A characterisation of the reconstructed birth-death process through time rescaling

Anastasia Ignatieva [*†‡], Jotun Hein [†¶], Paul A. Jenkins [‡§¶]

May 6, 2020

Abstract

The dynamics of a population exhibiting exponential growth can be modelled as a birth-death process, which naturally captures the stochastic variation in population size over time. In this article, we consider a supercritical birth-death process, started at a random time in the past, and conditioned to have n sampled individuals at the present. The genealogy of individuals sampled at the present time is then described by the reversed reconstructed process (RRP), which traces the ancestry of the sample backwards from the present. We show that a simple, analytic, time rescaling of the RRP provides a straightforward way to derive its inter-event times. The same rescaling characterises other distributions underlying this process, obtained elsewhere in the literature via more cumbersome calculations. We also consider the case of incomplete sampling of the population, in which each leaf of the genealogy is retained with an independent Bernoulli trial with probability ψ, and we show that corresponding results for Bernoulli-sampled RRPs can be derived using time rescaling, for any values of the underlying parameters. A central result is the derivation of a scaling limit as ψ approaches 0, corresponding to the underlying population growing to infinity, using the time rescaling formalism. We show that in this setting, after a linear time rescaling, the event times are the order statistics of n logistic random variables with mode $\log(1/\psi)$; moreover, we show that the inter-event times are approximately exponentially distributed.

Keywords: birth-death, reconstructed process, Bernoulli sampling, time rescaling

1 Introduction

The coalescent is a widely used model describing the genealogy of a sample taken from a population, arising as the scaling limit of numerous population models (Hein et al., 2005). A key assumption of the basic coalescent is that the population size is large but constant or deterministically changing through time, although there are stochastic formulations (Kaj and Krone, 2003; Parsons et al., 2010). For some species, the dynamics of a population where individuals replicate and die independently of each other may be more naturally modelled as a birth-death process, which captures the stochastic nature and rapid growth of the population size (Boskova et al., 2014; Stadler et al., 2015). The simple linear birth-death process (BDP) studied by Kendall (1948) is a popular neutral population model, in which individuals independently divide at rate λ and die at rate μ. A realisation of this process can be represented as a tree relating the individuals, with bifurcations corresponding to birth events, and terminating branches corresponding to death events. The process models the entire population, creating a birth-death tree such as that shown on the left of Figure 1, where lineages can go extinct before the present. The genealogy of surviving individuals can then be obtained by pruning these extinct lineages, shown in the middle panel. The process tracing out the genealogy is termed the *reconstructed process (RP)* (Nee et al., 1994).

[*]anastasia.ignatieva@warwick.ac.uk
[†]Department of Statistics, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK
[‡]Department of Statistics, University of Warwick, Coventry CV4 7AL, UK
[§]Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
[¶]The Alan Turing Institute, British Library, London NW1 2DB, UK

Gernhard (2008a) considered the RP backwards in time, conditioning it on having n extant individuals at the present and a given time of origin T. Gernhard (2008a) noted a correspondence between this conditioned reconstructed process and a point process with i.i.d. speciation times; this is termed a *coalescent point process (CPP)* as introduced by Aldous and Popovic (2005) for critical branching processes. With this formulation, and using the results of Thompson (1975), Gernhard (2008a) then derived the density of bifurcation times in the RP, conditioned on T. Then, using an improper uniform $(0, \infty)$ prior on T and integrating, Gernhard (2008a) obtained an expression for the density of the k-th bifurcation time. In this article, we consider the time to origin to be random, similarly assuming a uniform prior on T, and condition on the sample size n at the present.

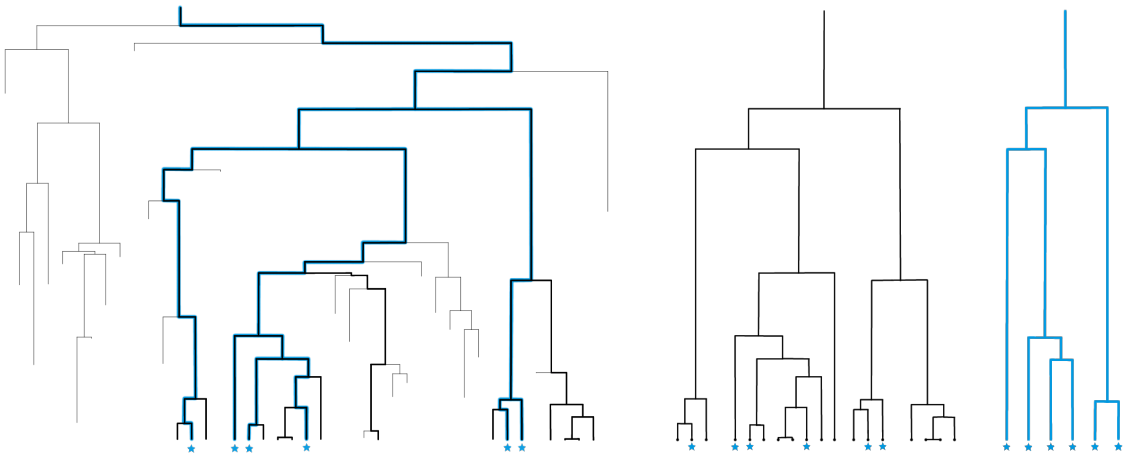

Figure 1: Left: birth-death tree with $\lambda = 0.1, \mu = 0.05$ and 18 individuals surviving to present time. Middle: corresponding RRP with complete sampling. Right: RRP with incomplete sampling: each surviving individual is sampled independently with fixed probability $\psi = \frac{1}{3}$; blue stars indicate sampled individuals

Birth-death models differ from the coalescent in that they must explicitly incorporate the sampling regime used in obtaining the sample. Two main sampling regimes have been considered in the literature: Bernoulli-type sampling (where each extant individual is sampled independently with some fixed probability ψ), and n-sampling (n individuals are sampled from the full population, of size N conditioned to be greater than n). Stadler (2009) analysed the conditioned reconstructed process with Bernoulli sampling and derived the joint density of bifurcation times for the sample (conditioned on the time of origin, or with a uniform prior). Wiuf (2018) and Stadler and Steel (2019) further looked at the correspondence between a complete and incompletely sampled process, by transforming the parameters. Lambert (2018) showed that there is a relationship between CPPs with Bernoulli sampling and CPPs with n-sampling: to simulate a CPP with n-sampling, one can first draw a realisation y of a random variable with a specific density, and then simulate a CPP under Bernoulli sampling with y as the sampling probability. We focus on Bernoulli-type sampling for fixed values of ψ, and then consider the limit $\psi \to 0$, which corresponds to the underlying population size (in the complete tree) growing to infinity.

There is also a substantial related body of work concerning Bienaymé-Galton-Watson (BGW) processes, considered in either discrete or continuous time, in which individuals reproduce independently according to a specified offspring distribution (in continuous time, when the number of offspring is either zero or two, we obtain the special case of a birth-death process). Work on reduced trees (tracing the ge-

nealogy of a sample) goes back several decades; Fleischmann and Siegmund-Schultze (1977) showed that the reduced tree associated with a BGW process is itself a time-inhomogeneous BGW process, similar to the result of Nee et al. (1994) for the reconstructed birth-death process. Several papers have considered the question of coalescence times for a finite sample (e.g. O'Connell, 1995; Harris et al., 2017; Grosjean and Huillet, 2018; Burden and Soewongsono, 2019). Similar to our treatment is that of O'Connell (1995), who derives an expression for the coalescence time of a sample of size 2, as a fraction of the time since origin of the process; Harris et al. (2017) generalise these results to any sample size, and consider continuous-time BGW processes sampled after time T, with n-sampling. Our work differs from the latter in the treatment of time to origin – rather than assuming that sampling happens a fixed time after the origin of the process, we treat the time of origin as random. Moreover, we consider Bernoulli type sampling, so the connection between our results is limited to the setting of taking the sampling probability to 0, which is conceptually similar to taking T to infinity. We focus specifically on results for birth-death processes, with any sample size, while Harris et al. (2017) limit the exposition of their results explicitly applied to birth-death processes in the limit $T \to \infty$ to $n = 2$. Our treatment of time to origin is more similar to the work of Burden and Soewongsono (2019), who consider the infinite-population limit of near-critical Galton-Watson process, arriving at the Feller diffusion; the time of most recent common ancestor is treated as random. However, we do not consider the diffusion approximation, and focus on supercritical rather than near-critical processes.

1.1 The birth-death process (BDP)

Consider a birth-death process $\mathcal{B}$ with birth rate λ and death rate $\mu \neq \lambda$ (which we will shorten as BDP(λ, μ)), with time measured in units β. The process starts with one individual at time 0 and is run for time T since origin, at which point all n extant individuals are sampled. We assume a uniform (improper) prior on T, reiterating that this choice of prior is not novel, and has been treated, for instance, in Aldous and Popovic (2005) for the critical case $\lambda = \mu$, and Gernhard (2008a) and Wiuf (2018) for the supercritical case. In this section, we give calculations showing that with this choice of improper prior, after conditioning on the number of sampled individuals n, the time since origin T_n^* of the conditioned process is random with a particular, proper distribution. We then show that the BDP(λ, μ) sampled a time T_n^* since origin and conditioned to have n sampled individuals is dual to the BDP(μ, λ), initialised with n individuals and run until first hitting state 0. This is not a new result, but it is crucial to the idea of considering the reconstructed process backwards in time from sampling, so we include it for completeness.

