Journal of Theoretical Biology 352 (2014) 60–70

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics

Amaury Lambert [a,b,*], Helen K. Alexander [c], Tanja Stadler [c]

[a] Sorbonne Universités, UPMC Univ Paris 06, UMR 7599 Laboratoire de Probabilités et Modèles Aléatoires, France
[b] Collège de France, UMR 7241 Center for Interdisciplinary Research in Biology, F-75005 Paris, France
[c] Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, 8092 Zürich, Switzerland

HIGHLIGHTS

- We introduce an epidemiological model with age-dependent removal/sampling rates.
- This framework allows for arbitrary lifetime distributions and heterochronic data.
- We show that viral phylogenies can be represented by a Markovian coalescent point process.
- We derive the likelihood of a phylogenetic tree for parameter inference.
- This method facilitates fast simulation of phylogenetic trees under the model.

ARTICLE INFO

Article history:
Received 23 August 2013
Received in revised form
22 February 2014
Accepted 25 February 2014
Available online 4 March 2014

Keywords:
Birth–death process
Contour process
Coalescent point process
Influenza
HIV

ABSTRACT

The reconstruction of phylogenetic trees based on viral genetic sequence data sequentially sampled from an epidemic provides estimates of the past transmission dynamics, by fitting epidemiological models to these trees. To our knowledge, none of the epidemiological models currently used in phylogenetics can account for recovery rates and sampling rates dependent on the time elapsed since transmission, i.e. age of infection.

Here we introduce an epidemiological model where infectives leave the epidemic, by either recovery or sampling, after some random time which may follow an arbitrary distribution.

We derive an expression for the likelihood of the phylogenetic tree of sampled infectives under our general epidemiological model. The analytic concept developed in this paper will facilitate inference of past epidemiological dynamics and provide an analytical framework for performing very efficient simulations of phylogenetic trees under our model. The main idea of our analytic study is that the non-Markovian epidemiological model giving rise to phylogenetic trees growing vertically as time goes by can be represented by a Markovian "coalescent point process" growing horizontally by the sequential addition of pairs of coalescence and sampling times.

As examples, we discuss two special cases of our general model, described in terms of influenza and HIV epidemics. Though phrased in epidemiological terms, our framework can also be used for instance to fit macroevolutionary models to phylogenies of extant and extinct species, accounting for general species lifetime distributions.

1. Introduction

Phylogenetic trees, which are reconstructed from genetic data, describe the genealogical relationships within a population. The analysis of these trees can provide important insights into the underlying population dynamic processes. For instance, in a group of species descending from a common ancestor, one can construct a tree based on homologous gene(s) sequenced from these species, and thus infer speciation and extinction rates (Nee et al., 1994). As another example, viral genetic sequences extracted from patient

* Corresponding author at: SMILE group, CIRB - Collège de France, 11 place Marcelin Berthelot, F-75005 Paris, France. Tel.: +33 1 44 27 13 91.
E-mail addresses: amaury.lambert@upmc.fr (A. Lambert), helen.alexander@env.ethz.ch (H.K. Alexander), tanja.stadler@env.ethz.ch (T. Stadler).
URLS: http://www.proba.jussieu.fr/pageperso/amaury/index.htm (A. Lambert), http://www.tb.ethz.ch/people/helenal (H.K. Alexander), http://www.tb.ethz.ch/people/tstadler (T. Stadler).

http://dx.doi.org/10.1016/j.jtbi.2014.02.031
0022-5193

samples can provide information on the rate at which an infectious disease transmits within the host population (Stadler et al., 2012).

Maximum likelihood and Bayesian inference are common techniques for estimating such parameters, given a model of the underlying population dynamics. However, the complexity of models that can be applied is limited by the need to derive the likelihood of a phylogenetic tree. Until recently, approaches using a birth–death model framework were limited to death rates of individuals being independent of the age of the individual (see e.g. Nee et al., 1994; Morlon et al., 2011; Stadler, 2011; Etienne et al., 2012 for species phylogenies and Stadler et al., 2012, 2013 for virus phylogenies). Meanwhile, a coalescent-based framework, which epidemiological applications have more widely used (e.g. Pybus et al., 2001; Drummond et al., 2002; Pomeroy et al., 2008; de Silva et al., 2012; Dearlove and Wilson, 2013), does not separately estimate birth and death rates (Stadler et al., 2012).

For phylogenetic trees in which all tips are sampled at one point in time, such as extant species phylogenies, Lambert (2010) and Lambert and Stadler (2013) introduced a framework to calculate the likelihood of a phylogenetic tree accounting for general lifetime distributions. Here we build upon this approach to additionally allow for sequential sampling. Sequential sampling allows analysis e.g. of virus sequence data obtained throughout the course of an epidemic. In the model exposition and worked examples to follow, we focus on an epidemic model in which "births" (branching events) represent the transmission events and "deaths" represent the events of becoming non-infectious either with or without sampling. The model also applies to non-epidemic scenarios in which individuals are sampled at different time points, for instance when dated fossils are included in a species tree.

Allowing age-dependent death/recovery and sampling agrees with the common observation that lifetimes (time being infectious, in the epidemic model) are not generally exponential. For example the infectious period of influenza typically lasts for 5–7 days, according to the U.S. Centers for Disease Control and Prevention (http://www.cdc.gov/flu/about/disease/spread.htm). Extending the model to age-dependent removal will allow the use of genetic sequence data to quantify the death/recovery dynamics more accurately and to test whether parameter estimates (such as the basic reproductive number, R_0) have been biased by the more simplistic assumption of age-independent removal rates. Furthermore, our approach will allow for rapid simulation of phylogenies under age-dependent death/recovery rates even for huge epidemic outbreaks, thus allowing for efficient investigation of the impact of age-dependent rates on the structure of the phylogenetic tree.

The structure of the paper is as follows. First we introduce more precisely the general model of infection and sampling. The forward-in-time (vertical) process which determines the phylogeny is non-Markovian due to the age-dependent removal rates. We then describe the jumping chronological contour process (JCCP, or simply "contour process" for short), a systematic way of exploring trees. The contour process analysis (horizontal) makes use of a Markovian process giving rise to the phylogeny by sequentially adding pairs of coalescence and sampling times in a way that only depends on the previous sampling time.

We proceed to apply Lévy process theory in order to obtain explicit expressions for the Markov process transition probabilities in terms of the so-called scale function associated with the contour process. This leads to the key result of the paper, an explicit formula for the likelihood of a given sampled tree as a function of the parameters of the population dynamic process (Theorem 6.3).

Two worked examples then illustrate application of the general mathematical results: an influenza model, where the lifetime of individuals is not dependent upon whether they leave the epidemic by recovery or sampling; and an HIV model, where sampling occurs after some exponential time during the (independently distributed) infectious period.

We conclude the paper by discussing future challenges of creating a computational inference tool based on this theoretical framework that can be used to analyze pathogen genetic sequence data collected during an epidemic.

2. Model of infection and sampling

We model the dynamics of a population of infectives by a (possibly non-Markovian) branching process, where

- Each infective independently gives birth to a new infective at constant rate b.
- Each infective is removed from the population at rate ρ_1, because of *recovery*, and independently, at rate ρ_2, because of *sampling/detection*, where ρ_1 and ρ_2 are functions of the time elapsed since transmission, hereafter called the *age* of the infective.

The process is assumed to start with one infective at time 0 and is stopped after an overall time duration of t. Let us comment on these assumptions.

Here, a "birth" event is interpreted as the transmission of the infection to a susceptible individual. The assumption that transmissions occur independently and at a constant rate (branching property) is due to the implicit assumption that susceptibles are in excess. The branching property implies in particular that the population of infectives either becomes extinct or asymptotically grows exponentially. Note that we do not assume any latent period, that is, the new infective is assumed to be infectious immediately after infection.

Here, the "death" of an individual is the removal of an individual from the infective population. An infective is removed from the population either because he/she naturally recovers (or actually dies) without detection, or because his/her infection is detected, and by assumption, immediate behavioral changes or successful treatment prevents any further transmission after detection. Removals due to recovery are said to be of type 1, and removals following detection are said to be of type 2. At a detection time, a sample is simultaneously taken from the removed individual and is included in the phylogeny, hence the synonymy between detection and sampling.

The assumptions on ρ_1 and ρ_2 are equivalent to saying that individuals with 'age' a leave the epidemic at rate $\rho(a) := \rho_1(a) + \rho_2(a)$ (i.e., an individual is removed at the first point of a time-dependent Poisson process with instantaneous rate ρ, where time is reset at birth), and that upon leaving the epidemic at age a, they leave it by recovery (without sampling) with probability $\rho_1(a)/\rho(a)$ and by sampling with probability $\rho_2(a)/\rho(a)$.

