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

By Amaury Lambert *,1,2, Helen K. Alexander 3, and Tanja Stadler 4

* Corresponding author

1 UPMC Univ Paris 06
2 Laboratoire de Probabilités et Modèles Aléatoires CNRS UMR 7599
3 Collège de France
4 Center for Interdisciplinary Research in Biology CNRS UMR 7241

Paris, France

Phone number: +33 1 44 27 13 91
E-mail: amaury.lambert@upmc.fr
URL: http://www.proba.jussieu.fr/pageperso/amaury/index.htm

Institute of Integrative Biology
ETH Zürich
Universitätsstrasse 16
8092 Zürich
Switzerland
E-mail: helen.alexander@env.ethz.ch
URL: http://www.tb.ethz.ch/people/helenal

Institute of Integrative Biology
ETH Zürich
Universitätsstrasse 16
8092 Zürich
Switzerland
E-mail: tanja.stadler@env.ethz.ch
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.

Here we introduce an epidemiological model where infectives leave the epidemic, either by 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, namely an application to influenza and an application to HIV. 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.

Running head. Phylogenies with age-dependent death and sampling.

Key words and phrases. Branching process; birth–death process; contour process; coalescent point process; Lévy process; scale function; epidemiology; influenza; HIV.

Word count. 6000 words approximately, including supporting information (appendix).

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 [11]. As another example, viral
genetic sequences extracted from patient samples can provide information on the rate at which an infectious disease transmits in the host population [14].

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 were limited to death rates of individuals being independent of the age of an individual (see e.g. [11, 10, 13, 4] for species phylogenies and [14, 15] for virus phylogenies).

For phylogenetic trees in which all tips are sampled at one point in time, e.g. extant species phylogenies, Lambert [7] introduced a framework to calculate the likelihood of a phylogenetic tree accounting for general lifetime distributions (see also [8]). 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 transmission events and “deaths” represent 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 Center for Disease Control (http://www.cdc.gov/flu/about/disease/spread.htm). Extending the model to age-dependent removal will allow quantifying the death/recovery dynamics more accurately based on genetic sequencing data, 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 rapidly simulating phylogenies under age-dependent death/recovery rates even for huge epidemic outbreaks, thus allowing 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 giving rise to the phylogeny is non-Markovian due to 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 only
depending 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 the application of the general mathematical results:
the influenza model, where the lifetime of individuals is not dependent upon whether they leave
the epidemic by recovery or sampling; and the HIV model, where sampling occurs after some
exponential time during the (independently distributed) infectious period.

We conclude the paper by discussing future challenges in putting the theoretical framework
into a computational inference tool, such that the model can be used to analyze pathogen genetic
sequence data collected during an epidemic.

2 Model of infection and sampling

We model by a (possibly) non-Markovian branching process the dynamics of a population of
infectives. The process is assumed to start with one infected individual and has an overall time
duration of $t$.

A “birth” event is interpreted as the infection of a susceptible individual, where suscepti-
bles are supposed to be in excess, so that individuals can be assumed to give birth independently
(branching property) and at constant rate, say $b$. The branching property means in particular
that the population of infectives is, on average, exponentially growing or declining (or constant).
The new infective is assumed to be infectious immediately after infection.

The “death” of an individual is the removal of an individual from the infective population.
Individuals may become non-infectious because of actual death, recovery, successful treatment,
or behaviour changes. Upon removal, individuals may be sampled (type 2; i.e. included into the
phylogeny) or may not be sampled (type 1).

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 from other individuals.

Individuals of type 1 live a duration distributed as $V_1$ after which they are removed by becoming non-infectious. Individuals of type 2 live a duration distributed as $V_2$ after which they are simultaneously sampled and removed, meaning each death of an individual of type 2 coincides with a sampling event.