1.1.1 Prior on the time of origin

Let $\mathcal{B}_n$ denote the process $\mathcal{B}$ conditioned to have n sampled individuals, and denote by $\mathcal{B}_{n,t}$ this process with the sampling step happening at time t since origin. Here we will consider both the subcritical ($\mu > \lambda$) and supercritical ($\lambda > \mu$) cases. Let $N(t)$ denote the number of individuals alive in $\mathcal{B}_n$ at time t since origin. Then the generating function of $N(t)$ is given by (Athreya and Ney, 1972, Chapter III, Section 5):

$$G(s) = \mathbb{E}(s^{N(t)}) = \frac{\mu(s-1)e^{(\lambda-\mu)t} - \lambda s + \mu}{\lambda(s-1)e^{(\lambda-\mu)t} - \lambda s + \mu}.$$

Then

$$p_t := \mathbb{P}(N(t) = 0) = G(0) = \frac{\frac{\mu}{\lambda-\mu}(e^{(\lambda-\mu)t} - 1)}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t} - 1)},$$

and

$$\mathbb{P}(N(t)=j)=(1-p_t)\left(1-\frac{\lambda}{\mu}p_t\right)p_t^{j-1}.$$

Analogously to Aldous and Popovic (2005, Section 2), define the probability measure

$$\mathbb{P}^*(\mathcal{B}_n\in\cdot):=\frac{\int_0^\infty\mathbb{P}(\mathcal{B}_{n,t}\in\cdot)\mathbb{P}(N(t)=n)dt}{\int_0^\infty\mathbb{P}(N(t)=n)dt}. \tag{1.1}$$

Making the observation that $\mathbb{P}(N(t)=j)=\frac{1}{\mu}\left(\frac{d}{dt}p_t\right)(p_t)^{j-1}=\frac{1}{j\mu}\frac{d}{dt}(p_t^j)$, we obtain

$$\int_0^\infty\mathbb{P}(N(t)=n)\,dt=\frac{1}{n\mu}\left[p_t^n\right]_0^\infty=\begin{cases}\frac{1}{n\mu} & \text{if } \mu>\lambda\\ \frac{1}{n\mu}\left(\frac{\mu}{\lambda}\right)^n & \text{if } \lambda>\mu.\end{cases}$$

Then the function

$$f_{T_n^*}(t)=\frac{\mathbb{P}(N(t)=n)}{\int_0^\infty\mathbb{P}(N(t)=n)dt}=\begin{cases}\frac{n\mu e^{(\mu-\lambda)t}\left[\frac{\mu}{\mu-\lambda}(e^{(\mu-\lambda)t}-1)\right]^{n-1}}{\left[1+\frac{\mu}{\mu-\lambda}(e^{(\mu-\lambda)t}-1)\right]^{n+1}} & \text{if } \mu>\lambda\\ \frac{n\lambda e^{(\lambda-\mu)t}\left[\frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t}-1)\right]^{n-1}}{\left[1+\frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t}-1)\right]^{n+1}} & \text{if } \lambda>\mu\end{cases} \tag{1.2}$$

is a probability density on $[0,\infty)$ for the time since origin T_n^*, and (1.1) can be rewritten as

$$\mathbb{P}^*(\mathcal{B}_n\in\cdot)=\int_0^\infty f_{T_n^*}(t)\,\mathbb{P}(\mathcal{B}_{n,t}\in\cdot)dt.$$

Thus, after conditioning on the sample size n, $f_{T_n^*}(t)$ is a proper density for T_n^*.

1.1.2 Time reversal

The population size of the BDP(λ,μ) is a continuous-time Markov chain with the transition rates

$$q_{i,i+1}=\lambda i,\;\; q_{i,i-1}=\mu i.$$

As above, denote by $\{N(t),\ 0\le t\le T_n^*\}$ the corresponding process associated with the complete tree, counting the population size up to the time of sampling, making the jump from 0 to 1 at time 0. Consider also the continuous-time Markov chain $\{\widehat{N}_n(s),\ 0\le s\le T_n\}$, which has the reversed transition rates

$$q_{i,i+1}=\mu i,\;\; q_{i,i-1}=\lambda i,$$

started in state $\widehat{N}_n(0)=n$ and run until the first hitting time T_n of state 0. Then, mirroring Aldous and Popovic (2005, Lemma 2) for the critical case, we have the following:

Lemma 1.1.

$$\{N(T_n^*-s),\ T_n^*\ge s\ge 0\}\overset{d}{=}\{\widehat{N}_n(s),\ 0\le s\le T_n\},$$

and in particular $T_n^*\overset{d}{=}T_n$*, where* $\overset{d}{=}$ *denotes equality in distribution.*

Proof. Fix event times $s_0,\dots,s_M$, with $s_M>s_{M-1}>\dots s_1>s_0=0$, and positive integers $k_M=1,k_{M-1},\dots,k_2,k_1=n$, with $|k_m-k_{m-1}=1|$ and setting $k_{M+1}=0$. The sequence of k_m's describes a population size trajectory of a realisation of the birth-death process; reading from left to right, this has

$b+1$ increases of size 1, and $n-1+b$ decreases of size 1 for some integer $b \geq 0$ with $n+2b = M$. Then the event

> {as s decreases, $N(T_n^* - s)$ jumps from k_{m+1} to k_m for $s \in [s_m, s_m + ds_m]$ (for all $M \geq m \geq 1$) and makes no other jumps}

has measure

$$ds_M \cdot \prod_{m=M}^{2} \left(e^{-k_m(\lambda+\mu)(s_m - s_{m-1})} \, k_m \, ds_{m-1} \right) \cdot \lambda^{b+1} \, \mu^{n-1+b} \cdot e^{-k_1 s_1}, \tag{1.3}$$

where ds_M comes from the uniform prior. For the reversed process $\widehat{N}_n(s)$, the event

> {as s increases, $\widehat{N}_n(s)$ jumps from k_m to k_{m+1} in the interval $s \in [s_m, s_m + ds_m]$ (for all $1 \leq m \leq M$) and makes no other jumps}

has probability

$$\prod_{m=1}^{M} \left(e^{-k_m(\lambda+\mu)(s_m - s_{m-1})} \, k_m ds_m \right) \cdot \mu^{n-1+b} \, \lambda^{b+1}, \tag{1.4}$$

as reading the sequence of k_m's from right to left, there are $n-1+b$ increases of size 1 and $b+1$ decreases of size 1. The measure (1.3) is $1/k_1 = 1/n$ times (1.4), so after conditioning the probability measures of the two events are equal. □

This demonstrates the duality between the BDP(λ, μ) started with 1 individual at time 0 and reaching n individuals after the random time T_n^* (running "forwards" to the time of sampling), and the BDP(μ, λ), started from n individuals at time 0 and run until it reaches state 0 (running "backwards in time" from the sample). We next consider the reconstructed process, which tracks the genealogy of only the sampled individuals, and make use of the duality between the forwards-in-time and backwards-in-time formulations.

1.2 The reversed reconstructed process (RRP)

The RP (forwards in time) describes the number of lineages in the BDP(λ, μ), which will have at least one surviving descendant in the sample. Nee et al. (1994) identified that the RP forwards in time is generated by an underlying time-inhomogeneous pure birth process, with birth rate per lineage at time t given by:

$$\begin{aligned} \lambda P_1(t,T) &:= \lambda \cdot \mathbb{P}(\text{a single lineage born at time } t \text{ is not extinct by time } T) \\ &= \frac{\lambda(\lambda-\mu)}{\lambda - \mu e^{-(\lambda-\mu)(T-t)}} \\ &= \frac{\lambda e^{(\lambda-\mu)(T-t)}}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)(T-t)} - 1)}, \end{aligned} \tag{1.5}$$

where T is the time of sampling and $P_1(t,T)$ is given by Kendall (1948). The state of the process at time t is the number of individuals alive at t with at least one descendant at the time of sampling T, with events corresponding to transitions from state j to $j+1$, $j \geq 1$.

It is advantageous to consider the process running backwards in time from the present, conditioning on the sample size n, and not explicitly conditioning on the time of origin of the process (which is generally unknown, and for which we impose a uniform improper prior). In line with the discussion in Section 1.1.2, we will thus consider the properties of the *reversed reconstructed process (RRP)*, which is defined as

the process tracking the genealogy of the initial population of the BDP(μ, λ), initialised at n individuals and run until the first hitting time of state 0. It is straightforward to show, similarly to Lemma 1.1, that the RP with birth rate (1.5) run for time T_n^* and reaching state n at the time of sampling is dual to the RRP which is started in state n at time 0 and stopped at the first hitting time of state 0, with death rate obtained by replacing $T - \tau$ by τ in (1.5) to account for the time reversal. Note that the time index τ increases into the past, and $\tau = 0$ denotes the time of sampling. The RRP is thus an inhomogeneous pure-death process, with death rate per lineage given by:

$$m_\beta(\tau) = \frac{\lambda e^{(\lambda-\mu)\tau}}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau} - 1)}. \tag{1.6}$$

To obtain the death rate of the RRP with Bernoulli sampling (where each lineage is sampled with a fixed probability ψ at time 0), measured in time units of γ, we replace $P_1(\tau, T)$ with the relevant probability $P_\psi(t, T)$ as derived by Yang and Rannala (1997):

$$P_\psi(t,T) = \frac{\psi(\lambda-\mu)}{\psi\lambda - (\mu - (1-\psi)\lambda)e^{-(\lambda-\mu)(T-t)}} = \frac{\psi e^{(\lambda-\mu)(T-t)}}{1 + \frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)(T-t)} - 1\right)},$$

which, following the same reasoning as for the case of complete sampling, gives the RRP death rate:

$$m_\gamma(\tau) = \frac{\psi\lambda e^{(\lambda-\mu)\tau}}{1 + \frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau} - 1)}. \tag{1.7}$$

Note that for the case of a subcritical process (with $\lambda < \mu$), the population process backwards in time is supercritical. To ensure that the population reaches a common ancestor, we thus need to condition this process on ultimate extinction; it can be shown that this is equivalent to swapping the birth and death rate (Waugh, 1958), indeed this is clear from (1.2) for the time of origin. Thus, the RRP death rate in the subcritical case will be the same as (1.7) but with λ and μ interchanged.

For the case of a critical branching process, measured in time units of α, the death rate is given by taking the limit $\lambda \to \mu$ in (1.7):

$$m_\alpha(\tau) = \frac{\psi\lambda}{1 + \psi\lambda\tau}.$$

1.3 Overview

In this paper we consider the RRP as a backwards in time inhomogeneous pure-death process, as described in Section 1.2 above. We show that properties of the RRP are easily derived using this formulation. We use this to re-derive several results, such as densities of event times, which have been given elsewhere in the literature, but stress that the resulting proofs are significantly simpler and more intuitive.