Mathematically, we can equivalently assume that the type (1 or 2) is chosen upon infection (birth) with probabilities c_1 and $c_2 = 1 - c_1$ respectively, independently of other individuals. Individuals of type 1 live for a duration distributed as V_1 after which they are removed without being sampled. Individuals of type 2 live for a duration distributed as V_2 after which they are simultaneously sampled and removed. This description agrees with the previous one if one sets for $i = 1,2$:

$$c_i := \int_0^\infty \rho_i(z) e^{-\int_0^z \rho(a)\, da}\, dz \quad \text{and} \quad P(V_i \in dz) := c_i^{-1}\, \rho_i(z) e^{-\int_0^z \rho(a)\, da}\, dz.$$

Our analyses and results apply to the general model just described, but we will later use the following two cases as examples. In the first case (framed as an influenza model), V_1 and V_2 are

identically distributed, meaning that the duration of infectiousness does not depend on being sampled. In the second case (framed as an HIV model), natural infectious lifetimes are distributed as some random variable V, while sampling is assumed to occur after some independent exponential duration with parameter μ, meaning that individuals are sampled with a constant rate μ while being infectious. The type of an individual is determined by the first event to occur (removal with or without sampling).

The binary random tree, embedded in continuous time, of this two-type population can be viewed as a two-type *splitting tree*, where in addition the tip of every edge corresponding to the life of an individual of type 2 is *marked* as a sampling point, see Fig. 1. Splitting trees (Geiger and Kersting, 1997; Lambert, 2009, 2010; Lambert and Trapman, 2013) are those random trees generated by a so-called *homogeneous, binary Crump–Mode–Jagers process* (CMJ), that is, a branching process where individuals give birth singly and at constant rate b, during lifetimes that are independent and identically distributed (iid), distributed as some random variable V, which is not necessarily exponentially distributed. In particular, the process counting the total population size is not necessarily Markovian. The law of a splitting tree is characterized by the measure $\pi(\cdot):=bP(V\in\cdot)$ usually called the lifespan measure.

Here, the law of our two-type splitting tree is characterized by the knowledge of the two lifespan measures $\pi_1:=bc_1P(V_1\in\cdot)$ and $\pi_2:=b(1-c_1)P(V_2\in\cdot)$. Notice that regardless of types/marks, the genealogical tree of the whole population (i.e., on both sampled and unsampled individuals) is a splitting tree with lifespan measure $\pi:=\pi_1+\pi_2$.

We call the *sampled tree* the part of the marked splitting tree which is *spanned by its marks and the root*, that is, the phylogenetic tree of samples (i.e., when all lineages without sampled descendants are pruned). See Fig. 1b for a graphical representation. Assuming that the sampled tree can be reconstructed exactly from the patient samples, our goal is to provide a method for computing the probability density (likelihood) of a sampled tree for given parameters under our model. The method can also be used to compute the posterior likelihood of the parameters given the data, in a Bayesian framework where parameters are given a prior distribution. The likelihood thus allows us to infer parameters of the epidemiological process from the sampled tree using maximum likelihood or Bayesian methodology.

From now on, we assume that the tree is embedded in the plane, employing the natural *orientation* where each daughter edge sprouts to the right of its mother edge (see Fig. 1). Our next step is to describe a process which allows us to systematically explore plane splitting trees, and elucidates how plane sampled trees under our model may be represented simply by successive pairs of coalescence and sampling times.

3. The contour process

In Lambert (2010), Lambert has considered the so-called jumping chronological contour process (JCCP), or simply contour process, of the plane splitting tree truncated up to height (time) t. This process can be seen as the path of a ball that follows an outline of the oriented tree, decreasing at unit speed along its edges (which are vertical and embedded in the plane), and jumping instantaneously to the tip of the daughter edge when reaching a node. Fig. 2 shows the contour process associated with the tree in Fig. 1a.

The contour process can also be seen as an alternative representation of the transmission process. The ball starts at the "death" of an infective and slips back until the corresponding infective transmits. Due to the transmission being a Poisson process, we can have the ball slip backward in time until transmission, rather than forward in time until transmission. At transmission, the ball jumps to the time of "death" of the newly infected individual, and again the ball slips back until the next transmission occurs. Once the ball

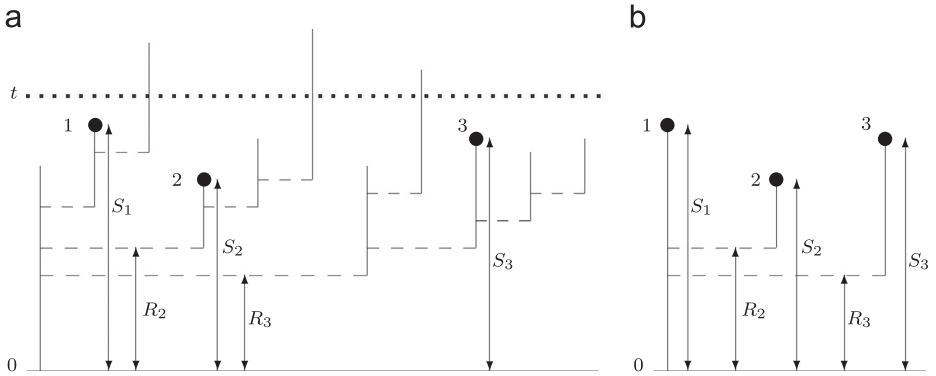

Fig. 1. (a) The oriented phylogeny of the epidemics showing transmission events (horizontal dashed lines) and sampling events (black dots), for 3 infectives sampled before present time t (dotted line), and 3 infectives alive at time t; (b) Sampling times (S_i) and coalescence times (R_i) characterizing the oriented sampled tree (see main text).

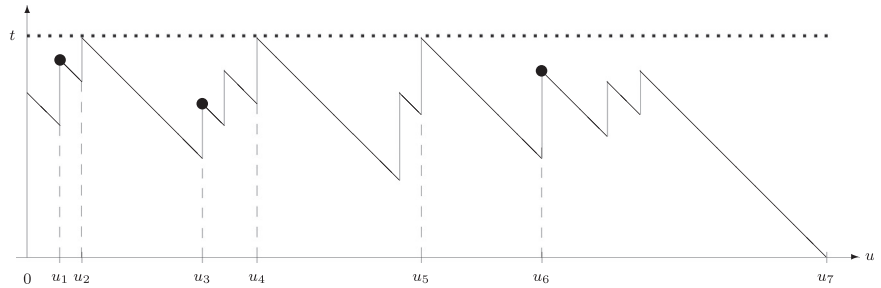

Fig. 2. The marked contour process, with jumps in solid line, which is associated to the marked tree of Fig. 1. Exploration time is denoted by u, and times u_1 to u_6 are all jump times of the contour process corresponding to lifetimes of individuals who are either alive at t or sampled before t. The process terminates at time u_7.

reaches the time of infection of the current infective, it returns to the donor in the infection event of consideration.

Observe that the number of visits of t by the contour process is exactly the number of individuals in the population at time t. Details can be found in Lambert (2010) and Lambert and Trapman (2013). We now seek to uncover the law of this process under our model.

Now let X denote the stochastic process with derivative -1 almost everywhere, which jumps at rate bc_1, with jump sizes distributed as V_1. In probabilistic terms, X is a *compound Poisson process* with *jump measure* π_1 *compensated* at rate -1. In the absence of sampled individuals, we have shown (Lambert, 2010, Theorem 4.3) that the contour process has exactly the same law as the process X reflected below t (meaning sent back to exactly t whenever it overshoots), and killed upon hitting 0.

From now on, X will denote this stochastic process, which, properly reflected and killed, is the contour process of the population *on unsampled individuals*. The idea is that the subpaths between sampled individuals, into which we will later break up the process, can be seen as independent realizations of X. We denote the law of X by P, writing P_x when conditioning on $X_0 = x$. Nevertheless, unless otherwise specified, the denomination 'contour process' will be reserved for the contour process of the *whole* population, that we will denote by Z. It is straightforward that when forgetting about the types of individuals, Z is just the compound Poisson process with jump measure π compensated at rate -1 and reflected below t. Notice that Z is reflected and killed, whereas X is not.

The statement regarding Z can also be seen by the following argument. Recall that regardless of their types, individuals give birth to type 2 individuals at rate bc_2. Since the contour process visits the tree at unit speed, by the lack-of-memory property of the exponential distribution, it is easy to see that the contour process of the two-type splitting tree can be obtained from X by *adding* jumps, whose sizes are distributed as V_2, and which occur after independent exponential random variables with parameter bc_2 (further reflecting this new process under t and killing it upon hitting 0). By analogy with the representation in Fig. 2, we will call these jumps the *marked jumps* of the contour process Z.