These assumptions are consistent with the natural framework where, for an individual who was infected a time units ago (i.e., with ‘age’ $a$), recovery occurs with the instantaneous rate $\rho_1(a)$ and sampling occurs with the rate $\rho_2(a)$, independently. This is 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 $\rho$, where time is reset at birth), and that upon leaving the epidemic at age $a$, they leave it by recovery with probability $\rho_1(a)/\rho(a)$ and by sampling with probability $\rho_2(a)/\rho(a)$. This is exactly the same framework as described above, 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 (influenza model), $V_1$ and $V_2$ are identically distributed, meaning the duration of infectiousness does not depend on being sampled. In the second case (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 $\mu$, meaning individuals are sampled with a constant rate $\mu$ 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 Figure 1. Splitting trees [5, 6, 7, 9] 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.
Here, the law of our two-type splitting tree is characterized by the knowledge of the two lifespan measures $\pi_1 := bc_1 P(V_1 \in \cdot)$ 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 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 Figure 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 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 Figure 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 pairs of coalescence and sampling times.

3 The Contour Process

In [7], 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. Figure 2 shows the contour process associated to the tree in Figure 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 transmission being a Poisson process, we can have the ball slip backward in time until transmission, rather than forward in time until transmission. At trans-
mission, 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 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 [7, 9]. 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 \( \pi_1 \) compensated at rate \(-1\). In the absence of sampled
individuals, we have shown [7, 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.

Now when we additionally consider sampled individuals, 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 Figure 2, we will call these jumps the
marked jumps of the contour process. It is straightforward that this new process is just the
compound Poisson process with jump measure \( \pi \) compensated at rate \(-1\). However, we stick to
the previous two-type description in order to keep track of births of type 2 individuals.
4 The 2D coalescent point process

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 (only) time at which this individual is 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 $i$).

Our first remark is that the pairs $(R_i, S_i)$ characterize the (plane) sampled tree, as seen in Figure 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 two-dimensional coalescent point process [1, 7, 8].

Straightforward consequences of the definition of the JCCP 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 (which can be seen as the jump size of a marked jump at exploration time 0).

Now by the Markov property of the contour process, the pairs $(R_i, S_i)$ form a killed Markov chain, where the transition probability only depends on the second component, thanks to the fact that $S_i$ is the new starting point of the marked contour process. A killed Markov chain is a Markov chain with a possibly finite (random) lifetime. More specifically, the transition kernel $p(x, \cdot)$ of a killed Markov chain $X$ 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 $X_n = x$, 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 now characterize the transitions of this Markov chain in terms of the contour process. To get rid of the reflection at $t$, we apportion the path of the contour process into all subpaths
terminating as soon as a marked jump appears or as the path exits \((0, t]\). We classify subpaths according to the four following events:

- \(A\) – exit from the bottom of \((0, t]\), i.e. hit 0, before the first marked jump;
- \(B\) – arrival of a marked jump with terminal value in \((0, t]\) before exit of \((0, t]\);
- \(C\) – exit from the top of \((0, t]\), i.e. overshoot \(t\), strictly before the next marked jump;
- \(C'\) – exit from the top of \((0, t]\) at the next marked jump.

Notice that the events \(A, B, C, C'\) form a partition. The path on Figure 2 is apportioned into 7 subpaths delineated by times \(0 < u_1 < \cdots < u_7\), which are respectively of types \(B, CC', B, CC', B, CC', A\) (where \(CC'\) means: \(C \cup C'\)). Generally speaking, the last subpath, and the last subpath only, is of type \(A\), and by the Markov property, there is a geometric number of subpaths of type \(CC'\) between two consecutive marked jumps, each corresponding to subpaths of type \(B\). Our objective is now to compute the joint law of the infimum of this concatenation of subpaths of type \(CC'\) and of the terminal value of the concluding subpath of type \(B\).