Noting that there is a time rescaling between a time-reversed Yule rate 1 process and the RRP of the birth-death population model, we propose a new simulation algorithm for (incompletely) sampled RRPs using time rescaling. This is an alternative to existing algorithms (Hartmann et al., 2010; Stadler, 2011), which instead utilise a coalescent point process (CPP) formulation. We discuss the relationship between these two approaches. Further, we demonstrate the relationship between completely and incompletely sampled RRPs through time rescaling. In related work, e.g. Stadler and Steel (2012), the approach taken of transforming birth and death rates meant that results could be derived only for a restricted set of parameter values, in particular for $1 - \psi \leq \mu/\lambda \leq 1$; this is especially restrictive when ψ is small. Here,

we show instead that the completely and incompletely sampled RRPs are time-rescaled versions of each other, so distributions for the incompletely sampled case can be derived using a change of variables. We use this to complete the proof for the length of a randomly chosen pendant edge in Stadler and Steel (2012) for all parameter values.

Next we consider the scenario in which the underlying population size in a birth-death process grows to infinity, but a finite sample of size n is obtained. This can be thought of as taking the limit $\psi \to 0$ for the Bernoulli sampling probability; we discuss the connection with the limit as the total population size tends to infinity for n-sampling, using results of Lambert (2018). We describe in detail the time transformation between the RRP in this setting to a time-reversed Yule rate 1 process; in this scenario, there are two distinct timescales, separating the time of the first event from the events nearer the root of the tree. The RRP tree becomes star-shaped: the terminal branch lengths tend to infinity, while the inter-event times at the top of the tree are approximately exponentially distributed on a shorter timescale, with rate depending on the remaining number of lineages. We then use the time rescaling formalism to derive, analytically, the density of the inter-event times, both for any $\psi \in (0, 1]$ and in the limit $\psi \to 0$; both results are new to the best of our knowledge. We then show that in the limit $\psi \to 0$, the event times are distributed as the order statistics of n logistic random variables, with mode $\log(1/\psi)$ (after a simple, linear, time rescaling). Further, we show that the inter-event times (thus distributed as the spacings between consecutive order statistics of n logistic random variables) are approximately exponentially distributed, with error bounded by $1/n$ in terms of Kolmogorov-Smirnov distance. We also show that the expectation of inter-event times agrees exactly with the expectation under this approximation.

The paper is structured as follows. In Section 1.4 below, we introduce the notation used throughout. In Section 2, we state several known results for inhomogeneous birth-death processes which we will rely on, and review time rescaling for these processes. In Section 3, we consider the RRP of birth-death processes with Bernoulli sampling. In Section 4, we focus on the limit of the sampling probability going to 0. Finally, discussion is presented in Section 5. Proofs can be found in Appendix A.

Illustrations of trees throughout were made using the R package `ape` (Paradis and Schliep, 2018).

1.4 Notation

Table 1 summarises the notation used throughout. For instance, $\mathrm{BDP}(\lambda, \mu, \psi)$ denotes the birth-death population process where each individual divides independently with rate λ, dies independently with rate $\mu < \lambda$, with the rates measured in time units γ; at time 0, each surviving individual is sampled with a fixed probability ψ. The corresponding RRP, i.e. the process tracing out the genealogy of the sample from this population backwards in time from 0, is denoted by $X_\psi^\gamma := (X_\psi^\gamma(\tau) : \tau \geq 0)$. We write X_ψ^ξ to denote the same process, but with time rescaled to units of $\xi = g(\gamma)$ for some time transformation g, i.e. $X_\psi^\xi(g(\tau)) = X_\psi^\gamma(\tau)$. We denote the death rates of X_ψ^γ and X_ψ^ξ by m_γ and m_ξ, respectively, with the subscripts denoting the time scale on which the rates are measured.

A table summarising the properties of each RRP is given in Appendix B for reference.

2 Background

We briefly review relevant known results which we will rely on throughout the paper.

Population process	**Time unit**	**Notation**	**RRP notation**
Yule process, birth rate 1	t	Yule(1)	Y
Critical branching process, birth = death rate λ, sampling probability ψ	α	CBP(λ, ψ)	Z_ψ^α
Birth-death process, birth rate λ, death rate μ, complete sampling	β	BDP$(\lambda, \mu, 1)$	X_1^β
Birth-death process, birth rate λ, death rate μ, sampling probability ψ	γ	BDP(λ, μ, ψ)	X_ψ^γ
Birth-death process, birth rate λ', death rate μ', with $\lambda' - \mu' = 1$ sampling probability ψ	δ	BDP(λ', μ', ψ)	X_ψ^δ

Table 1: Summary of notation

2.1 Inhomogeneous pure-death processes

We briefly state relevant known results concerning inhomogeneous pure-death processes. Consider a time-inhomogeneous pure-death process, with time measured in units ξ, starting with n individuals alive at time 0. Each individual dies independently at rate $m_\xi(\tau)$; if there are j individuals at time τ, the intensity is $jm_\xi(\tau)$. The rate function of the process is given by

$$\rho_\xi(\tau) = \int_0^\tau m_\xi(x)dx.$$

The transition probabilities, i.e. the probability of going from n to j individuals in time τ, are given by a binomial distribution (Bailey, 1964, p.112):

$$P_{nj}(\tau) = \begin{cases} \binom{n}{j}\left(1 - e^{-\rho_\xi(\tau)}\right)^{n-j}\left(e^{-\rho_\xi(\tau)}\right)^j & \text{for } j \leq n, \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

with $e^{-\rho_\xi(\tau)}$ being the probability that a lineage has not died by time τ. The distribution of time to origin is (Bailey, 1964, p.112):

$$F_{T_n}(\tau) = P(T_n < \tau) = \left(1 - e^{-\rho_\xi(\tau)}\right)^n, \tag{2.2}$$

and, by differentiating, the pdf is

$$f_{T_n}(\tau) = nm_\xi(\tau)e^{-\rho_\xi(\tau)}\left(1 - e^{-\rho_\xi(\tau)}\right)^{n-1}. \tag{2.3}$$

The density of the time of the k-th event is given by

$$\begin{aligned} f_{T_k}(\tau) &= \binom{n}{k} \cdot \underbrace{km_\xi(\tau)e^{-\rho_\xi(\tau)}\left(1 - e^{-\rho_\xi(\tau)}\right)^{k-1}}_{k\text{-th lineage dies at } \tau} \cdot \underbrace{\left(e^{-\rho_\xi(\tau)}\right)^{n-k}}_{n-k \text{ survive for at least } \tau} \\ &= \binom{n}{k} k\, m_\xi(\tau)\left(1 - e^{-\rho_\xi(\tau)}\right)^{k-1}\left(e^{-\rho_\xi(\tau)}\right)^{n-k+1}. \end{aligned} \tag{2.4}$$

2.2 Time rescaling

Consider a pure-death inhomogeneous process with death rate $m_\xi(\tau)$, with time measured in units of ξ. Suppose that time is rescaled in units of $\zeta = g(\xi)$, where g is strictly monotonic and differentiable. The death rate of the process then becomes, using a change of variables:

$$m_\zeta(\tau) = m_\xi(g^{-1}(\tau))\left|\frac{d}{d\tau}\, g^{-1}(\tau)\right|.$$

The time rescaling theorem, due to Meyer (1971) and Papangelou (1972), states that any inhomogeneous point process with an integrable intensity function can be rescaled to a Poisson process with unit rate. The RRP can be thought of as a point process, with intensity given by its inhomogeneous death rate times the number of lineages. If the RRP (of any population process) has death rate $m_\xi(\tau)$, then rescaling time via the transformation $g = \rho_\xi$ rescales the RRP to a homogeneous pure-death process with death rate per lineage equal 1 (a time-reversed Yule rate 1 process).

2.3 Time-reversed Yule rate 1 process

We define the time-reversed Yule rate 1 process as a pure death process where each lineage dies independently at rate 1, denoted Y. This is the RRP of a forwards-in-time Yule process with birth rate 1. The inter-event time during which there are exactly j lineages is exponentially distributed with rate j. Using (2.3), the time to origin has density:

$$f_{T_n}(\tau) = ne^{-\tau}(1-e^{-\tau})^{n-1},$$

and using (2.4), the time to k-th event has density:

$$f_{T_k}(\tau) = \binom{n}{k}\, k\, \left(1-e^{-\tau}\right)^{k-1}\left(e^{-\tau}\right)^{n-k+1}. \tag{2.5}$$

The expectation of time to origin is $\sum_{j=1}^{n} \frac{1}{j}$. These results are identical to those derived by Gernhard (2008b).

3 Birth-death process with Bernoulli sampling

We now consider in detail the RRP X_ψ^γ of a supercritical birth-death process. Using the formulation introduced in Section 1.2, we first re-derive some known properties of the process, which will be readily available from the results given in Section 2. Then, using the fact that the RRP X_ψ^γ is a time rescaling of the RRP associated with a Yule rate 1 process, we propose a simulation algorithm. Finally, we discuss the relationship between completely and incompletely sampled RRPs through time rescaling.

3.1 Properties of the process

Set $T_0 = 0$ and for $k \in \{1, \ldots, n\}$ denote by T_k the time of the k-th event, backwards from the present time 0. At T_k, the number of lineages decreases from $n-k+1$ to $n-k$. For $k \in \{0, \ldots n-1\}$, let $W_k := T_{k+1} - T_k$ denote the inter-event time.

3.1.1 Transition probabilities and densities of event times

We use the pure-death process formulation of X_ψ^γ to derive distributions characterising this process. The transition probabilities are, using (2.1):

$$P_{ij}(\tau) = \begin{cases} \binom{i}{j}\big(1-e^{-\rho_\gamma(\tau)}\big)^{i-j}\big(e^{-\rho_\gamma(\tau)}\big)^{j} & \text{for } j \le i \\ 0 & \text{otherwise,} \end{cases}$$

where, by integrating the death rate in (1.7):

$$\rho_\gamma(\tau) = \int_0^\tau m_\gamma(x)dx = \log\!\left(1 + \frac{\psi\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\tau}-1\Big)\right), \tag{3.1}$$

and

$$e^{-\rho_\gamma(\tau)} = \frac{1}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}.$$

For $\tau \to \infty$ and fixed $\psi \in (0,1]$, we have $\rho_\gamma(\tau) \to \infty$ and $e^{-\rho_\gamma(\tau)} \to 0$, so $P_{ij}(\tau) \to 0$ for all $j \neq 0$, and $P_{i0}(\tau) \to 1$. This implies that two individuals sampled at the present will eventually find a common ancestor in the past with probability 1.