4. The coalescent point process with sampling times

In this section we show that pairs of consecutive sampling times and coalescence times in the sampled phylogeny extracted from the contour process give rise to a so-called *coalescent point process*. This observation will allow us to provide an expression for the probability of the sampled tree.

Assume that we label sampled individuals (i.e., type 2 individuals) $1, 2, \ldots$ in the order of the contour, that is, from left to right. We denote by S_i the *sampling time* of individual i, which is, by assumption, the time at which this individual is removed from the infective population (and simultaneously sampled). We further denote by R_i the *coalescence time* between individuals $i-1$ and i, that is, the time at which their most recent common ancestor in the epidemic transmitted the disease to an ancestor of i (which can be assumed, for practical applications, to also be the coalescence time between the pathogens carried by $i-1$ and i).

Straightforward consequences of the definition of the contour process are the following:

1. the sampling time S_i is the value of the contour process at its i-th marked jump;
2. the coalescence time R_i is the infimum of the contour process between the $(i-1)$-th and the i-th marked jump.

In the special case when the progenitor is sampled (before time t), S_1 is actually the lifetime of the progenitor.

More formally, let σ_i denote the time of the i-th marked jump. Formally, if the progenitor of the epidemic is not sampled, then $\sigma_0 := 0$ and for any $i \geq 1$

$$\sigma_i = \inf\{s > \sigma_{i-1} : Z_{s-} < Z_s < t \text{ and this jump of } Z \text{ at time } s \text{ is marked}\},$$

with the convention that $\sigma_i = +\infty$ when there is no marked jump after σ_{i-1}. Otherwise, if the progenitor is sampled, $\sigma_1 := 0$ and the previous definition only holds for $i \geq 2$. If $\sigma_i < \infty$, then

$$S_i = Z(\sigma_i) \quad \text{and} \quad R_i = \inf\{Z(s); \sigma_{i-1} \leq s < \sigma_i\};$$

otherwise if $\sigma_i = +\infty$, then $(R_i, S_i) = (0, 0)$. Our first remark is that the pairs (R_i, S_i) characterize the (plane) sampled tree, as seen in Fig. 1. By analogy with phylogenies spanned by extant individuals (where one can consider $S_i = t$ for all i), we will say that (R_i, S_i) form a *coalescent point process with sampling times* (Aldous and Popovic, 2005; Lambert, 2010; Lambert and Stadler, 2013).

Since each σ_i is a stopping time for Z, observe, by the strong Markov property of Z, that conditional on $\sigma_i < \infty$ and $Z(\sigma_i) = x$, the subpath $\{Z(s); s \geq \sigma_i\}$ is independent of the subpath $\{Z(s); s \leq \sigma_i\}$. In particular, conditional on S_i, the pair (R_{i+1}, S_{i+1}), which is a function of $\{Z(s); s \geq \sigma_i\}$, is independent of $\{Z(s); s \leq \sigma_i\}$. Since all the pairs $\{(R_j, S_j); j \leq i\}$ are functions of $\{Z(s); s \leq \sigma_i\}$, we deduce that conditional on $S_i = x > 0$, the pair (R_{i+1}, S_{i+1}) is independent of $\{(R_j, S_j); j \leq i\}$. In other words, the pairs (R_i, S_i) form a Markov chain, where the transition probability only depends on the second component.

More accurately, they form a *killed Markov chain*, that is, a Markov chain with a possibly finite (random) lifetime, which is the first i such that $(R_i, S_i) = (0, 0)$. More specifically, the transition kernel $p(x, \cdot)$ of a killed Markov chain M with values in some space E is a sub-probability kernel, in the sense that $p(x, E) \leq 1$. Then at each time step n, conditional on $M_n = x$, the Markov chain is killed (has lifetime n) with probability $1 - p(x, E)$, and with probability $p(x, E)$, makes a transition according to the probability kernel $p(x, \cdot)/p(x, E)$.

We can record the previous discussion in the next statement.

Lemma 4.1. *The pairs (R_i, S_i) form a killed Markov chain whose transition probability only depends on the second component. In addition, if we set for any $x \in (0, t]$, $y \in (0, x)$ and $z \in (y, t)$,*

$$p(x; dy\, dz) := \mathbb{P}\left(\inf_{0 \leq s < \sigma_1} Z(s) \in dy, Z(\sigma_1) \in dz | Z_0 = x \right),$$

then for any $i \geq 1$,

$$\mathbb{P}(R_{i+1} \in dy, S_{i+1} \in dz | S_i = x) = p(x; dy\, dz),$$

and conditional on $S_i = x$, the Markov chain is killed at step i with probability

$$k(x) := 1 - \int_{(0,x)} \int_{(y,t)} p(x; dy\, dz) = \mathbb{P}(\sigma_1 = +\infty | Z_0 = x).$$

From now on, we will use the notation $\mathbb{P}_x$ to denote the law of Z when $Z_0 = x$ (recall that P_x is the law of X when $X_0 = x$). With a slight abuse of notation, we define the random pair (R,S) by

$$P_x(R \in dy, S \in dz) := p(x; dy\, dz),$$

and $(R, S) := (0, 0)$ with probability $k(x)$.

Our goal is now to give a finer characterization of the distribution of (R,S), i.e., of the transitions of the coalescent point process with sampling times. For any $y \in [0, t)$, we can classify paths of the contour process according to the three following events:

A'_y – exit of (y,t) from the bottom, i.e. hit y before hitting t and before the first marked jump;

B'_y – arrival of a marked jump (with terminal value in $(0,t)$) before exit of (y,t);
C'_y – exit of (y,t) from the top and before the first marked jump.

Notice that the events A'_y, B'_y, C'_y form a partition. If we denote by $T_{\mathcal{A}}$ the first hitting time of the set $\mathcal{A}$, that is,

$$T_{\mathcal{A}} = \inf\{s > 0 : Z(s) \in \mathcal{A}\},$$

and if we write σ for σ_1, we can express the events A'_y, B'_y and C'_y as follows:

$$A'_y = \{T_y < T_t \wedge \sigma\}, \quad B'_y = \{\sigma < T_t \wedge T_y\}, \quad C'_y = \{T_t < T_y \wedge \sigma\},$$

where we used the usual notation $a \wedge b = \min(a,b)$. Now an iterative application of the strong Markov property of Z at its successive hitting times of t shows that

$$P_x(R > y, S \in dz) = \mathbb{P}_x(B'_y, Z(\sigma) \in dz) + \mathbb{P}_x(C'_y) \sum_{n \geq 0} \mathbb{P}_t(C'_y)^n \mathbb{P}_t(B'_y, Z(\sigma) \in dz)$$

$$= \mathbb{P}_x(B'_y, Z(\sigma) \in dz) + \frac{\mathbb{P}_x(C'_y)}{1 - \mathbb{P}_t(C'_y)} \mathbb{P}_t(B'_y, Z(\sigma) \in dz).$$

Similarly,

$$k(x) = \mathbb{P}_x(A'_0) + \frac{\mathbb{P}_x(C'_0)}{1 - \mathbb{P}_t(C_0{}')} \mathbb{P}_t(A'_0).$$

Apportioning the paths of Z as we just did will now allow us to express the law of (R,S) in terms of the law of the Lévy process X rather than that of the contour process Z. We stick to the notation $T_{\mathcal{A}}$ for the first hitting time of the set $\mathcal{A}$ by X. Now recall that the paths of Z can be obtained by adding independent jumps, distributed as V_2, to the paths of X, at rate

$$q := bc_2,$$

and further reflecting those paths below t. As a consequence, if $\mathbf{e}$ denotes an independent exponential random variable with parameter q, and if V_2 is assumed to be independent of $\mathbf{e}$ and X, the events A'_y, B'_y, C'_y have the same law under $\mathbb{P}_x$, respectively, as the events A_y, B_y, C_y under P_x, where

$$A_y := \{T_y < T_{(t,+\infty)} \wedge \mathbf{e}\},$$

$$B_y := \{\mathbf{e} < T_y \wedge T_{(t,+\infty)}, X_{\mathbf{e}} + V_2 \leq t\},$$

$$C_y := \{T_{(t,+\infty)} < T_y \wedge \mathbf{e}\} \cup \{\mathbf{e} < T_y \wedge T_{(t,+\infty)}, X_{\mathbf{e}} + V_2 > t\}.$$

More precisely, we arrive at the following statement, where $\underline{X}$ and $\overline{X}$ denote respectively the infimum and supremum processes of X, that is,

$$\underline{X}_s = \inf_{0 \leq u \leq s} X_u \quad \text{and} \quad \overline{X}_s = \sup_{0 \leq u \leq s} X_u.$$

Proposition 4.2. *Let $x \in (0,t]$, $y \in (0,x)$ and $z \in (y,t)$. Then*

$$P_x(R > y,\ S \in dz) = P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t, X_{\mathbf{e}} + V_2 \in dz) + \frac{P_x(C_y)}{1 - P_t(C_y)} P_t(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t, X_{\mathbf{e}} + V_2 \in dz),$$

and

$$k(x) = P_x(A_0) + \frac{P_x(C_0)}{1 - P_t(C_0)} P_t(A_0).$$

We will now use the fact that X is a Lévy process in order to obtain explicit expressions for the above probabilities, finally leading to an explicit expression for the probability of a sampled tree in Theorem 6.3. In the following section, we first introduce the necessary background results on Lévy processes.

