 - z u_t) + (\lambda u_t^2 - \gamma u_t + \mu) z}{(1 - z u_t)^2} p_t^k \\
&= \frac{\lambda(-2kz u_t - (2k-1) z^2 u_t^2) + \gamma(kz - (k-1) z^2 u_t) - \mu z^2}{(1 - z u_t)^2} p_t^k \quad .
\end{aligned}
$$

We then turn to the second component,

$$
(\mu z^3 - \gamma z^2 + \lambda z) \partial_z f = \frac{\lambda z u_t - \gamma z^2 u_t + \mu z^3 u_t}{(1 - z u_t)^2} p_t^k \quad .
$$

And the third one,

$$
\begin{aligned}
(\mu z^2 - \gamma k z + (2k-1)\lambda) f &= \frac{(1 - z u_t)(\mu z^2 - \gamma k z + (2k-1)\lambda)}{(1 - z u_t)^2} p_t^k \\
&= \frac{\lambda((2k-1) - (2k-1) z u_t) + \gamma(-kz + k z^2 u_t) + \mu(z^2 - z^3 u_t)}{(1 - z u_t)^2} p_t^k \quad .
\end{aligned}
$$

And the fourth and final one,

$$\begin{aligned} -(2k-1)\lambda f(t,0) &= \frac{-\lambda(2k-1)(1-zu_t)^2}{(1-zu_t)^2}p_t^k \\ &= \frac{\lambda(-(2k-1)z^2u_t^2+2(2k-1)zu_t-(2k-1))}{(1-zu_t)^2}p_t^k \quad . \end{aligned}$$

Putting everything together, we can now check that indeed,

$$-z\partial_t f+(\mu z^3-\gamma z^2+\lambda z)\partial_z f+(\mu z^2-\gamma kz+(2k-1)\lambda)f-(2k-1)f(t,0)=0 \quad .$$

While the branching and ψ-sampling with removal updates do not change anything to this solution, all the others do. Further work is thus needed to look for other solutions to this same PDE with different initial conditions.