Recall that the first marked jump occurs after an exponential random variable that we denote by \(e\), with rate parameter

\[ q := bc_2, \]

at which time the contour process has a jump distributed as \(V_2\). Now recall that \(X\) denotes the contour process in the absence of sampled individuals. Throughout the paper, \(V_2\) is assumed independent of \(e\) and \(X\). We denote by \(T_A\) the first hitting time of the set \(A\) by \(X\), and by \(T := T_0 \land T_{(t, +\infty)}\) the first exit time of \((0, t]\) by \(X\), where we use the notation \(a \land b = \min(a, b)\). We further denote by \(\overline{X}\) and \(\underline{X}\), respectively, the infimum and supremum processes of \(X\), that is,

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

We now express the events \(A, B, C, C'\) using the preceding notation. By the strong Markov property of \(X\), it is sufficient to characterize each of these events in terms of one single path of \(X\):

\[ A = \{T_0 < T_{(t, +\infty)} \land e\}, \]
\[ B = \{e < T, X_e + V_2 \leq t\}, \]
\[ C = \{T_{(t, +\infty)} < T_0 \land e\}, \]
\[ C' = \{ e < T, X_e + V_2 > t \} = \{ X_e > 0, X_e \leq t, X_e + V_2 > t \}. \]

On \( C \cup C' \), it will sometimes be useful to call \( \tau \) the first time at which the contour process is reflected.

Recall that the pairs \((R_i, S_i)\) form a Markov chain, where the transition probability only depends on the second component, which also is the new starting point of the contour. For this reason, we will define the pair \((R, S)\) by

\[
P_x(R \in dy, S \in dz) := P(R_2 \in dy, S_2 \in dz | S_1 = x) =: p(x; dy\, dz).
\]

At each step \( i \), conditional on \( S_i = x \), the Markov chain can be killed with probability

\[
k(x) = 1 - \int_{[0,x]} \int_{[y,t]} p(x; dy\, dz).
\]

Now we decompose the contour process into its excursions (subpaths) below \( t \) until the first excursion, say \( \epsilon \), hitting 0 (killing, type A) or possessing a marked jump (sampling, type B). In particular,

- \( S \in dz \) if the first excursion of type \( A \cup B \) is actually of type \( B \) and its marked jump ends in \( dz \);
- \( R > y \) if the infimum of the contour process until \( \epsilon \) is larger than \( y \), where the infimum has to be taken over the geometrically distributed number of excursions of the process (type \( C \cup C' \)) preceding \( \epsilon \);
- \( \epsilon \) can either be the very first excursion (starting from \( x \)) or any other excursion (starting from \( t \)).

Therefore, we have

\[
P_x(R > y, S \in dz) = P_x(X_e > y, X_e \leq t, X_e + V_2 \in dz)
+ P_x(X_e > y, C \cup C') \sum_{n \geq 0} (P_t(X_e > y, C \cup C'))^n P_t(X_e > y, X_e \leq t, X_e + V_2 \in dz).
\]

With the same line of reasoning,

\[
k(x) = P_x(A) + P_x(C \cup C') \sum_{n \geq 0} (P_t(C \cup C'))^n P_t(A)
\]
Rewriting the summations, we arrive at the following statement.

**Proposition 4.1.** Let \( x \in (0, t] \), \( y \in [0, x) \) and \( z \in (y, t) \). Then

\[
P_x(R > y, S \in dz) = P_x(\overline{X}_e > y, \overline{X}_e \leq t, X_e + V_2 \in dz)
\]

\[
+ P_x(\overline{X}_e > y, C \cup C') \frac{P_l(X_e > y, \overline{X}_e \leq t, X_e + V_2 \in dz)}{1 - P_l(\overline{X}_e > y, C \cup C')},
\]

and

\[
k(x) = P_x(A) + P_x(C \cup C') P_l(A \mid A \cup B).
\]

We will 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 [2, 3, 9]. We state these results in terms of an arbitrary compound Poisson process \( Y \) with jump measure \( \pi \) 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 \( T_A \) of \( A \) and extremum processes \( \overline{Y} \) and \( \underline{Y} \)). It can be convenient to characterize the law of this process by its Laplace exponent \( \psi \) defined by

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

The function \( \psi \) is differentiable and convex and we denote by \( \eta \) its largest root. Then \( \psi \) is increasing on \([\eta, +\infty)\) and we denote by \( \phi \) its inverse on this set, so that \( \phi \) 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. [2]), in the form

\[
P_x(T_0 < T_{(t, +\infty)}) = \frac{W(t - x)}{W(t)},
\]

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

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

Equation (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.