5. Lévy processes and scale functions

The standard results presented in this section can be found in Bertoin (1996, 1997), and Lambert and Trapman (2013). We state these results in terms of an arbitrary compound Poisson process Y with jump measure π on $(0,+\infty)$ with total mass b, compensated at rate -1. We stick to the notation defined earlier for X (law P_x when started from x, first hitting time $T_{\mathcal{A}}$ of $\mathcal{A}$ and extremum processes $\underline{Y}$ and $\overline{Y}$). It can be convenient to characterize the law of this process by its Laplace exponent ψ defined by

$$\psi(\lambda) := \lambda - \int_0^\infty \pi(dx)(1 - e^{-\lambda x}), \quad \lambda \geq 0. \tag{1}$$

The function ψ is differentiable and convex and we denote by η its largest root. Then ψ is increasing on $[\eta, +\infty)$ and we denote by ϕ its inverse on this set, so that ϕ is a bijection from $[0,\infty)$ to $[\eta,\infty)$.

The probability of exit of an interval (from the bottom or from the top) by Y has a simple expression (see e.g. Bertoin, 1996) in the form

$$P_x(T_0 < T_{(t,+\infty)}) = \frac{W(t-x)}{W(t)}, \quad t \geq x \geq 0, \tag{2}$$

where the so-called *scale function* W is the non-negative, non-decreasing, differentiable function such that $W(0) = 1$, characterized by its Laplace transform

$$\int_0^\infty dx\, e^{-\lambda x} W(x) = \frac{1}{\psi(\lambda)}, \quad \lambda > \eta. \tag{3}$$

Eq. (2) gives the probability that Y exits $(0,t]$ from the bottom of the interval. The following formula gives the Laplace transform of T on this event, where T denotes the first exit time of $(0,t]$, that is,

$$T = T_0 \wedge T_{(t,+\infty)}.$$

For any $q > 0$,

$$E_x(e^{-qT} \mathbf{1}_{\{T_0 < T_{(t,+\infty)}\}}) = \frac{W^{(q)}(t-x)}{W^{(q)}(t)}, \tag{4}$$

where the so-called *q-scale function* $W^{(q)}$ is the non-negative, nondecreasing, differentiable function such that $W^{(q)}(0) = 1$, characterized by its Laplace transform

$$\int_0^\infty dx\, e^{-\lambda x} W^{(q)}(x) = \frac{1}{\psi(\lambda) - q}, \quad \lambda > \phi(q). \tag{5}$$

Note that $W^{(0)} \equiv W$. Last, the q-resolvent of the process killed upon exiting $(0,t]$ is given by the following formula:

$$u^q_t(x,z)\, dz := E_x\left(\int_{s=0}^T ds\, e^{-qs} \mathbf{1}_{\{Y_s \in dz\}}\right) = \frac{W^{(q)}(t-x) W^{(q)}(z)}{W^{(q)}(t)} - \mathbf{1}_{\{z \geq x\}} W^{(q)}(z-x). \tag{6}$$

Observe that by the Fubini–Tonelli Theorem

$$q u^q_t(x,z)\, dz = E_x\left(\int_{s=0}^T \mathbf{1}_{\{\mathbf{e} \in ds\}} \mathbf{1}_{\{Y_s \in dz\}}\right) = P_x(\mathbf{e} < T, Y_{\mathbf{e}} \in dz) = P_x(\underline{Y}_{\mathbf{e}} > 0, \overline{Y}_{\mathbf{e}} \leq t, Y_{\mathbf{e}} \in dz), \tag{7}$$

where $\mathbf{e}$ denotes an independent exponential random variable with parameter q. The previous formula is key to computing the probabilities involved in Proposition 4.2 (see Appendix). We will use the following useful lemma (proved in the Appendix) several times.

Lemma 5.1. *For any $z, q \geq 0$,*

$$\int_0^z W^{(q)}(z-x)\pi(dx) = (q+b) W^{(q)}(z) - W^{(q)'}(z).$$

6. The likelihood of the sampled tree

We now apply the results from Section 5 to the process X (the contour process on nonsampled individuals), in order to give an explicit formula for the probabilites displayed in Proposition 4.2. Let ψ_1 be the Laplace exponent of X:

$$\psi_1(\lambda) = \lambda - \int_0^\infty bc_1 P(V_1 \in dx)(1-e^{-\lambda x}),$$

and $W_1^{(q)}$ the q-scale function associated with ψ_1 and defined in (5), required now for the specific $q = bc_2$. Note that all formulae given in the previous section hold for a general q, but that from now on we will always assume $q = bc_2$. We will use the following definitions:

$$C_1^{(q)}(z) := q \int_0^z W_1^{(q)}(z-u) P(V_2 \in du), \tag{8}$$

and

$$U_1^{(q)}(z) := 1 + \int_0^z C_1^{(q)}(x)\, dx = 1 + q \int_0^z W_1^{(q)}(z-u) P(V_2 \le u)\, du. \tag{9}$$

The last equality comes from an application of Fubini–Tonelli theorem and a change of variable. Notice in particular that $U_1^{(q)\prime} = C_1^{(q)}$. Then we have the following results, for which proofs can be found in the Appendix.

Lemma 6.1. *Let $x \in (0,t]$, $y \in [0,x)$ and $z \in (y,t)$. Then*

$$P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \le t, X_{\mathbf{e}} + V_2 \in dz) = \left(\frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} C_1^{(q)}(z-y) - \mathbf{1}_{\{z \ge x\}} C_1^{(q)}(z-x) \right) dz.$$

Lemma 6.2. *Let $x \in (0,t]$ and $y \in [0,x)$. Then*

$$P_x(C_y) = U_1^{(q)}(t-x) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} U_1^{(q)}(t-y).$$

We can now state the main result of this paper.

Theorem 6.3. *The sequence $S_1, (R_2,S_2), (R_3,S_3), \dots$ is a killed Markov chain where the transition probability only depends on the second component (S_i), and for any $x \in (0,t]$, $y \in [0,x)$ and $z \in (y,t)$, the starting point has distribution*

$$P(S_1 \in dz) = c_2 P(V_2 \in dz) + \left(c_2 \int_0^z P(V_2 \in du) W_1^{(q)\prime}(z-u) - \frac{C_1^{(q)}(z) C_1^{(q)}(t)}{b U_1^{(q)}(t)} \right) dz,$$

the transition probability $p(x; dy\, dz) = P_x(R \in dy, S \in dz)$ is characterized by

$$P_x(R > y, S \in dz) = \left(C_1^{(q)}(z-y) \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t-y)} - \mathbf{1}_{\{z \ge x\}} C_1^{(q)}(z-x) \right) dz,$$

and the killing probability is

$$k(x) = \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t)}. \tag{10}$$

The probability p that at least one individual is sampled before time t (i.e., the sequence is not empty) is given by

$$p = \int_0^t P(S_1 \in dz) = \frac{C_1^{(q)}(t)}{b U_1^{(q)}(t)}. \tag{11}$$

When the chain is conditioned upon the number n of sampled individuals, it remains a Markov chain $((R_i,S_i); 1 \le i \le n)$, but the transition probability becomes $p(x; dy\, dz)/(1-k(x))$, which now integrates to 1.

The formula for the transition probability is a direct consequence, by elementary calculus, of Proposition 4.2 and Lemmas 6.1 and 6.2. The remaining statements are proved in the Appendix.