The distribution of time to origin, using (2.2), is given by:

$$F_{T_n}^\psi(\tau) = P(T_n < \tau) = \big(1-e^{-\rho_\gamma(\tau)}\big)^n = \left(\frac{\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\right)^n,$$

and its density, using (2.3), is

$$\begin{aligned} f_{T_n}^\psi(\tau) &= n\cdot\frac{\psi\lambda e^{(\lambda-\mu)\tau}}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\cdot\frac{1}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\cdot\left(\frac{\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\right)^{n-1} \\ &= n\psi\lambda e^{(\lambda-\mu)\tau}\frac{\Big[\frac{\psi\lambda}{\lambda-\mu}\big(e^{(\lambda-\mu)\tau}-1\big)\Big]^{n-1}}{\Big[1+\frac{\psi\lambda}{\lambda-\mu}\big(e^{(\lambda-\mu)\tau}-1\big)\Big]^{n+1}}. \end{aligned} \tag{3.2}$$

Note that this agrees with (1.2) for the case $\psi = 1$. This result is also obtained in Stadler (2009, Lemma 3.1). We note that although the outcome is identical, the derivation given above is significantly simpler, and follows directly from the properties of the RRP as a stochastic process. In particular, the distribution function is immediately obtained from knowing the death rate; moreover, to obtain the pdf we do not need to integrate over the prior for the time of origin, as this is implicit in the time reversal.

Using (2.4), the waiting time to the k-th event is given by:

$$\begin{aligned} f_{T_k}^\psi(\tau) &= \binom{n}{k}\, k\, m_\gamma(\tau)\big(1-e^{-\rho_\gamma(\tau)}\big)^{k-1}\big(e^{-\rho_\gamma(\tau)}\big)^{n-k+1} \\ &= \binom{n}{k}\, k\, \frac{\psi\lambda e^{(\lambda-\mu)\tau}}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\left(\frac{\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\right)^{k-1}\left(\frac{1}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau}-1)}\right)^{n-k+1} \\ &= \binom{n}{k}\, k\, \psi\lambda e^{(\lambda-\mu)\tau}\frac{\Big[\frac{\psi\lambda}{\lambda-\mu}\big(e^{(\lambda-\mu)\tau}-1\big)\Big]^{k-1}}{\Big[1+\frac{\psi\lambda}{\lambda-\mu}\big(e^{(\lambda-\mu)\tau}-1\big)\Big]^{n+1}}. \end{aligned} \tag{3.3}$$

This agrees with the result derived in Gernhard (2008a, Theorem 4.1) for the case of complete sampling; we again note that the result follows almost immediately from the properties of the RRP, which removes the need for deriving the related distributions by hand.

3.1.2 Simulating from the RRP

As described in Section 2.2, applying the time transformation $g_1 = \rho_\gamma$ rescales the RRP X_ψ^γ to the time-reversed Yule rate 1 process Y. From (3.1), this transformation is given by:

$$\begin{aligned} t = g_1(\gamma) &= \log\Big(1 + \frac{\psi\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\gamma} - 1\Big)\Big), \\ \gamma = g_1^{-1}(t) = \rho_\gamma^{-1}(t) &= \frac{1}{\lambda-\mu}\log\Big(1 + \frac{\lambda-\mu}{\psi\lambda}\Big(e^{t} - 1\Big)\Big), \end{aligned} \tag{3.4}$$

and we have that

$$X_\psi^\gamma(g_1^{-1}(\tau)) = Y(\tau) \text{ and } X_\psi^\gamma(\tau) = Y(g_1(\tau)),$$

by which we mean that X_ψ^γ rescaled in time units $g_1(\gamma)$ has the same death rate as Y. To see why this works, the death rate of X_ψ^γ when measured in units $t = g_1(\gamma)$ becomes:

$$m_t(\tau) = m_\gamma(g_1^{-1}(\tau))\left|\frac{d}{d\tau}\, g_1^{-1}(\tau)\right| = m_\gamma(\rho_\gamma^{-1}(\tau))\left|\frac{d}{d\tau}\, \rho_\gamma^{-1}(\tau)\right| = m_\gamma(\rho_\gamma^{-1}(\tau))\Big/ m_\gamma(\rho_\gamma^{-1}(\tau)) = 1.$$

Note also that in the complete process, birth, death, and sampling events affect all individuals with equal probability, so the RRP trees we study have the same law in topology as Yule and coalescent trees (Aldous, 1996), and can thus be generated backwards in time by merging pairs of lineages selected uniformly at random. This suggests that to simulate from X_ψ^γ, first we can simulate from Y, and then rescale the event times using the transformation given by (3.4). The method is summarised as Algorithm 1. This provides an alternative to the algorithms of Hartmann et al. (2010) and Stadler (2011), where first the time of origin is drawn from its distribution, and then the coalescent point process formulation is used to obtain the event times.

Algorithm 1 Simulating from the RRP X_ψ^γ

Given n individuals at time 0:

1. Draw $\widetilde{W}_j \sim \text{Exp}(n-j)$ for $j = 0, \ldots, n-1$, being the waiting times of Y.
2. Compute the event times $\widetilde{T}_{j+1} = \sum_{i=0}^{j} \widetilde{W}_i$.
3. Rescale the event times as $T_k = \frac{1}{\lambda-\mu}\log\Big(1 + \frac{\lambda-\mu}{\psi\lambda}\Big(\exp\Big(\widetilde{T}_k\Big) - 1\Big)\Big)$ for $k = 1, \ldots, n$.
4. Construct a tree from $T_1, \ldots, T_n$ by choosing a pair of lineages uniformly at random to coalesce at each event time.

Note that one can first derive distributions of interest for Y, and then use the change of variables given by (3.4) to obtain the equivalent results for X_ψ^γ. We will use this to derive the distribution of inter-event times W_k, analytically, in Section 4.3.

3.1.3 Relationship with coalescent point processes

Gernhard (2008a) gives the following CPP formulation for a supercritical process. To simulate an

RRP for a sample of size n, first condition on the sample size and a time of origin T_n (possibly drawn from the distribution (2.2)), and then draw the times of the $n-1$ bifurcations in the tree i.i.d. from some specific density depending on T_n. Lambert and Stadler (2013) further give this density for the case of Bernoulli sampling. In a sense, conditioning on the time of origin, the event times can thus be simulated "horizontally", one-by-one for each sampled lineage, rather than "vertically", i.e. forwards or backwards in time.

The formulation of the RRP as a pure-death process also allows for simulation of the RRP lineage-by-lineage, conditioning on the sample size but not on the time of origin (producing a tree including the root edge). Because each lineage dies independently from the others, in order to simulate from X_ψ^γ for a sample of size n, we can simulate the death times of each of the n lineages independently, and then merge the lineages uniformly at random at each event time to create the tree. The death time of one lineage has density:

$$f^\psi_{T_{(1)}}(\tau) = \frac{\psi\lambda e^{(\lambda-\mu)\tau}}{\left[1+\frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)\tau}-1\right)\right]^2}, \tag{3.5}$$

which is obtained from (3.2) by substituting $n=1$; this can be simulated by drawing from an exponential rate 1 density, and rescaling time using (3.4). Therefore the relationship between CPP and the pure-death formulation is very direct. With the pure-death formulation, each of the n lineages dies independently with the same death rate. Once we also condition on a time of origin T_n, the lineages still die independently, with death rate amended so that each event happens before T_n. The latter is exactly the CPP formulation of Gernhard (2008a).

The CPP formulation described in Lambert and Stadler (2013) also gives a method for simulating a Bernoulli RRP without conditioning on the sample size, as follows. Given a time of origin T, draw realisations $H^\psi_1, \ldots, H^\psi_N$ of a random variable H^ψ, with the stopping criterion that H^ψ_N is the first realisation that is greater than T. Then the $H^\psi_1, \ldots, H^\psi_{N-1}$ are the event times up to the MRCA for a sample of N lineages in a Bernoulli sampled RRP, conditioned on time of origin T. Note that in this case, setting $p = P(H^\psi > T)$, the number of sampled lineages is geometric with mass function $(1-p)^{n-1}p$, and the density of H^ψ given in Lambert and Stadler (2013, p.122) is exactly that in (3.5).

The pure-death formulation of the RRP highlights two differences between the genealogy of a birth-death process and the coalescent. Firstly, viewing the basic coalescent as a backwards in time pure-death process with rate $\frac{j(j-1)}{2}$ when there are j lineages, at each point in time the death rate of each individual lineage depends on the total number of lineages remaining; this dependence cannot be removed by conditioning on the time of origin (for $n > 2$). This implies that the process cannot be simulated by drawing the death time of each lineage independently from some density, as for the RRP. This supports the conjecture of Lambert and Stadler (2013) that the coalescent does not have a CPP representation.

Secondly, the coalescent with variable population size, as described by Griffiths and Tavare (1994), can be described as an inhomogeneous pure-death process, where the death rate is quadratic in the number of lineages and depends on a population size function. Because the death rate of the RRP is linear in the number of lineages, there is no population size function which would equate the two models.

3.2 Relationship between completely and incompletely sampled RRPs

Stadler (2009) noted that there is a relationship between the RRP of the incompletely sampled BDP(λ, μ, ψ), and the RRP of the completely sampled BDP$(\widehat{\lambda}, \widehat{\mu}, 1)$, through the following transformation

of the birth and death parameters:

$$\widehat{\lambda} = \psi\lambda, \;\; \widehat{\mu} = \mu - \lambda(1-\psi). \tag{3.6}$$

Substituting (3.6) as the birth and death rates into (1.6) gives (1.7). Thus, the resulting process looks like the RRP of an incompletely sampled BDP(λ, μ, ψ) population process. However, as noted by Stadler and Steel (2012), $\widehat{\mu}$ can be negative (in particular, for very small values of ψ); for instance, with the parameters used in Figure 1, $\widehat{\mu} = -1/60$. In this case, the interpretation as an RRP of some birth-death process is problematic. Stadler and Steel (2012, 2019) discuss that when distributions are derived for the completely sampled process, this reparameterisation trick can be used to obtain the equivalent distributions for a process with incomplete sampling, but only for $\frac{\mu}{\lambda} \geq 1 - \psi$. Thus, this method of transforming the birth and death rates does not always produce a valid mapping between completely and incompletely sampled RRPs.