For any $q > 0$,

$$E_x \left( e^{-qT} 1_{T_0 < T_{(t, +\infty)}} \right) = \frac{W(q)(t-x)}{W(q)(t)},$$  \hspace{1cm} (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

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

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

Observe that by the Fubini–Tonelli Theorem

$$q u^q_t(x, z) dz = E_x \left( \int_{s=0}^T 1_{e \in ds} 1_{Y_s \in dz} \right) = P_x (e < T, Y_e \in dz) = P_x (Y_e > 0, \gamma_e \leq t, Y_e \in dz),$$  \hspace{1cm} (7)

where $e$ denotes an independent exponential random variable with parameter $q$. The previous formula is key to computing the probabilities involved in Proposition 4.1 (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 probabilities displayed in Proposition...
Let $\psi_1$ 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 $\psi_1$ 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

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

and

$$U_1(q)(z) := 1 + \int_0^z C_1(q)(x) dx = 1 + q \int_0^z W_1(q)(z - u) P(V_2 \leq u) du. \quad (9)$$

The last equality comes from an application of Fubini–Tonelli theorem and a change of variable. Notice in particular that $U_1(q) = C_1(q)$. 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

$$P_z(X_e > y, X_e \leq t, X_e + V_2 \in dz) = \left(\frac{W_1(q)(t - x)}{W_1(q)(t - y)} C_1(q)(z - y) - 1_{\{z \geq x\}} C_1(q)(z - x)\right) dz.$$

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

$$P_x(X_e > y, C \cup C') = U_1(q)(t - x) - \frac{W_1(q)(t - x)}{W_1(q)(t - y)} U_1(q)(t - y).$$

We can now state the main result of this article.

**Theorem 6.3.** The sequence $S_1, (R_2, S_2), (R_3, S_3), \ldots$ 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

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

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

$$P_x(R > y, S \in dz) = \left(C_1(q)(z - y) \frac{U_1(q)(t - x)}{U_1(q)(t - y)} - 1_{\{z \geq x\}} C_1(q)(z - x)\right) dz,$$
and the killing probability is
\[ k(x) = \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t)}. \] (10)

The probability \( p \) that at least one individual is sampled before time \( t \) (i.e., the sequence is not empty), is given by
\[ p = \int_0^t P(S_1 \in dz) = \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)}. \] (11)

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

The formula for the transition probability is a direct consequence, by elementary calculus, of Proposition 4.1 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
\[ C_1^{(q)'}(z) = q g_2(z) + q \int_0^z W_1^{(q)'}(z-u) g_2(u) du, \] (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
\[ g(z) = b^{-1} \left( C_1^{(q)'}(z) - \frac{C_1^{(q)}(z)C_1^{(q)}(t)}{U_1^{(q)}(t)} \right). \] (13)

The second consequence is that the transition probability has density, say \( f \),
\[ P_x(R \in dy, S \in dz) = p(x; dydz) =: f(x; y, z) dydz, \]
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 \( U_1^{(q)'} = C_1^{(q)} \), we get
\[ f(x; y, z) = \frac{U_1^{(q)}(t-x)}{U_1^{(q)}(t-y)} \left[ C_1^{(q)'}(z-y) - C_1^{(q)}(z-y) \frac{C_1^{(q)}(t-y)}{U_1^{(q)}(t-y)} \right]. \] (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 $(y_i)_{2 \leq i \leq n}$ and sampling times $(z_i)_{1 \leq i \leq n}$, 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 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. 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, respectively. We apply our mathematical results to these cases, by specifically deriving the expressions required for the likelihood.

7.1 Influenza

In the case of influenza, we assume that, after a random amount of time, an infective either recovers without sampling, with a certain probability $c_1$ which does not depend on the time elapsed, or with probability $c_2$, recovers with sampling, which typically happens for the severe cases in the hospital. Thus we assume that $V_1$ and $V_2$ are equal in distribution. 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 Equations (12), (13), (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 the case of HIV, each individual has a “natural” infectious lifetime, denoted $V$, having an arbitrary distribution. This lifetime would apply if there were no intervention. However, individuals are additionally sampled after some independent exponential duration, say $e'$, with rate parameter $\mu$. Individuals are removed from the infectious class upon sampling, due e.g. to successful treatment or behavior change concomitant with the intervention. We note that a model with constant rates of both “natural death” and sampling, as in [14], is a special case of this model; see the computations for the Markovian case at the end of this section.