In the rest of this section, *we assume that V_2 has a density*, say g_2, in the sense that $P(V_2 \in du) = g_2(u)\, du$, so that $C_1^{(q)}$ is differentiable with derivative

$$C_1^{(q)\prime}(z) = q g_2(z) + q \int_0^z W_1^{(q)\prime}(z-u) g_2(u)\, du, \tag{12}$$

where the first term comes from differentiating the integral as a function of its upper bound and the second one comes from differentiating the function of z inside the integral. The first consequence is that S_1 has a density, say g, given by

$$g(z) = b^{-1} \left(C_1^{(q)\prime}(z) - \frac{C_1^{(q)}(z) C_1^{(q)}(t)}{U_1^{(q)}(t)} \right). \tag{13}$$

The second consequence is that the transition probability has density, say f,

$$P_x(R \in dy, S \in dz) = p(x; dy\, dz) =: f(x; y, z)\, dy\, dz,$$

where by differentiating the expression given in Theorem 6.3 for $P_x(R > y, S \in dz)/dz$ with respect to y and recalling that $U_1^{(q)\prime} = C_1^{(q)}$, we get

$$f(x; y, z) = \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t-y)} \left[C_1^{(q)\prime}(z-y) - C_1^{(q)}(z-y) \frac{C_1^{(q)}(t-y)}{U_1^{(q)}(t-y)} \right]. \tag{14}$$

Then we can directly write down the likelihood of a given oriented tree as follows:

Corollary 6.4. *For any given oriented tree $\mathcal{T}$ with coalescence times $(y_i)_{2 \le i \le n}$ and sampling times $(z_i)_{1 \le i \le n}$, where tips are labeled from left to right, the likelihood $\mathcal{L}_S(\mathcal{T})$ of this tree under the general epidemiological model observed at time t, conditional on at least one sampled individual, is*

$$\mathcal{L}_S(\mathcal{T}) = \frac{g(z_1) k(z_n)}{p} \prod_{i=2}^{n} f(z_{i-1}; y_i, z_i),$$

where k and p are given respectively by (10) and (11) in Theorem 6.3, and g and f are given respectively by (13) and (14).

Alternatively, we can condition on the number n of sampled individuals ($n \ge 1$). Applying the remark in Theorem 6.3, we obtain the conditional likelihood $\mathcal{L}_n(\mathcal{T})$

$$\mathcal{L}_n(\mathcal{T}) = \frac{g(z_1) k(z_n)}{p} \prod_{i=2}^{n} \frac{f(z_{i-1}; y_i, z_i)}{1 - k(z_{i-1})}.$$

7. Worked examples

For illustration, we now describe two specific cases of the general model, meant as simplistic descriptions of influenza and HIV epidemics. We apply our mathematical results to these cases by deriving the expressions required for the likelihood under certain simplifying assumptions. We emphasize, however, that any practitioner working with data should carefully consider whether the assumptions made below are suitable for their disease of interest, or whether a different specification of the general model would be more appropriate.

7.1. Influenza

In one special case of the model, which we envision as a description of influenza, we assume that the outcome of an infection is clear after a certain (random) amount of time, on the order of a few days: either the individual has a normal, mild case of influenza, and recovers at this point (without sampling), or it becomes apparent that the infection is severe and the individual is

hospitalized, with concomitant sampling of their virus. We assume that hygiene and isolation measures in the hospital imply that such an individual is also removed from the infectious population. We suppose that the former outcome occurs with given probability c_1 and the latter with probability $c_2 = 1 - c_1$; realistically for influenza, c_1 is likely to be much larger than c_2, except when restricted to particular sub-populations at risk (e.g. the elderly). Furthermore, we assume that V_1 and V_2 are equal in distribution, i.e. the time until the outcome is "decided" does not depend on the outcome itself. Recall that, as per the general model, all infective individuals are assumed to have the same transmission rate (b) regardless of the outcome (severity) of their infection.

The following statement is a straightforward consequence of Lemma 5.1 and is needed for practical applications of Theorem 6.3. In the case when $V_1 = V_2$ has a density, it is recommended to use Eqs. (12)–(14) for such practical applications.

Proposition 7.1. *In the influenza model, we have*

$$C_1^{(q)}(z) = (c_2/c_1)(bW_1^{(q)}(z) - W_1^{(q)\prime}(z)),$$

so that

$$U_1^{(q)}(z) = 1 + (c_2/c_1)\left(1 + b\int_0^z W_1^{(q)}(s)\,ds - W_1^{(q)}(z)\right).$$

7.2. HIV

In a second special case of the model, which we envision as a description of HIV, we suppose that individuals have a "natural" infectious lifetime, denoted as V; we leave the distribution of V arbitrary in the following derivations. This lifetime would apply if there were no intervention, and since HIV is incurable, the end of this lifetime would indeed correspond to the individual's death. Without treatment, the time until death (shortly after the onset of AIDS) is typically several years (median around 10 years but with substantial variation, Collaborative Group on AIDS Incubation and HIV Survival, 2000).

We then apply sampling at a constant rate, μ, "on top" of these natural infectious lifetimes. That is, sampling occurs after an independent exponential duration, say $\mathbf{e}'$, with rate parameter μ; if this time falls before the end of an individual's "natural" lifetime, he/she is considered to be sampled. A constant rate is not entirely unreasonable since, outside the relatively short acute phase of the infection, the disease is asymptomatic until the onset of AIDS. Importantly, we assume that sampling occurs concomitantly with an effective intervention (successful drug treatment and/or behavioral adjustments) that permanently prevents this individual from transmitting further, i.e. removes him/her from the infectious population. That is, (i) all individuals receiving interventions that may affect the "natural" lifetime defined above have their virus sampled (otherwise the definition of lifetime for non-sampled individuals should be adjusted accordingly), and (ii) any such individual immediately and completely ceases any transmission. Assumption (i) requires that viral sequencing occurs upon diagnosis, which is now common for drug resistance testing in some countries. Assumption (ii) is of course never perfectly accurate, but can be partially justified by the high efficacy of modern drug treatment regimens when available, and has been used in previous HIV data analysis (Stadler et al., 2012). These assumptions may be reasonable in some resource-rich countries, but likely not in many resource-poor countries. Note that the model used in Stadler et al. (2012) is in fact a special case of the model described here, where both "natural death" (removal without sampling) and removal with sampling occur at constant rates. We deal with this Markovian case at the end of the section, but first present derivations for an arbitrary distribution of "natural" lifetime V.

Setting $V^\mu := \min(V, \mathbf{e}')$, where V and $\mathbf{e}'$ are assumed to be independent, we have the probability of sampling:

$$c_2 = P(V^\mu = \mathbf{e}') = P(\mathbf{e}' < V) = 1 - E(e^{-\mu V})$$

which we can rewrite as

$$c_2 = P(V > \mathbf{e}') = \int_{(0,\infty]} \mu e^{-\mu r} P(V > r)\,dr = 1 - \int_{(0,\infty]} e^{-\mu r} P(V \in dr)$$
$$= 1 - c_1. \tag{15}$$

Furthermore,

$$P(V_1 \in dr) = c_1^{-1} e^{-\mu r} P(V \in dr) \quad \text{and} \quad P(V_2 \in dr) = c_2^{-1} \mu e^{-\mu r} P(V > r)\,dr. \tag{16}$$

Taking $\psi(\lambda) := \lambda - b\int_0^\infty (1 - e^{-\lambda r}) P(V \in dr)$ and manipulating Eq. (15) yield

$$c_2 = \frac{\mu - \psi(\mu)}{b},$$

while,

$$\psi_1(\lambda) := \lambda - bc_1 \int_0^\infty (1 - e^{-\lambda r}) P(V_1 \in dr)$$
$$= \lambda - bc_1 + b\int_0^\infty e^{-(\lambda+\mu)r} P(V \in dr) = \psi(\lambda+\mu) - \psi(\mu).$$

Now since $\psi_1(\lambda) = \psi(\lambda+\mu) - \psi(\mu)$ and $q = bc_2 = \mu - \psi(\mu)$, notice that $\psi_1(\lambda) - q = \psi(\lambda+\mu) - \mu$.

The following statement is needed for practical applications of Theorem 6.3. The proof is found in the appendix.