To avoid this issue, instead of transforming the birth and death parameters directly, we use a transformation of time, and demonstrate the relationship between the RRPs X_ψ^γ and X_1^β. We do not introduce restrictions on the values of the parameters (λ, μ, ψ), so this allows distributions derived for the completely sampled process to be transformed for the case of incomplete sampling.

3.2.1 Time transformation from X_ψ to X_1

Define the transformation of time units g_2 as:

$$\begin{aligned} \beta = g_2(\gamma) &= \frac{1}{\lambda - \mu} \log\Big(1 + \psi(e^{(\lambda-\mu)\gamma} - 1)\Big), \\ \gamma = g_2^{-1}(\beta) &= \frac{1}{\lambda - \mu} \log\left(1 + \frac{1}{\psi}(e^{(\lambda-\mu)\beta} - 1)\right). \end{aligned} \tag{3.7}$$

This is a valid time transformation with $\gamma = 0 \iff \beta = 0$, and $\gamma = \beta$ when $\psi = 1$. Using a change of variable in (1.7), we compute the death rate:

$$\begin{aligned} m_\beta(\tau) &= m_\gamma(g_2^{-1}(\tau)) \cdot \left|\frac{dg_2^{-1}(\tau)}{d\tau}\right| \\ &= \frac{\psi\lambda(1 + \frac{1}{\psi}(e^{(\lambda-\mu)\tau} - 1))}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau} - 1)} \cdot \frac{\frac{1}{\psi}e^{(\lambda-\mu)\tau}}{1 + \frac{1}{\psi}(e^{(\lambda-\mu)\tau} - 1)} \\ &= \frac{\lambda e^{(\lambda-\mu)\tau}}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\tau} - 1)}. \end{aligned}$$

This is the death rate of the completely sampled RRP X_1^β as given in (1.6). Thus, we have the relationship:

$$\begin{aligned} X_1^\beta(\tau) &= X_\psi^\gamma(g_2^{-1}(\tau)), \\ X_1^\beta(g_2(\tau)) &= X_\psi^\gamma(\tau). \end{aligned}$$

The RRP of a BDP(λ, μ, ψ) process is a time rescaled version of the RRP of a completely sampled BDP$(\lambda, \mu, 1)$ process. In effect, introducing incomplete sampling is equivalent to non-linearly rescaling the RRP of the BDP$(\lambda, \mu, 1)$ process using the time transformation (3.7).

3.2.2 Deriving results for X_ψ from X_1

Using the time transformation approach, distributions can be derived for X_1^β with complete sampling, and then the equivalent distribution results for X_ψ^γ can be obtained through a simple change of variables. As an example, Stadler and Steel (2012) derive the density of the length of a randomly chosen pendant edge (an edge adjacent to a sampled individual) for an incompletely sampled tree with the restriction $1-\psi \leq \frac{\mu}{\lambda} \leq 1$; we complete the proof for the case $0 \leq \frac{\mu}{\lambda} \leq 1-\psi$.

Proposition 3.1. *The density of a the length of a randomly chosen pendant edge, E, of the RRP X_ψ^γ for any $0 \leq \mu < \lambda$ and $\psi \in (0,1]$ is*

$$f_E^\psi(\tau) = \frac{2\psi\lambda(\lambda-\mu)^3 e^{(\lambda-\mu)\tau}}{\left(\lambda\psi e^{(\lambda-\mu)\tau} - [\mu - \lambda(1-\psi)]\right)^3}.$$

Proof. Mooers et al. (2012) give the density of the length of a pendant edge of a completely sampled RRP X_1^β as:

$$f_E^1(\tau) = \frac{2\lambda(\lambda-\mu)^3 e^{(\lambda-\mu)\tau}}{(\lambda e^{(\lambda-\mu)\tau} - \mu)^3}. \tag{3.8}$$

Using the time rescaling (3.7) and a change of variable, for X_ψ^γ this becomes:

$$\begin{aligned} f_E^\psi(\tau) &= f_E^1(g_2(\tau)) \left|\frac{d\, g_2(\tau)}{d\tau}\right| \\ &= \frac{2\lambda(\lambda-\mu)^3[1+\psi(e^{(\lambda-\mu)\tau}-1)]}{\left(\lambda\left[1+\psi(e^{(\lambda-\mu)\tau}-1)\right]-\mu\right)^3} \cdot \frac{\psi e^{(\lambda-\mu)\tau}}{1+\psi(e^{(\lambda-\mu)\tau}-1)} \\ &= \frac{2\psi\lambda(\lambda-\mu)^3 e^{(\lambda-\mu)\tau}}{\left(\lambda\psi e^{(\lambda-\mu)\tau} - [\mu-\lambda(1-\psi)]\right)^3}. \end{aligned}$$

□

Equivalence with the result of Stadler and Steel (2012, Section 4) for $\frac{\mu}{\lambda} \geq 1-\psi$ is easily checked by substituting the birth rate $\widehat{\lambda}$ and death rate $\widehat{\mu}$ into (3.8).

4 Sampling from large populations

We now consider the setting where the total population size is very large compared to the sample size n. This is a scenario often encountered in practice when collecting genetic data, particularly from viral or bacterial populations, when the population size is unknown but can be presumed very large. An example will be mentioned within the discussion in Section 5.

This situation is to be distinguished from the limit as the sample size grows to infinity, which has been considered in Wiuf (2018). The scenario of interest here is when the total population tends to infinity, but a finite sample of size n is obtained. This can be interpreted as either the Bernoulli sampling probability ψ going to 0, or the total population size growing to infinity in the case of n-sampling. In the following section we will discuss why the two regimes are conceptually similar.

In this section, for the sake of readability of the expressions, we rescale time linearly in units of $\delta = (\lambda-\mu)\gamma$, and write $\lambda' = \frac{\lambda}{\lambda-\mu}, \mu' = \frac{\mu}{\lambda-\mu}$ with $\lambda' - \mu' = 1$. This simplifies the formulas, and is easy

to reverse within any derived expressions. The RRP on this timescale is denoted X_ψ^δ, with death rate

$$m_\delta(\tau) = \frac{\psi\lambda' e^\tau}{1 + \psi\lambda'(e^\tau - 1)}.$$

The time transformation between X_ψ^δ and Y is given by $g_3 = \rho_\delta$, with

$$t = g_3(\delta) = \log\big(1 + \psi\lambda'(e^\delta - 1)\big), \tag{4.1}$$

$$\delta = g_3^{-1}(t) = \rho_\delta^{-1}(t) = \log\left(1 + \frac{1}{\psi\lambda'}(e^t - 1)\right), \tag{4.2}$$

4.1 Sampling method

Lambert (2018) showed the following relationship between the two sampling scenarios when considered from a CPP perspective. Bernoulli sampled trees can be generated using the CPP formulation; that is, conditioning on a time of origin T, the event times are i.i.d. according to a specific density (as described in Section 3.1.3). For n-sampling, if we were to first generate a CPP tree with complete sampling (conditioned to have size at least n), and then choose n lineages uniformly at random, then this would not have a CPP formulation (Lambert and Stadler, 2013). However, the genealogy of such an n-sample can be obtained by first drawing a sampling probability $\Psi = y$ from a specific improper prior, and then generating a Bernoulli CPP of size n with sampling probability y. The improper prior has the form (Lambert, 2018, Theorem 3):

$$\frac{n(1-a)y^{n-1}}{(1-a(1-y))^{n+1}},$$

where $a = P(H < T)$ is the probability that the random variable corresponding to event times (in the complete tree) takes a value less than the specified time of origin.

The underlying population (of the complete tree) growing to infinity can be seen to correspond to the time of origin of the complete process growing to infinity, and thus the probability $a = P(H < T)$ approaching 1. In this case, the improper prior on Ψ tends to a point mass at $y = 0$. We do not explicitly condition on T, however this argument implies that the behaviour of the RRP for Bernoulli sampling with $\psi \to 0$, and for n-sampling when the underlying population grows to infinity, should be the same.

4.2 Relationship between X_ψ^δ and Y for small ψ

We examine the effect of the time rescaling between X_ψ^δ and the time-reversed Yule rate 1 process Y, when $\psi \to 0$. In the following, we assume that λ' is fixed and very small compared to $1/\psi$.

Consider the time rescaling given by (4.1): the process X_ψ^δ rescaled in units of $g_3(\delta)$ is a time-reversed Yule rate 1 process. This rescaling is illustrated in Figure 2 for a small value of ψ; the left panel shows a realisation of X_ψ^δ for $n = 10$. The right panel shows the same tree, but the intervals delineated by blue lines in the left panel are rescaled to intervals of equal length in the right panel.

Using the identity $\log(1 + x) = \log(x) + \log(1 + 1/x)$ and a Taylor expansion in $\psi\lambda'$ around 0, we obtain from (4.2):

$$\delta - \log\left(\frac{1}{\psi\lambda'}\right) = \log\big(e^t - 1\big) + \mathcal{O}(\psi\lambda'). \tag{4.3}$$

For small t, we have that $e^t - 1 \approx t$ and the transformation behaves as $\delta - \log(1/(\psi\lambda')) \approx \log(t)$. For large t, we have $\log(e^t - 1) \approx t$, so $\delta - \log(1/(\psi\lambda')) \approx t$. Thus, there are two time regimes, with a smooth transition between them.