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

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

which we can rewrite as:

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

Furthermore,

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

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

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

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

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

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

and the initial distribution of \( S_1 \) is given by:

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

We make further computations in the Markovian case, that is, when the “natural” lifetime of individuals ends at constant rate \( d \). Then \( \pi(dr) = bde^{-dr} \, dr \) and

\[ \psi(\lambda + \mu) - \mu = \frac{Q(\lambda)}{\lambda + \mu + d}, \]

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

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

where \( \alpha_1 \) and \( \alpha_2 \) are both positive. Using \( \alpha_2 - \alpha_1 = b - d - \mu \), 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 leads to the same expression for the likelihood as derived previously using methods particular to this Markovian case [14].

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 two-dimensional coalescent point process 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 [12, 14]. A special section is dedicated to this case in the Appendix.

We envision to use 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 maximizing the probability 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 orien-
tations 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 phy-
logenetic 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 sam-
ples. 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.
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 roll back 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. AL was financially supported by grant MANEGE ‘Modèles Aléatoires en Écologie, Génétique et Évolution’ 09-BLAN-0215 of ANR (French national research agency). AL also thanks the Center for Interdisciplinary Research in Biology (Collège de France) for funding. HA received support from ETH Zürich. TS thanks the Swiss National Science foundation for funding (SNF grant #PZ00P3 136820).

References

[1] D. Aldous and L. Popovic. A critical branching process model for biodiversity. *Advances in Applied Probability*, 37(4):1094–1115, 2005.

[2] J. Bertoin. Lévy Processes, volume 121 of Cambridge Tracts in Mathematics, 1996.

[3] J. Bertoin. Exponential decay and ergodicity of completely asymmetric Lévy processes in a finite interval. *The Annals of Applied Probability*, pages 156–169, 1997.

[4] R. Etienne, B. Haegeman, T. Stadler, T. Aze, P. Pearson, A. Purvis, and A. Phillimore. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society London B*, 279(1732):1300–1309, 2012.

[5] J. Geiger and G. Kersting. Depth-first search of random trees, and Poisson point processes in classical and modern branching processes (Minneapolis, 1994) IMA Math. Appl. Vol. 84, 1997.
[6] A. Lambert. The allelic partition for coalescent point processes. *Markov Processes and Related Fields*, 15:359–386, 2009.

[7] A. Lambert. The contour of splitting trees is a Lévy process. *The Annals of Probability*, 38(1):348–395, 2010.

[8] A. Lambert and T. Stadler. Macro-evolutionary models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Under review*, 2013.

[9] A. Lambert and P. Trapman. Splitting trees stopped when the first clock rings and Vervaat’s transformation. *Journal of Applied Probability, in press*, 2013.

[10] H. Morlon, T. Parsons, and J. Plotkin. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences USA*, 108(39):16327–16332, 2011.

[11] S. C. Nee, R. M. May, and P. Harvey. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society London B*, 344:305–311, 1994.

[12] T. Stadler. Sampling-through-time in birth-death trees. *J. Theo. Biol.*, 267(3):396–404, 2010.

[13] T. Stadler. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences USA*, 108(15):6187–6192, 2011.

[14] T. Stadler, R. D. Kouyos, V. von Wyl, S. Yerly, J. Böni, P. Bürgisser, T. Klimkait, B. Joos, P. Rieder, D. Xie, H. F. Günthard, A. Drummond, S. Bonhoeffer, and the Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution*, 29:347–357, 2012.