Proposition 7.2. *In the HIV model, we have*

$$C_1^{(q)}(z) = \mu \int_0^z e^{-\mu x} W_1^{(q)\prime}(z-x)\,dx$$
$$= \mu\left(W_1^{(q)}(z) - 1 - \int_0^z \mu e^{-\mu x}(W_1^{(q)}(z-x) - 1)\,dx\right),$$

$$U_1^{(q)}(z) = 1 + \mu \int_0^z dx\, e^{-\mu x}(W_1^{(q)}(z-x) - 1) = W_1^{(q)}(z) - \mu^{-1} C_1^{(q)}(z),$$

and the initial distribution of S_1 is given by

$$P(S_1 \in dz) = \frac{\mu}{b}\left(W_1^{(q)\prime}(z) - \frac{C_1^{(q)}(z) W_1^{(q)}(t)}{U_1^{(q)}(t)}\right)dz.$$

We make further computations in the Markovian case, that is, when the "natural" lifetime of individuals ends at constant rate δ. Then $\pi(dr) = b\delta e^{-\delta r}\,dr$ and

$$\psi(\lambda+\mu) - \mu = \frac{Q(\lambda)}{\lambda + \mu + \delta},$$

where $Q(\lambda) = \lambda^2 + \lambda(\mu + \delta - b) - b\mu$. Then the polynomial Q has two distinct real roots

$$-\alpha_1 = \left(b - \delta - \mu - \sqrt{(\mu+\delta-b)^2 + 4b\mu}\right)/2 \quad \text{and}$$
$$\alpha_2 = \left(b - \delta - \mu + \sqrt{(\mu+\delta-b)^2 + 4b\mu}\right)/2,$$

where both α_1 and α_2 are positive. Using $\alpha_2 - \alpha_1 = b - \delta - \mu$, we get

$$\frac{1}{\psi(\lambda+\mu) - \mu} = \frac{1}{\alpha_1 + \alpha_2}\left[\frac{\alpha_2 - b}{\lambda + \alpha_1} + \frac{\alpha_1 + b}{\lambda - \alpha_2}\right],$$

so that

$$W_1^{(q)}(x) = \frac{\alpha_2 - b}{\alpha_1 + \alpha_2} e^{-\alpha_1 x} + \frac{\alpha_1 + b}{\alpha_1 + \alpha_2} e^{\alpha_2 x}, \quad x \geq 0.$$

We demonstrate in the Appendix that applying Theorem 6.3 with the above expression for $W_1^{(q)}(x)$ leads to the same expression for the likelihood as derived previously using methods particular to this Markovian case (Stadler et al., 2012).

8. Discussion

We introduced a stochastic population dynamics model giving rise to phylogenetic trees with sequentially sampled tips. The lifetime of the individuals within the population may follow an arbitrary distribution, while the production of "daughter" individuals occurs with a constant rate. We showed that the coalescent point process with sampling times formed by pairs of coalescence and sampling times in the left-to-right order satisfies the Markov property. We characterized the law of this Markov chain, providing a framework to calculate the likelihood of a phylogenetic tree, as displayed in Theorem 6.3 and especially in Corollary 6.4.

Evaluating the likelihood of a phylogenetic tree requires the numerical evaluation of the function $W_1^{(q)}$. This evaluation can be performed either by solving the inverse Laplace transform in (3) or the integro-differential equation in Lemma 5.1. We leave the numerical challenges for a future study. However, for the special case of exponentially distributed lifetimes, analytic solutions for the inverse Laplace transform and thus also for the likelihood of the sampled tree are available (Stadler, 2010; Stadler et al., 2012). A special section is dedicated to this case in the Appendix.

We envision using the model on epidemiological data in the following way. Pathogen genetic sequencing data from different hosts is used to reconstruct the genealogical relationship of the data, i.e. the phylogenetic tree. This phylogenetic tree is treated as a proxy for the transmission tree (i.e. branching events are transmission events). We do not deal with this reconstruction and assume for our method that the reconstructed tree is provided. We then assume that the model introduced in this paper gave rise to the transmission tree, and want to fit the model to the tree using the likelihood function. There are two ways to do the fitting. First, the likelihood of the tree can be used for determining maximum likelihood parameter estimates for a given sampled phylogenetic tree, by optimizing the expression for the likelihood of the sampled tree over the parameters. Second, the likelihood together with prior distributions on the model parameters can be used in a Bayesian framework to obtain the posterior distribution of parameters given a sampled tree.

We stress that real data (i.e. sequences, sampling times and/or the associated phylogenetic tree) do not come with the information on the orientation of the tree. However, different orientations lead to different likelihoods, since different orientations can give rise to different precise pairings of successive coalescence and sampling times (R and S). An additional computational challenge is thus to sum the likelihood over all valid (R,S) pairings.

The second useful application of our framework is concerned with the simulation of phylogenetic trees. If simulating the model forward in time, one must simulate many non-sampled individuals, and thus it takes much computational time to obtain the required number of samples. However, using the Markov chain property of our coalescent and sampling time pairs, we can sample once from the distribution for the starting point and $n-1$ times from the distribution specifying the Markov chain in order to obtain a tree on n tips of a specified tree age (observation time t).

So far we had to assume a constant birth rate. Generalizing the results to time-dependent birth rates, as well as death/sampling rates, should be conceptually straightforward: the ball in the contour process is simply rolling back towards 0 with a varying speed. However, generalizing to age-dependent birth rates, i.e. an arbitrary distribution of time until birth of a new individual, will most likely be unachievable with the current framework, as we can no longer let the ball in the contour process roll back along an edge without knowing the age of the individual it represents.

We conclude by emphasizing that our analyses were performed with an epidemiological application in mind; however, any implementation may also be useful for analyzing phylogenetic trees with sequentially sampled tips arising in different applications, such as species phylogenies with fossil tips.

Acknowledgments

A.L. was financially supported by Grant MANEGE 'Modèles Aléatoires en Écologie, Génétique et Évolution' 09-BLAN-0215 of ANR (French national research agency). A.L. also thanks the *Center for Interdisciplinary Research in Biology* (Collège de France) for funding. H.A. received support from ETH Zürich. T.S. thanks the Swiss National Science foundation for funding (SNF Grant #PZ00P3 136820).

Appendix A. Proofs

A.1. Proof of Lemma 5.1

By an integration by parts, the Laplace transform (as a function of $\lambda > \phi(q)$) of the non-negative function $z \mapsto W^{(q)'}(z) + \int_0^z W^{(q)}(z-x)\pi(dx)$ equals

$$[e^{-\lambda z}W^{(q)}(z)]_0^\infty + \frac{\lambda}{\psi(\lambda)-q} + \frac{\int_0^\infty \pi(dx)e^{-\lambda x}}{\psi(\lambda)-q}$$
$$= -1 + \frac{\lambda}{\psi(\lambda)-q} + \frac{\psi(\lambda)-\lambda+b}{\psi(\lambda)-q} = \frac{q+b}{\psi(\lambda)-q},$$

where we used successively the facts that the Laplace transform of $W^{(q)}$ is $1/(\psi(\lambda)-q)$, that the Laplace transform of a convolution product is the product of Laplace transforms, and that $W^{(q)}(0)=1$. Now the right-hand side is also the Laplace transform of the non-negative function $z \mapsto (q+b)W^{(q)}(z)$. □

A.2. Proof of Lemma 6.1

Set

$$H^{(q)}(x,t;dz) := P_x(\underline{X}_{\mathbf{e}} > 0, \overline{X}_{\mathbf{e}} \le t, X_{\mathbf{e}} + V_2 \in dz).$$

By (7), defining u_t^q the q-resolvent of the process X killed upon exiting $(0,t]$, we get

$$H^{(q)}(x,t;dz) = q\int_0^z u_t^q(x,dr)P(V_2 \in dz - r),$$

so by Eqs. (6) and (8),

$$H^{(q)}(x,t;dz)/dz = \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t)}C_1^{(q)}(z) - \mathbf{1}_{\{z \ge x\}}C_1^{(q)}(z-x).$$

In conclusion,

$$P_x(\underline{X}_{\mathbf{e}} > 0, \overline{X}_{\mathbf{e}} \le t, X_{\mathbf{e}} + V_2 \in dz)$$
$$= \left(\frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t)}C_1^{(q)}(z) - \mathbf{1}_{\{z \ge x\}}C_1^{(q)}(z-x)\right)dz. \quad (17)$$

Invariance by translation yields the result. □

A.3. Proof of Lemma 6.2

Integrating over z the equality in the previous lemma and applying Eq. (9) yield

$$P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \le t, X_{\mathbf{e}} + V_2 \le z)$$

$$= \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)}(U_1^{(q)}(z-y)-1) - \mathbf{1}_{\{z\geq x\}}(U_1^{(q)}(z-x)-1). \quad (18)$$

Now observe that

$$P_x(C_y) = P_x(T_{(t,+\infty)} < T_y \wedge \mathbf{e}) + P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t, X_{\mathbf{e}} + V_2 > t).$$

For the first term, we get

$$P_x(T_{(t,+\infty)} < T_y \wedge \mathbf{e}) = 1 - P_x(\mathbf{e} < T_y \wedge T_{(t,+\infty)}) - P_x(T_y < T_{(t,+\infty)} \wedge \mathbf{e})$$
$$= 1 - P_x(\mathbf{e} < T_y \wedge T_{(t,+\infty)}) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)},$$

where the last equality is due to (4). For the second term, we have

$$P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t, X_{\mathbf{e}} + V_2 > t) = P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t) - P_x(\underline{X}_{\mathbf{e}} > y, \overline{X}_{\mathbf{e}} \leq t, X_{\mathbf{e}} + V_2 \leq t)$$
$$= P_x(\mathbf{e} < T_y \wedge T_{(t,+\infty)}) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)}(U_1^{(q)}(t-y)-1) + (U_1^{(q)}(t-x)-1),$$

where the last equality follows by applying (18) with $z=t$.