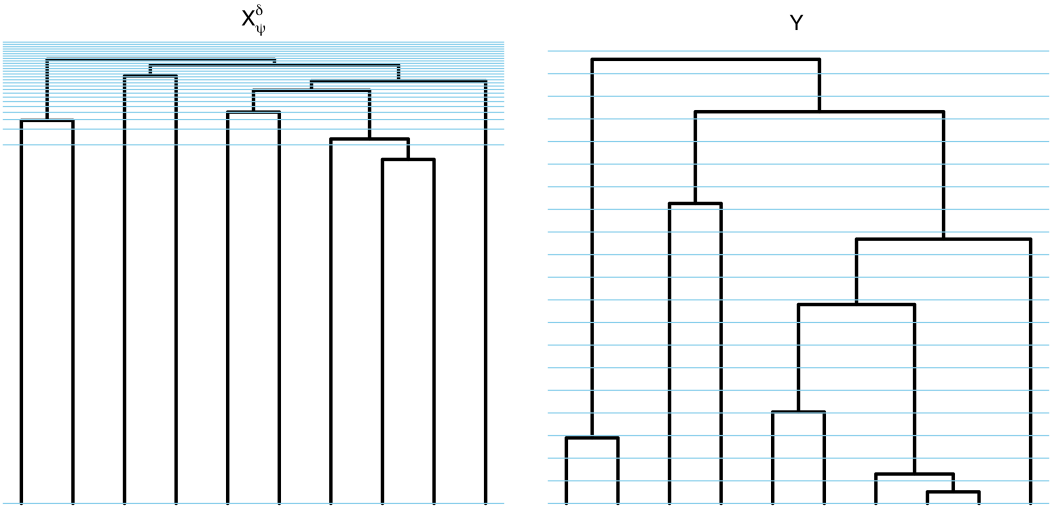

Figure 2: Left: realisation of X^{δ}_{ψ} with $\psi = e^{-20}, \lambda' = 2$. Right: same tree, rescaled in time units given by (4.1). Intervals delineated by blue lines in the left panel are rescaled to intervals of equal length in the right panel.

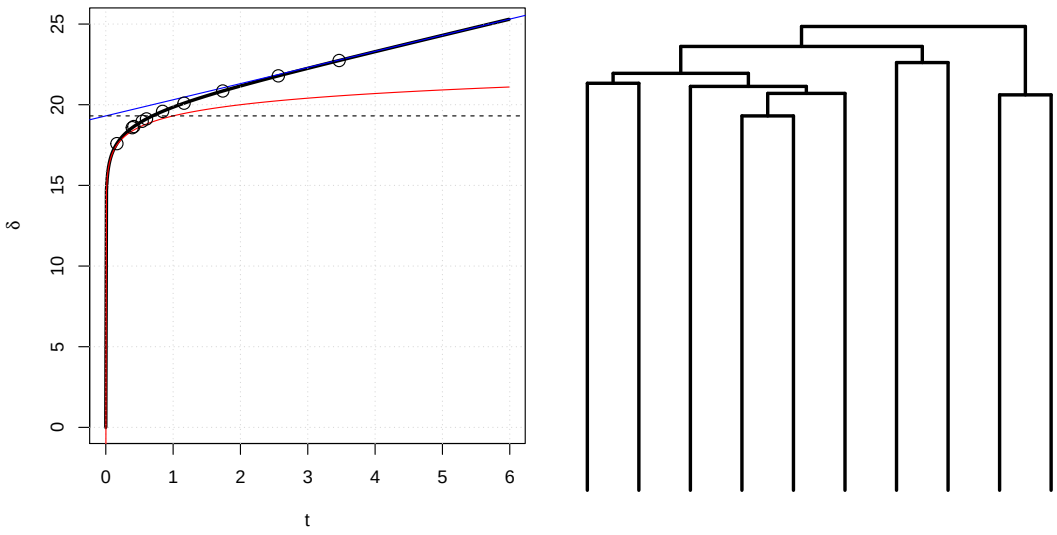

Figure 3: Left: solid black line shows time rescaling between δ and t (time units of Y). In blue: line $\delta = t - \log(\psi\lambda')$. In red: curve $\delta = \log(t) - \log(\psi\lambda')$. Dashed black line shows $\delta = -\log(\psi\lambda')$. Dots show simulated event times. Right: tree corresponding to the simulated event times. Parameters used: $\psi = e^{-20}, \lambda' = 2, \mu' = 1$.

This can be understood as follows. Under Bernoulli sampling, the sample size n is of order ψN, where N is the underlying population size in the complete tree. In the limit $\psi \to 0$, N is therefore $\mathcal{O}(\psi^{-1})$; this is very large compared to n, and no coalescences happen for a very long time: the probability of going from n to $n-1$ individuals in time τ is, from (2.1):

$$P_{n,n-1}(\tau) = \left(1 - e^{-\rho_\delta(\tau)}\right)\left(e^{-\rho_\delta(\tau)}\right)^{n-1},$$

where

$$e^{-\rho_\delta(\tau)} = \frac{1}{1 + \psi\lambda'(e^\tau - 1)}$$

is the probability of no event happening. This is very close to 1 until τ grows to the order of $\log(1/\psi)$.

With the time transformation above, a step of one unit of t approximately corresponds to taking a time step of $\log(1/\psi)$ in units of δ. At this point, $e^{-\rho_\delta(\tau)} \approx (1+\lambda')^{-1}$ and the sample starts to coalesce. Then steps in t become roughly equal to steps in δ. In essence, we zoom back to a time when the underlying population was of order n, and then slow back down to linear time.

Figure 3 shows an example of the time rescaling (4.2) for $\psi = e^{-20}$, $\lambda' = 2$, $\mu' = 1$. The left panel shows δ against t; the horizontal axis is the time scale of the Y process, the vertical axis is the time scale of X^δ_ψ. The red line shows the curve $\delta = \log(t) + \log(1/(\psi\lambda'))$; the blue line shows $\delta = t + \log(1/(\psi\lambda'))$. The circles indicate a set of simulated event times for a sample size $n = 10$. For instance, the time to first event is $T_1 \sim \text{Exp}(n)$ on the horizontal axis; this is rescaled using (4.2) to get the corresponding time on the vertical axis. The right panel shows the corresponding RRP tree.

As $\psi \to 0$, $\log(1/(\psi\lambda')) \to \infty$, so the rescaled time of the first event in units of δ grows to infinity, and the reconstructed tree of X^δ_ψ becomes star-shaped. The terminal branches dominate the tree, but the inter-event times near the origin of the tree are still approximately exponentially distributed with rate depending on the remaining number of lineages, as the time rescaling for large t is approximately linear.

4.3 Density of inter-event times in the limit $\psi \to 0$

We now derive, analytically, the density of inter-event times, first for any $\psi \in (0, 1]$, then for the limit $\psi \to 0$.

Theorem 4.1. *The density of waiting times $W_k = T_{k+1} - T_k$ between events k and $k+1$, $k \in \{0, \ldots, n-1\}$, for the RRP X^δ_ψ with $\psi \in (0, 1]$, is:*

$$f^\psi_{W_k}(w) = \frac{(n-k)}{(n+1)} e^{-(n-k)w}\Big[(n+1)\,{}_2F_1(n-k+1, n-k+1; n+1; (1-\psi\lambda')(1-e^{-w})) \\ - (1-\psi\lambda')(n-k+1)\,{}_2F_1(n-k+1, n-k+2; n+2; (1-\psi\lambda')(1-e^{-w}))\Big], \quad (4.4)$$

where ${}_2F_1$ is the ordinary hypergeometric function. In the case of a critical branching process with birth and death rate λ and RRP Z^α_ψ, this becomes:

$$\hat{f}^\psi_{W_k}(v) = \frac{(n-k+1)(n-k)}{n+1} \psi\lambda \cdot {}_2F_1(n-k+1, n-k+2; n+2; -\psi\lambda v).$$

Note that for $k = 0$, $f^\psi_{W_0}(w)$ reduces to the density of the first event, obtained by substituting $k = 1$ in (3.3). We have the following case for $\psi \to 0$:

Corollary 4.1. *The density of waiting times W_k between events k and $k+1$, $k \in \{1, \ldots, n-1\}$, in the limit $\psi \to 0$, is:*

$$f^0_{W_k}(w) = \frac{k(n-k)}{n+1} e^{-(n-k)w} {}_2F_1(n-k+1, n-k+1; n+2; 1-e^{-w}). \tag{4.5}$$

This is not a density for $k = 0$, i.e. for the waiting time to the first event; recall that for $\psi \to 0$ the first event time goes to infinity.

Note that using the transformation (Erdélyi et al., 1953, p.64)

$${}_2F_1(a, b; c; z) = (1-z)^{c-a-b} {}_2F_1(c-a, c-b; c; z),$$

the densities of the k-th and $(n-k)$-th waiting times are equal:

$$\begin{aligned} f^0_{W_k}(w) &= \frac{k(n-k)}{n+1} e^{-(n-k)w} {}_2F_1(n-k+1, n-k+1; n+2; 1-e^{-w}) \\ &= \frac{k(n-k)}{n+1} e^{-(n-k)w} e^{-(2k-n)w} {}_2F_1(k+1, k+1; n+2; 1-e^{-w}) \\ &= \frac{k(n-k)}{n+1} e^{-kw} {}_2F_1(k+1, k+1; n+2; 1-e^{-w}) \\ &= f^0_{W_{n-k}}(w). \end{aligned} \tag{4.6}$$

This is an interesting property of the RRP tree in the limit. The inter-event times are symmetric, for instance the time it takes to go from $n-1$ to $n-2$ lineages, and the time it takes for the last lineage to die, have the same distribution.

To gain some insights into why this is true, consider the event times of the time-reversed Yule rate 1 process, which are distributed as the order statistics of n exponential rate 1 random variables, say $X_1 \leq X_2 \leq \ldots \leq X_n$. The form of equation (4.3) implies that in the limit $\psi \to 0$, the k-th event time T_k can be obtained via the transformation $T_k = \log(1/(\psi\lambda')) + \log\big(e^{X_k} - 1\big)$. If $X \sim \text{Exp}(1)$, then $\log\big(e^X - 1\big)$ has the standard logistic distribution (George and Mudholkar, 1981). It thus follows that, in the limit, the shifted event time defined as $T'_k := T_k - \log(1/(\psi\lambda'))$ is distributed as the k-th order statistic of n draws from the standard logistic distribution, which has pdf

$$f^0_{T'_{(1)}}(\tau') = \frac{e^{\tau'}}{(1+e^{\tau'})^2}. \tag{4.7}$$

Note that this is equivalent to saying that T_k is distributed as the k-th order statistic of n draws from the logistic distribution with location parameter (mode) $\log(1/(\psi\lambda'))$ and scale 1. The same conclusion can also be reached by considering the coalescent point process density (3.5), writing $\tau = \tau' + \log(1/(\psi\lambda'))$ and taking the limit $\psi \to 0$, which gives the density (4.7).