[15] T. Stadler, D. Kühnert, S. Bonhoeffer, and A. Drummond. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences USA*, 2012.
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

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

where we used successively the facts that the Laplace transform of $W^q(z)$ 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 $z \mapsto (q + b)W^q(z)$. \hfill $\Box$

A.2 Proof of Lemma 6.1

Set

$$H^q(x,t;dz) := P_x(X_e > 0, X_e \leq t, X_e + V_2 \in dz).$$

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

$$H^q(x,t;dz) = q \int_0^t u^q_t(x,dr) P(V_2 \in dz - r)$$

so by Equations (6) and (8),

$$H^q(x,t;dz)/dz = \frac{W^q_1(t - x)}{W^q_1(t)} C^q_1(z) - 1_{\{z \geq x\}} C^q_1(z - x),$$

In conclusion,

$$P_x(X_e > 0, X_e \leq t, X_e + V_2 \in dz) = \left(\frac{W^q_1(t - x)}{W^q_1(t)} C^q_1(z) - 1_{\{z \geq x\}} C^q_1(z - x)\right) dz. \quad (17)$$

Invariance by translation yields the result.
A.3 Proof of Lemma 6.2

Integrating over $z$ the equality in the previous lemma and applying Equation (9) yields

$$P_x(X > y, X + V_2 \leq z) = W_1^{(q)}(t-x) \left(U_1^{(q)}(z-y) - 1\right) - 1_{\{z \geq x\}} \left(U_1^{(q)}(z-x) - 1\right).$$

(18)

Noting that $\tau \equiv e$ on $C'$, we deduce

$$P_x(X > y, C') = 1 - P_x(X > y, X + V_2 > t)$$

$$= P_x(X > y, X \leq t) - P_x(X > y, X \leq t, X + V_2 \leq t)$$

$$= P_x(e < T_y \land T_{(t,+)}) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} \left(U_1^{(q)}(t-y) - 1\right) + \left(U_1^{(q)}(t-x) - 1\right).$$

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

$$P_x(X > y, C) = P_x(T_{(t,+)} < T_y \land e)$$

$$= 1 - P_x(e < T_y \land T_{(t,+)}) - P_x(T_y < T_{(t,+)}) \land e$$

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

where the last equality is due to (18).

Since $C$ and $C'$ are mutually exclusive, we can sum the last two sets of equations to obtain

$$P_x(X > y, C \cup C') = 1 - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t-y)} \left(U_1^{(q)}(t-y) - 1\right) + \left(U_1^{(q)}(t-x) - 1\right)$$

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

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 4.1 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 4.1. Taking $y = 0$ in Lemma 6.2, we get

$$P_x(C \cup C') = U_1^{(q)}(t-x) - \frac{W_1^{(q)}(t-x)}{W_1^{(q)}(t)} U_1^{(q)}(t).$$
Also $P_x(A \cup B) + P_x(C \cup C') = 1$ and by (7),

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

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

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

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,

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

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

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

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

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

$$= bc_1 P(V_1 < t) + q \int_0^t du P(V_2 < u)(bW_1^{(q)}(t - u) - W_1^{(q)'}(t - u))$$

$$= bc_1 P(V_1 < t) + b(U_1^{(q)}(t) - 1) - q \int_0^t du P(V_2 < u)W_1^{(q)'}(t - u)$$

$$= bc_1 P(V_1 < t) + b(U_1^{(q)}(t) - 1) + bc_2 P(V_2 < t) - q \int_0^t P(V_2 \in du)W_1^{(q)}(t - u)$$

$$= b(c_1 P(V_1 < t) + c_2 P(V_2 < t)) + b(U_1^{(q)}(t) - 1) - C_1^{(q)}(t).$$

Similarly,

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

$$= q \int_0^z P(V_2 \in du) (bW_1^{(q)}(z - u) - W_1^{(q)'}(z - u))$$

$$= bC_1^{(q)}(z) - q \int_0^z P(V_2 \in du) W_1^{(q)'}(z - u).$$

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

$$P(S_1 \in dz) = c_2 P(V_2 \in dz) + \left( c_2 \int_0^z P(V_2 \in du) W_1^{(q)'}(z - u) - \frac{C_1^{(q)}(z)C_1^{(q)}(t)}{bU_1^{(q)}(t)} \right) dz,$$

which was to be proved.