As a conclusion,

$$P_x(C_y) = 1 - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)}(U_1^{(q)}(t-y)-1) + U_1^{(q)}(t-x) - 1$$
$$= -\frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} U_1^{(q)}(t-y) + U_1^{(q)}(t-x),$$

which was the announced result. □

A.4. Proof of Theorem 6.3

Recall that the formula for the transition probability is a direct consequence of Proposition 4.2 and Lemmas 6.1 and 6.2.

The computation of the killing probability can be obtained by two methods. The first method uses the formula in Proposition 4.2. Taking $y=0$ in Lemma 6.2, we get

$$P_x(C_0) = U_1^{(q)}(t-x) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t)} U_1^{(q)}(t).$$

Also by (7),

$$P_x(A_0) = \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t)},$$

which suffices to terminate the computation. The second method uses the fact that $1-k(x)$ is the total mass of the measure $p(x;\cdot)$. Taking $y=0$ in the transition probability, one gets

$$P_x(S \in dz) = \left(C_1^{(q)}(z) \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t)} - \mathbf{1}_{\{z\geq x\}} C_1^{(q)}(z-x) \right) dz. \quad (19)$$

The present alternative proof ends integrating the last density over $[0,t]$ and using (9).

As a last step, we express the distribution of S_1. To compute the law of S_1, observe that either the progenitor of the genealogy is sampled before t, or otherwise, conditional on the lifetime x of this progenitor, S_1 is distributed according to $P_x(S \in \cdot)$. This can be written as follows, integrating over the different possible values of x, greater than t (in which case reflection occurs) or smaller than t:

$$P(S_1 \in dz) = c_2 P(V_2 \in dz) + (c_1 P(V_1 \geq t) + c_2 P(V_2 \geq t)) P_t(S \in dz) + \int_{(0,t)} c_1 P(V_1 \in dr) P_r(S \in dz).$$

From (19), we get, after some algebra,

$$P(S_1 \in dz) = c_2 P(V_2 \in dz) + \frac{C_1^{(q)}(z)}{U_1^{(q)}(t)}(c_1 P(V_1 \geq t) + c_2 P(V_2 \geq t)) + b^{-1} A^{(q)}(t)\, dz - b^{-1} B^{(q)}(z)\, dz,$$

where

$$A^{(q)}(t) := bc_1 \int_{(0,t)} P(V_1 \in dr) U_1^{(q)}(t-r)$$
$$= bc_1 P(V_1 < t) + bc_1 q \int_0^t P(V_1 \in dr) \int_0^{t-r} W_1^{(q)}(t-r-u) P(V_2 \leq u)\, du,$$

and

$$B^{(q)}(z) := bc_1 \int_{(0,z)} P(V_1 \in dr) C_1^{(q)}(z-r) = bc_1 q \int_{(0,z)} P(V_1 \in dr) \times \int_0^{z-r} W_1^{(q)}(z-r-u) P(V_2 \in du).$$

Using the commutativity of the convolution product and Lemma 5.1, and recalling that $q = bc_2 = b(1-c_1)$, we get

$$A^{(q)}(t) = bc_1 P(V_1 < t) + q \int_0^t du\, P(V_2 < u) \int_0^{t-u} W_1^{(q)}(t-r-u) bc_1 P(V_1 \in dr)$$
$$= bc_1 P(V_1 < t) + q \int_0^t du\, P(V_2 < u)(bW_1^{(q)}(t-u) - W_1^{(q)\prime}(t-u))$$
$$= bc_1 P(V_1 < t) + b(U_1^{(q)}(t) - 1) - q \int_0^t du\, P(V_2 < u) W_1^{(q)\prime}(t-u)$$
$$= bc_1 P(V_1 < t) + b(U_1^{(q)}(t) - 1) + bc_2 P(V_2 < t) - q \times \int_0^t P(V_2 \in du) W_1^{(q)}(t-u)$$
$$= b(c_1 P(V_1 < t) + c_2 P(V_2 < t)) + b(U_1^{(q)}(t) - 1) - C_1^{(q)}(t).$$

Similarly,

$$B^{(q)}(z) = q \int_0^z du\, P(V_2 \in du) \int_0^{z-u} W_1^{(q)}(z-r-u) bc_1 P(V_1 \in dr)$$
$$= q \int_0^z P(V_2 \in du)(bW_1^{(q)}(z-u) - W_1^{(q)\prime}(z-u))$$
$$= bC_1^{(q)}(z) - q \int_0^z P(V_2 \in du) W_1^{(q)\prime}(z-u).$$

Substituting the final expressions for $A^{(q)}$ and $B^{(q)}$ into the previous expression for $P(S_1 \in dz)$ finally yields

$$P(S_1 \in dz) = c_2 P(V_2 \in dz) + \left(c_2 \int_0^z P(V_2 \in du) W_1^{(q)\prime}(z-u) - \frac{C_1^{(q)}(z) C_1^{(q)}(t)}{bU_1^{(q)}(t)} \right) dz,$$

which was to be proved.

Finally, the probability p that at least one individual is sampled before time t is given by

$$p = \int_0^t P(S_1 \in dz)$$
$$= c_2 P(V_2 \leq t) + c_2 \int_0^t dz \int_0^z P(V_2 \in du) W_1^{(q)\prime}(z-u) - \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)} \int_0^t C_1^{(q)}(z)\, dz$$
$$= c_2 P(V_2 \leq t) + c_2 \int_0^t P(V_2 \in du)(W_1^{(q)}(t-u) - 1) - \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)}(U_1^{(q)}(t) - 1)$$
$$= c_2 P(V_2 \leq t) + c_2 \int_0^t P(V_2 \in du) W_1^{(q)}(t-u) - c_2 P(V_2 \leq t) - b^{-1} C_1^{(q)}(t) + \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)}$$
$$= \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)},$$

which is the announced result. □

A.5. Proof of Proposition 7.2

First, using the convolution rule for Laplace transforms and then Eq. (16), the Laplace transform (as a function of λ) of $C_1^{(q)}$ is

$$\begin{aligned}\frac{qE(e^{-\lambda V_2})}{\psi_1(\lambda)-q} &= \frac{\int_0^\infty b\mu e^{-\mu r}P(V>r)e^{-\lambda r}\,dr}{\psi_1(\lambda)-q}\\ &= \frac{b\mu}{\lambda+\mu}\frac{1-E(e^{-(\lambda+\mu)V})}{\psi_1(\lambda)-q}\\ &= \frac{\mu}{\lambda+\mu}\frac{\lambda+\mu-\psi(\lambda+\mu)}{\psi_1(\lambda)-q}\\ &= \frac{\mu}{\lambda+\mu}\left(-1+\frac{\lambda}{\psi_1(\lambda)-q}\right).\end{aligned}$$

Now since the first factor in the final product is the Laplace transform of the exponential density with parameter μ and the second factor is the Laplace transform of $W_1^{(q)\prime}$, we get (by the convolution rule) the first proposed expression for $C_1^{(q)}$. The second one follows by an integration by parts. By substituting the first expression for $C_1^{(q)}$ into Eq. (9), one obtains the first expression proposed for $U_1^{(q)}(z)$. The second follows by rearranging terms in the second expression for $C_1^{(q)}$.

Let us now compute the initial distribution of S_1. To this end, we compute an expression for $I(z):=\mu^{-1}c_2\int_0^z P(V_2\in du)W_1^{(q)\prime}(z-u)$. Applying Eq. (16) (the laws of V_1 and V_2), we get

$$\begin{aligned}I(z) &= \int_0^z e^{-\mu u}P(V>u)W_1^{(q)\prime}(z-u)\,du\\ &= [-W_1^{(q)}(z-u)e^{-\mu u}P(V>u)]_0^z - \int_0^z W_1^{(q)}(z-u)(c_2P(V_2\in du)\\ &\quad + c_1P(V_1\in du))\\ &= -\mu^{-1}c_2P(V_2\in dz)/dz + W_1^{(q)}(z) - b^{-1}C_1^{(q)}(z) - b^{-1}(bW_1^{(q)}(z)\\ &\quad - W_1^{(q)\prime}(z)),\end{aligned}$$

where the second equality is an integration by parts and the last one is due to Lemma 5.1 and Eq. (8). Then we get

$$\begin{aligned}\mu I(z) &= c_2\int_0^z P(V_2\in du)W_1^{(q)\prime}(z-u) = -c_2P(V_2\in dz)/dz\\ &\quad + \mu b^{-1}(-C_1^{(q)}(z)+W_1^{(q)\prime}(z)).\end{aligned}$$