The limiting density of T'_k can also be obtained by applying the rescaling $\delta = (\lambda - \mu)\gamma$ and writing $\lambda' = \frac{\lambda}{\lambda - \mu}$ in the density (3.3),

$$f^{\psi}_{T_k}(\tau) = \binom{n}{k} k \, \frac{\psi\lambda' e^{\tau} [\psi\lambda'(e^{\tau} - 1)]^{k-1}}{[1 + \psi\lambda'(e^{\tau} - 1)]^{n+1}},$$

writing $T'_k = T_k - \log(1/(\psi\lambda'))$ and taking the limit $\psi \to 0$ gives

$$f^0_{T'_k}(\tau') = \lim_{\psi \to 0} \binom{n}{k} k \, \frac{e^{\tau'} [e^{\tau'} - \psi\lambda']^{k-1}}{[1 + e^{\tau'} - \psi\lambda']^{n+1}} = \binom{n}{k} k \, \frac{[e^{\tau'}]^k}{[1 + e^{\tau'}]^{n+1}}, \tag{4.8}$$

which, again, is the density of the k-th order statistic for the standard logistic distribution.

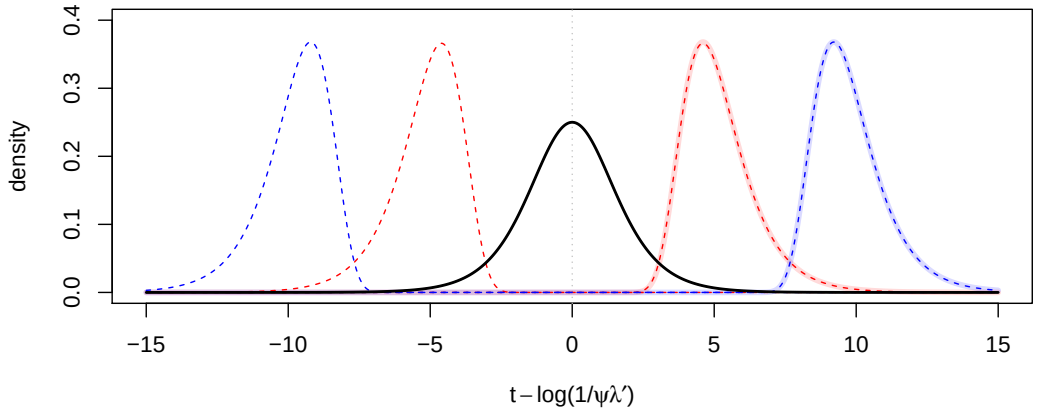

Figure 4: x-axis shows time shifted by $\log(1/(\psi\lambda'))$. Black solid line: standard logistic density (4.7). Dashed lines: density (4.8) of shifted time to first and last event for $n = 100$ (red) and $n = 10\ 000$ (blue). Faint solid lines: Gumbel density with parameters $(\log n, 1)$ for $n = 100$ (red) and $n = 10\ 000$ (blue).

As the logistic density (4.7) is symmetric around 0, the order statistics are also symmetric, with $T'_k \stackrel{d}{=} -T'_{n-k+1}$ (Arnold et al., 1992, pp. 26). This is illustrated in Figure 4: the black solid line shows the logistic density (4.7), and the red (blue) dashed lines show the densities of the first and last event times for $n = 100$ ($n = 10\ 000$). Thus, the densities of the event times T_k and T_{n-k+1} are symmetric around $\log(1/(\psi\lambda'))$.

Moreover, as $T'_{k+1} \stackrel{d}{=} -T'_{n-k}$, this demonstrates that the inter-event times $W_k = T_{k+1} - T_k = T'_{k+1} - T'_k$ and $W_{n-k} = T_{n-k+1} - T_{n-k} = T'_{n-k+1} - T'_{n-k}$ are equal in distribution. The density derived in Corollary 4.1 is hence that of the gap between the k-th and $(k+1)$-th order statistic of n standard logistic random variables. See for instance Mahmuod and Ragab (1973, pp. 84); their equation (4.1) gives the density of the gap between the k-th and $(k+1)$-th order statistics for the logistic distribution, which appears in very different form, but becomes the density in Corollary 4.1 after some algebra. We are not aware of a simpler expression for this particular density.

Corollary 4.2. *The distribution function of the waiting time W_k between events k and $k+1$, $k \in \{1, \ldots, n-1\}$, with $\psi \to 0$, is given by:*

$$F^0_{W_k}(w) = 1 - e^{-kw}\, {}_2F_1(k, k+1; n+1; 1 - e^{-w}). \tag{4.9}$$

Another interesting property of this distribution is that it does not depend on the scaled birth rate $\lambda' = \frac{\lambda}{\lambda - \mu}$, as this parameter only appears as a factor in $\psi\lambda'$. In particular, we can take $\lambda' = 1 \implies \mu' = 0$. Thus, the inter-event times for the RRP X^δ_ψ have the same distributions as those of an incompletely sampled time-reversed Yule rate 1 process, in the limit $\psi \to 0$.

4.4 Time to origin

We now consider the distribution of shifted time to origin $T'_n = T_n - \log(1/(\psi\lambda'))$ in the limit $\psi \to 0$. Integrating the density in (4.8) for $k = n$, the distribution function of T'_n is given by

$$F^0_{T'_n}(\tau') = (1 + e^{-\tau'})^{-n}.$$

As n grows, the density of time to origin shifts to the right away from $\log(1/(\psi\lambda'))$, so with high probability T'_n is much larger than 0. Figure 4 demonstrates this visually with examples of the density of T'_n for $n = 100$ and $n = 10\ 000$. Thus, for n large enough, this justifies introducing the approximation $1 + e^{-\tau'} \approx \exp\left(e^{-\tau'}\right)$, so the distribution of shifted time to origin can be approximated by

$$\widetilde{F}^0_{T'_n}(\tau') = \left[\exp\left(e^{-\tau'}\right)\right]^{-n} = \exp\left(-e^{-(\tau' - \log n)}\right).$$

This is a Gumbel distribution with location parameter (mode) $\log n$ and scale parameter 1. Figure 4 shows that this approximation provides a good fit, for $n = 100$ and $n = 10\ 000$.

This links to the results of Burden and Soewongsono (2019), who consider the diffusion limit (as the population size grows to infinity) of a near-critical Bienaymé-Galton-Watson process. Burden and Soewongsono (2019, Section 6) calculate numerically and plot the distribution of time to the MRCA, similarly shifted by the log of the population size at the time of sampling, and comment that as $n \to \infty$ this appears to converge to what looks like a Gumbel distribution. We have shown, analytically, that in the case of a supercritical birth-death process in the limit as $\psi \to 0$, the time to origin shifted by $\log(1/(\psi\lambda'))$ also converges to a Gumbel distribution, and in this case the location parameter depends on n.

4.5 Exponential approximation of inter-event times

Although Corollary 4.2 completely solves the question of what is the distribution of W_k as $\psi \to 0$, the appearance of ${}_2F_1$ in (4.9) somewhat obscures our insight into W_k. Here we show that these waiting times are well approximated by exponential distributions, so that the process is 'almost' Markov.

Consider an exponential approximation to $f^0_{W_k}(w)$ with rate $k(n-k)/n$:

$$\widetilde{f}^0_{W_k}(w) = \frac{k(n-k)}{n}\exp\left(-\frac{k(n-k)}{n}w\right), \qquad \widetilde{F}^0_{W_k}(w) = 1 - \exp\left(-\frac{k(n-k)}{n}w\right), \tag{4.10}$$

for $k \in \{1, \ldots, n-1\}$. We have the following result concerning the accuracy of this approximation:

Proposition 4.1. *Suppose the waiting time distribution W_k, with distribution function* (4.9) *for $\psi \to 0$, is approximated by an exponential distribution* (4.10)*. Then the approximation error is bounded, uniformly in k, in terms of Kolmogorov-Smirnov distance:*

$$\sup_w \left|F^0_{W_k}(w) - \widetilde{F}^0_{W_k}(w)\right| < \frac{1}{n}.$$

The density derived in Corollary 4.1 is nonintuitive, however this result shows that up to an error bounded by $1/n$, the distribution is actually approximately exponential. Note that the particular form of the exponential rate is such that $\widetilde{f}^0_{W_k}(w) = \widetilde{f}^0_{W_{n-k}}(w)$, so the symmetry between the k-th and $(n-k)$-th inter-event times is preserved in the approximation. Figure 5 shows an example of the (exact) density (4.4), for $\psi = 1$ on the left and very small ψ on the right; dotted lines in the latter case show the exponential approximations (4.10), demonstrating good agreement for $n = 100$.

Wiuf (2018) gives results for the expectation of time to origin, and recursions for calculating the expectation of the other event times, for the RRP with Bernoulli sampling (not in the limit $\psi \to 0$). We can use these results to show that the expectation under the exponential approximation, being $n/(k(n-k))$, is *exact* in the limit $\psi \to 0$ (for any n).

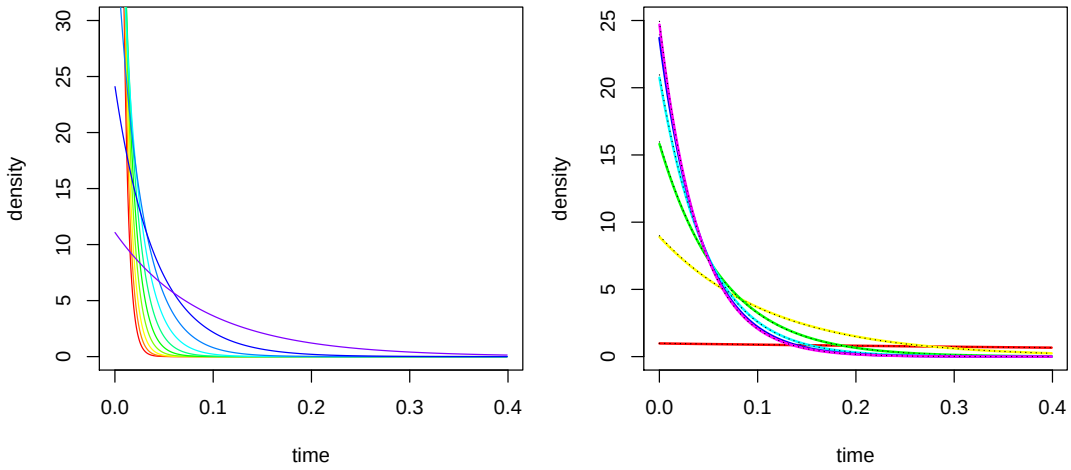

Figure 5: Inter-event time density, $n = 100$, $\lambda' = 2$, $\mu' = 1$. Left: with $\psi = 1$, colours (red to purple) correspond to event numbers $k = 0, 10, \ldots, 90$. Right: with $\psi = e^{-20}$, colours (red to purple) correspond to event numbers $k = 1, 10, 20, \ldots, 50$; dotted lines show exponential approximation (4.10).