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

$$p = \int_0^t P(S_1 \in dz)$$

$$= c_2 P(V_2 \leq t) + c_2 \int_0^t dz \int_0^z P(V_2 \in du) W_1^{(q)'}(z - u) - \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)} \int_0^t C_1^{(q)}(z) dz$$

$$= c_2 P(V_2 \leq t) + c_2 \int_0^t P(V_2 \in du) \left( W_1^{(q)}(t - u) - 1 \right) - \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)} \left( U_1^{(q)}(t) - 1 \right)$$

$$= c_2 P(V_2 \leq t) + c_2 \int_0^t P(V_2 \in du) W_1^{(q)}(t - u) - c_2 P(V_2 \leq t) - b^{-1}C_1^{(q)}(t) + \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)}$$

$$= \frac{C_1^{(q)}(t)}{bU_1^{(q)}(t)},$$

which is the announced result.
A.5 Proof of Proposition 7.2

First, using the convolution rule for Laplace transforms and then Equation (16), the Laplace transform (as a function of $\lambda$) of $C_1(q)$ is

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

Now since the first factor in the final product is the Laplace transform of the exponential density with parameter $\mu$ and the second factor is the Laplace transform of $W_1(q)'$, 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 Equation (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 $I(z) := \mu^{-1} c_2 \int_0^z P(V_2 \in du) W_1(q)'(z - u)$. Applying Equation (16) (the laws of $V_1$ and $V_2$), we get

$$
I(z) = \int_0^z e^{-\mu u} P(V > u)W_1(q)'(z - u) du \\
= [-W_1(q)(z - u)e^{-\mu u} P(V > u)]_0^z - \int_0^z W_1(q)(z - u)(c_2 P(V_2 \in du) + c_1 P(V_1 \in du)) \\
= -\mu^{-1}c_2 P(V_2 \in dz)/dz + W_1(q)(z) - b^{-1}C_1(q)(z) - b^{-1}(bW_1(q)(z) - W_1(q)'(z)),
$$

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

$$
\mu I(z) = c_2 \int_0^z P(V_2 \in du) W_1(q)'(z - u) = -c_2 P(V_2 \in dz)/dz + \mu b^{-1}(-C_1(q)(z) + W_1(q)'(z)).
$$
Using the general expression for the initial distribution of $S_1$ in Theorem 6.3, we get

$$P(S_1 \in dz) = \mu b^{-1}(-C_1^{(q)}(z) + W_1^{(q)'r}(z)) - \frac{C_1^{(q)}(z)C_1^{(q)}(t)}{bU_1^{(q)}(t)}$$

$$= \mu b^{-1}W_1^{(q)'r}(z) - \mu b^{-1}C_1^{(q)}(z) \frac{U_1^{(q)}(t) + \mu^{-1}C_1^{(q)}(t)}{U_1^{(q)}(t)}$$

$$= \mu \left(W_1^{(q)'r}(z) - \frac{C_1^{(q)}(z)W_1^{(q)'r}(t)}{U_1^{(q)}(t)} \right),$$

which ends the proof.

---

### A.6 Likelihood in the Markovian case

In the Markovian case, individuals die at constant rate $d$ and are sampled at constant rate $\mu$. In this competing-exponentials case, we have $c_2 = \frac{\mu}{\mu + d}$ and $P(V_2 \in dr) = (\mu + d)e^{-(\mu + d)r}dr$. 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 $\alpha_1$ and $\alpha_2$ are defined in Section 7.2.

$$\alpha_1 \alpha_2 = b\mu$$

$$\alpha_1 - \mu - d)(\alpha_2 + \mu + d) = -bd$$

$$\alpha_1 + b)(\alpha_1 - \mu - d) = -bd$$

$$\alpha_2 - b)(\alpha_2 + \mu + d) = -bd$$

$$\alpha_1(\alpha_2 - b)(\alpha_2 + \mu) = -bd\mu$$

$$\alpha_2(\alpha_1 + b)(\alpha_1 - \mu) = bd\mu$$
Then, using the definitions of $C_1^{(q)}$ and $U_1^{(q)}$ in Equations (8) and (9), and simplifying, we obtain in summary:

\[
W_1^{(q)}(x) = \frac{\omega_2 - b}{\omega_1 + \omega_2} e^{-\omega_1 x} + \frac{\omega_1 + b}{\omega_1 + \omega_2} e^{\omega_2 x}
\]

\[
W_1^{(q)}(x) = \frac{b}{\omega_1 + \omega_2} \left( (\omega_1 - \mu)e^{-\omega_1 x} + (\omega_2 + \mu)e^{\omega_2 x} \right)
\]

\[
C_1^{(q)}(x) = \frac{b\mu}{\omega_1 + \omega_2} \left( e^{\omega_2 x} - e^{-\omega_1 x} \right)
\]

\[
C_1^{(q)}(x) = \frac{b\mu}{\omega_1 + \omega_2} \left( \alpha_1 e^{-\omega_1 x} + \alpha_2 e^{\omega_2 x} \right)
\]

\[
U_1^{(q)}(x) = \frac{\alpha_2 e^{-\omega_1 x} + \alpha_1 e^{\omega_2 x}}{\omega_1 + \omega_2}
\]

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

\[
g(z_1) = \frac{\mu e^{\omega_2 z_1} \left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-z_1)} \right)}{\omega_2 + \omega_1 e^{(\omega_1 + \omega_2)t}}
\]

\[
k(z_n) = \frac{e^{\omega_1 z_n} \left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-z_n)} \right)}{\omega_2 + \omega_1 e^{(\omega_1 + \omega_2)t}}
\]

\[
p = \frac{\mu (e^{(\omega_1 + \omega_2)t} - 1)}{\omega_2 + \omega_1 e^{(\omega_1 + \omega_2)t}}
\]

\[
f(z_{i-1}; y_i, z_i) = \frac{\mu e^{\omega_1 (z_{i-1} - y_i)} e^{\omega_2 (z_i - y_i)} \left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-z_{i-1})} \right) \left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-z_i)} \right)}{\left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-y_i)} \right)^2}
\]

For direct comparison to the likelihood derived previously for the Markovian case [14], we consider the likelihood given the time of observation $(t)$ but not conditioned on sampling, which we denote $L(T)$. Substituting the above factors and simplifying, we have

\[
L(T) = g(z_1)k(z_n) \prod_{i=2}^{n} f(z_{i-1}; y_i, z_i)
\]

\[
= b^{n-1} \mu^n \frac{1}{e^{-(\omega_1 + \omega_2)t} \left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)t} \right)^2} \prod_{i=1}^{n} e^{-\left( \omega_1 + \omega_2 \right)(t-z_i)} \frac{\left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-z_i)} \right)^2}{\left( \omega_2 + \omega_1 e^{(\omega_1 + \omega_2)(t-y_i)} \right)^2}
\]

\[
(20)
\]

On the other hand, the likelihood was previously derived [14] as the following, adjusted to match present notation:

\[
L(T) = b^{n-1} \mu^n \frac{1}{q(t)} \prod_{i=1}^{n} q(t-z_i) \prod_{i=2}^{n} q(t-y_i)
\]

\[
(21)
\]
with the definitions

\[ q(x) = 2(1 - \gamma_2^2) + e^{-\gamma_1 x}(1 - \gamma_2)^2 + e^{\gamma_1 x}(1 + \gamma_2)^2, \]
\[ \gamma_1 = \sqrt{(b - d - \mu)^2 + 4b\mu}, \quad \gamma_2 = \frac{b - d - \mu}{\gamma_1} \]

Note that \( \gamma_1 = \alpha_1 + \alpha_2 \) and \( \gamma_2 = \frac{\alpha_1 - \alpha_2}{\alpha_1 + \alpha_2} \). We can thus rewrite,

\[ q(x) = 4e^{-\frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2}x} \left( \alpha_2 + \alpha_1 e^{\alpha_1 + \alpha_2}x \right)^2 \]

Cancelling the constant factors in \( q(\cdot) \), it immediately follows that Equations \([20]\) and \([21]\) precisely agree.
Figure 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).
Figure 2: The marked contour process, with jumps in solid line, which is associated to the marked tree of Figure 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 $u_7$. 