Using the general expression for the initial distribution of S_1 in Theorem 6.3, we get

$$\begin{aligned}P(S_1\in dz)/dz &= \mu b^{-1}(-C_1^{(q)}(z)+W_1^{(q)\prime}(z)) - \frac{C_1^{(q)}(z)C_1^{(q)}(t)}{bU_1^{(q)}(t)}\\ &= \mu b^{-1}W_1^{(q)\prime}(z) - \mu b^{-1}C_1^{(q)}(z)\frac{U_1^{(q)}(t)+\mu^{-1}C_1^{(q)}(t)}{U_1^{(q)}(t)}\\ &= \frac{\mu}{b}\left(W_1^{(q)\prime}(z) - \frac{C_1^{(q)}(z)W_1^{(q)}(t)}{U_1^{(q)}(t)}\right),\end{aligned}$$

which ends the proof. □

A.6. Likelihood in the Markovian case

In the Markovian case, individuals die at constant rate δ and are sampled at constant rate μ. In this competing-exponentials case, we have $c_2=\mu/(\mu+\delta)$ and $P(V_2\in dr)=(\mu+\delta)e^{-(\mu+\delta)r}\,dr$. The scale function $W_1^{(q)}$ was already presented in Section 7.2, and we now compute the remaining functions required for the expression of the likelihood. To obtain simple expressions, we note the following useful relationships, where α_1 and α_2 are defined in Section 7.2:

$$\begin{aligned}&\alpha_1\alpha_2 = b\mu\\ &(\alpha_1-\mu-\delta)(\alpha_2+\mu+\delta) = -b\delta\\ &(\alpha_1+b)(\alpha_1-\mu-\delta) = -b\delta\\ &(\alpha_2-b)(\alpha_2+\mu+\delta) = -b\delta\\ &\alpha_1(\alpha_2-b)(\alpha_2+\mu) = -b\delta\mu\\ &\alpha_2(\alpha_1+b)(\alpha_1-\mu) = b\delta\mu\end{aligned}$$

Then, using the definitions of $C_1^{(q)}$ and $U_1^{(q)}$ in Eqs. (8) and (9), and simplifying, we obtain

$$\begin{aligned}W_1^{(q)}(x) &= \frac{\alpha_2-b}{\alpha_1+\alpha_2}e^{-\alpha_1 x}+\frac{\alpha_1+b}{\alpha_1+\alpha_2}e^{\alpha_2 x}\\ W_1^{(q)\prime}(x) &= \frac{b}{\alpha_1+\alpha_2}((\alpha_1-\mu)e^{-\alpha_1 x}+(\alpha_2+\mu)e^{\alpha_2 x})\\ C_1^{(q)}(x) &= \frac{b\mu}{\alpha_1+\alpha_2}(e^{\alpha_2 x}-e^{-\alpha_1 x})\\ C_1^{(q)\prime}(x) &= \frac{b\mu}{\alpha_1+\alpha_2}(\alpha_1e^{-\alpha_1 x}+\alpha_2e^{\alpha_2 x})\\ U_1^{(q)}(x) &= \frac{\alpha_2e^{-\alpha_1 x}+\alpha_1e^{\alpha_2 x}}{\alpha_1+\alpha_2}.\end{aligned}$$

We can now proceed to calculate the factors involved in the likelihood (Corollary 6.4). Substituting the required functions and simplifying, we have

$$\begin{aligned}g(z_1) &= \frac{\mu e^{\alpha_2 z_1}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-z_1)})}{\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)t}}\\ k(z_n) &= \frac{e^{\alpha_1 z_n}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-z_n)})}{\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)t}}\quad p=\frac{\mu(e^{(\alpha_1+\alpha_2)t}-1)}{\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)t}}\\ &f(z_{i-1};y_i,z_i)\\ &= \frac{b\mu e^{\alpha_1(z_{i-1}-y_i)}e^{\alpha_2(z_i-y_i)}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-z_{i-1})})(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-z_i)})}{(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-y_i)})^2}.\end{aligned}$$

For direct comparison to the likelihood derived previously for the Markovian case (Stadler et al., 2012), we consider the likelihood given the time of observation (t) but not conditioned on sampling, which we denote as $\mathcal{L}(\mathcal{T})$. Substituting the above factors and simplifying, we have

$$\begin{aligned}\mathcal{L}(\mathcal{T}) &= g(z_1)k(z_n)\prod_{i=2}^n f(z_{i-1};y_i,z_i)\\ &= b^{n-1}\mu^n\frac{1}{e^{-(\alpha_1+\alpha_2)t}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)t})^2}\\ &\quad\times\frac{\prod_{i=1}^n e^{-(\alpha_1+\alpha_2)(t-z_i)}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-z_i)})^2}{\prod_{i=2}^n e^{-(\alpha_1+\alpha_2)(t-y_i)}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)(t-y_i)})^2}.\end{aligned}\tag{20}$$

On the other hand, the likelihood was previously derived (Stadler et al., 2012) as the following, adjusted to match present notation:

$$\mathcal{L}(\mathcal{T}) = b^{n-1}\mu^n\frac{1}{q(t)}\frac{\prod_{i=1}^n q(t-z_i)}{\prod_{i=2}^n q(t-y_i)}\tag{21}$$

with the definitions

$$\begin{aligned}&q(x) = 2(1-\gamma_2^2)+e^{-\gamma_1 x}(1-\gamma_2)^2+e^{\gamma_1 x}(1+\gamma_2)^2,\\ &\gamma_1=\sqrt{(b-\delta-\mu)^2+4b\mu},\quad \gamma_2=-\frac{b-\delta-\mu}{\gamma_1}.\end{aligned}$$

Note that $\gamma_1=\alpha_1+\alpha_2$ and $\gamma_2=(\alpha_1-\alpha_2)/(\alpha_1+\alpha_2)$. We can thus rewrite

$$q(x) = \frac{4e^{-(\alpha_1+\alpha_2)x}}{(\alpha_1+\alpha_2)^2}(\alpha_2+\alpha_1e^{(\alpha_1+\alpha_2)x})^2.$$

Cancelling the constant factors in $q(\cdot)$, it immediately follows that Eqs. (20) and (21) precisely agree.

References

Aldous, D., Popovic, L., 2005. A critical branching process model for biodiversity. Adv. Appl. Probab. 37 (4), 1094–1115.
Bertoin, J., 1996. Lévy processes. In: Cambridge Tracts in Mathematics, vol.e 121.
Bertoin, J., 1997. Exponential decay and ergodicity of completely asymmetric Lévy processes in a finite interval. Ann. Appl. Probab., 156–169

Collaborative Group on AIDS Incubation and HIV Survival, 2000. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Lancet 355, 1131–1137.
de Silva, E., Ferguson, N.M., Fraser, C., 2012. Inferring pandemic growth rates from sequence data. J. R. Soc. Interface 9, 1797–1808.
Dearlove, B., Wilson, D.J., 2013. Coalescent inference for infectious disease: meta-analysis of hepatitis C. Philos. Trans. R. Soc. B 368, 20120314.
Drummond, A., Nicholls, G., Rodrigo, A., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.
Etienne, R., Haegeman, B., Stadler, T., Aze, T., Pearson, P., Purvis, A., Phillimore, A., 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. Lond. B 279 (1732), 1300–1309.
Geiger, J., Kersting, G., 1997. Depth–first search of random trees, and Poisson point processes in classical and modern branching processes (Minneapolis, 1994). IMA Math. Appl. 84.
Lambert, A., 2009. The allelic partition for coalescent point processes. Markov Process. Relat. Fields 15, 359–386.
Lambert, A., 2010. The contour of splitting trees is a Lévy process. Ann. Probab. 38 (1), 348–395.
Lambert, A., Stadler, T., 2013. Birth-death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90, 113–128.
Lambert, A., Trapman, P., 2013. Splitting trees stopped when the first clock rings and Vervaat's transformation. J. Appl. Probab. 50 (1), 208–227.
Morlon, H., Parsons, T., Plotkin, J., 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA 108 (39), 16327–16332.
Nee, S.C., May, R.M., Harvey, P., 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B 344, 305–311.
Pomeroy, L., Bjørnstad, O., Holmes, E., 2008. The evolutionary and epidemiological dynamics of the paramyxoviridae. J. Mol. Evol. 66, 98–106.
Pybus, O., Charleston, M., Gupta, S., Rambaut, A., Holmes, E., Harvey, P., 2001. The epidemic behavior of the hepatitis C virus. Science 292, 2323–2325.
Stadler, T., 2010. Sampling-through-time in birth-death trees. J. Theor. Biol. 267 (3), 396–404.
Stadler, T., 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. USA 108 (15), 6187–6192.
Stadler, T., Kouyos, R.D., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H.F., Drummond, A., Bonhoeffer, S., 2012. and the Swiss HIV Cohort Study, 2012. Estimating the basic reproductive number from viral sequence data. Mol. Biol. Evol. 29, 347–357.
Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A., 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. USA 110 (1), 228–233.