Proposition 4.2. *The expectation of time to origin for $\psi \to 0$ is given by:*

$$\mathbb{E}(T_n) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1} \frac{1}{j} + \mathcal{O}(\psi). \tag{4.11}$$

This is an illuminating result, as the expectation is split into two parts. The first is $\log(1/(\psi\lambda'))$, corresponding to the first time rescaling regime, as described in Section 4.2. Near 0, a small step in t is equivalent to a step of order $\log(1/(\psi\lambda'))$ in units of δ. The second part is equivalent to the expectation of a sum of $n-1$ exponential waiting times with rate being the remaining number of lineages, corresponding to the second time rescaling regime, which is approximately linear.

This result agrees with the discussion in Section 4.3: recall that in the limit $\psi \to 0$, the shifted event time T'_k is distributed as the k-th order statistic of n standard logistic random variables, so $T_k = T'_k + \log(1/(\psi\lambda'))$ has expectation

$$\sum_{j=1}^{k-1} \frac{1}{j} - \sum_{j=1}^{n-k} \frac{1}{j} + \log\left(\frac{1}{\psi\lambda'}\right), \tag{4.12}$$

obtained by simplifying equation (4.8.6) in Arnold et al. (1992, p.82). Setting $k = n$, this becomes (4.11) up to the $\mathcal{O}(\psi)$ term. Notice also that using the Gumbel approximation for large n, as described in Section 4.4, gives the expectation of T'_n as $\log n + \widetilde{\gamma}$ (where $\widetilde{\gamma}$ is the Euler-Mascheroni constant). This is the limit of the harmonic sum in (4.11) as $n \to \infty$, so the expectations agree in this limit.

Wiuf (2018, Appendix D) derives a recursion for the expectations of event times, which in our notation becomes:

$$\mathbb{E}_n(T_k) = \frac{n}{n-k}\mathbb{E}_{n-1}(T_k) - \frac{k}{n-k}\mathbb{E}_n(T_{k+1}), \tag{4.13}$$

where $\mathbb{E}_n(T_k)$ denotes the expectation of the k-th event time if the sample is of size n at time 0. Using this and the expression for time to origin given by Proposition 4.2, we obtain the following result:

Proposition 4.3. *The expectation of waiting times between events is given by:*

$$\mathbb{E}(W_k) = \mathbb{E}(T_{k+1}) - \mathbb{E}(T_k) = \frac{n}{k(n-k)} + \mathcal{O}(\psi).$$

This agrees exactly with the expectation using the exponential approximation for $\psi \to 0$. This also agrees, up to the $\mathcal{O}(\psi)$ term, with the expectation of $T_{k+1} - T_k$ obtained using (4.12) in the limit $\psi \to 0$.

5 Discussion

In this paper, we have demonstrated that viewing the RRP as an inhomogeneous pure-death process allows for relatively simple and intuitive derivations of its properties. The time rescaling approach allows for results derived for completely sampled RRPs to be transformed to those for incomplete sampling, using a simple change of variables, with no restrictions on the parameter values. Moreover, the time rescaling between the time-reversed Yule rate 1 process and the RRP can be used to simulate the RRP in a straight forward way, by simulating each event time sequentially.

In the limit $\psi \to 0$, this rescaling can be decomposed into two timescales. The RRP tree becomes star-shaped, with terminal branch lengths tending to infinity, but inter-event times at the top of the tree are approximately exponential with a rate depending on n and the event number. This has interesting implications for data analysis, as it suggests that the number of singleton mutations in a small sample from a very large population tends to infinity, but the number of shared mutations does not. Indeed, the recent paper of Dinh et al. (2019) considers the expected frequency spectrum of mutations using a birth-death model with the infinite sites assumption. Although this is not explicitly discussed, the results of the simulations show that for small values of ψ, the expected number of singletons is orders of magnitude larger than that of mutations shared by multiple individuals. Taking the limit as $\psi \to 0$ in their equation (8), the expected number of singletons for X_ψ^δ grows to infinity, while for $k > 1$

$$\mathbb{E}\big[S_n(k)\big] = \theta \, \frac{n+k-1}{k(k-1)},$$

where θ is the mutation rate and $S_n(k)$ is the number of mutations with multiplicity k in the sample of size n. In applying their method to cancer data, Dinh et al. (2019) consider small values of ψ with the population size being very large compared to the sample size—our results presented in Section 4 provide an insight into the properties of the genealogy in this case.

As can be seen from our results and related work, properties of the genealogy of a sample obtained from a population following a birth-death process are notably different from those arising under the coalescent, particularly when the sample size is close to being of the same order as the population size. The coalescent is widely used in statistical inference for intra-host viral and bacterial populations (e.g. Dialdestoro et al., 2016). However, the choice of model should be appropriate to the relative scale of the biological application, and the individual-level population dynamics are arguably likely to be better modelled by a birth-death process. When considering the scenario of a small sample obtained from a very large population, the differences between the coalescent and the small-ψ limit of the birth-death model will carry through to the resulting inference. An important question is thus whether, for samples of viral or bacterial genetic sequencing data, birth-death models can provide better inference on the evolutionary dynamics of such populations. Answering this would require development of new methods for statistical inference that condition on the data and incorporate the natural processes governing such populations, such as high rates of mutation, recombination, and rapid demographic changes. This also

presents interesting challenges in making full use of the increasingly rich sequencing data available for viral and bacterial infections.

Acknowledgements

We thank two anonymous referees for their helpful comments. This work was supported by the OxWaSP CDT under the EPSRC grant EP/L016710/1, and by the Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

Aldous, D. (1996). Probability distributions on cladograms. In *Random Discrete Structures*, pp. 1–18. Springer.

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

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A first course in order statistics*, volume 54. Siam.

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes*. Springer-Verlag Berlin Heiderberg.

Bailey, N. T. (1964). *The elements of stochastic processes with applications to the natural sciences*. Wiley.

Boskova, V., Bonhoeffer, S. and Stadler, T. (2014). Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Computational Biology*, **10**(11), e1003913.

Burden, C. J. and Soewongsono, A. C. (2019). Coalescence in the diffusion limit of a Bienaymé–Galton–Watson branching process. *Theoretical Population Biology*, **130**, 50–59.

Dialdestoro, K., Sibbesen, J. A., Maretty, L., Raghwani, J., Gall, A., Kellam, P., Pybus, O. G., Hein, J. and Jenkins, P. A. (2016). Coalescent inference using serially sampled, high-throughput sequencing data from intrahost HIV infection. *Genetics*, **202**(4), 1449–1472.

Dinh, K. N., Jaksik, R., Kimmel, M., Lambert, A. and Tavaré, S. (2019). Statistical inference for the evolutionary history of cancer genomes. *bioRxiv*. doi:10.1101/722033. URL `https://www.biorxiv.org/content/early/2019/08/01/722033`.

Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G. and Bateman, H. (1953). *Higher transcendental functions*, volume 1. New York McGraw-Hill.

Fleischmann, K. and Siegmund-Schultze, R. (1977). The structure of reduced critical Galton-Watson processes. *Mathematische Nachrichten*, **79**(1), 233–241.

George, E. O. and Mudholkar, G. S. (1981). Some relationships between the logistic and the exponential distributions. In *Statistical Distributions in Scientific Work*, pp. 401–409. Springer.

Gernhard, T. (2008a). The conditioned reconstructed process. *Journal of Theoretical Biology*, **253**(4), 769–778.

Gernhard, T. (2008b). New analytic results for speciation times in neutral models. *Bulletin of Mathematical Biology*, **70**(4), 1082–1097.

Griffiths, R. C. and Tavare, S. (1994). Simulating probability distributions in the coalescent. *Theoretical Population Biology*, **46**(2), 131–159.

Grosjean, N. and Huillet, T. (2018). On the genealogy and coalescence times of Bienaymé–Galton–Watson branching processes. *Stochastic Models*, **34**(1), 1–24.

Harris, S. C., Johnston, S. G. and Roberts, M. I. (2017). The coalescent structure of continuous-time galton-watson trees. *arXiv preprint arXiv:1703.00299*.

Hartmann, K., Wong, D. and Stadler, T. (2010). Sampling trees from evolutionary models. *Systematic Biology*, **59**(4), 465–476.

Hein, J., Schierup, M. H. and Wiuf, C. (2005). *Gene genealogies, variation and evolution*. Oxford University Press.

Kaj, I. and Krone, S. M. (2003). The coalescent process in a population with stochastically varying size. *Journal of Applied Probability*, **40**(1), 33–48.

Kendall, D. G. (1948). On some modes of population growth leading to RA Fisher's logarithmic series distribution. *Biometrika*, **35**(1/2), 6–15.

Lambert, A. (2018). The coalescent of a sample from a binary branching process. *Theoretical Population Biology*, **122**, 30–35.

Lambert, A. and Stadler, T. (2013). Birth-death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology*, **90**, 113–128.

Mahmuod, M. and Ragab, A. (1973). On order statistics in samples drawn from the logistic distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, **4**(1), 81–88.

Meyer, P.-A. (1971). Demonstration simplifiee d'un theoreme de Knight. *Seminaire de Probabilites V Universite de Strasbourg, Lecture Notes in Mathematics*, **5**, 191–195.

Mooers, A., Gascuel, O., Stadler, T., Li, H. and Steel, M. (2012). Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Systematic Biology*, **61**(2), 195–203.

Nee, S., May, R. M. and Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **344**(1309), 305–311.

O'Connell, N. (1995). The genealogy of branching processes and the age of our most recent common ancestor. *Advances in Applied Probability*, **27**(2), 418–442.

Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, **165**, 483–506.

Paradis, E. and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.

Parsons, T. L., Quince, C. and Plotkin, J. B. (2010). Some consequences of demographic stochasticity in population genetics. *Genetics*, **185**, 1345–1354.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, **261**(1), 58–66.

Stadler, T. (2011). Simulating trees with a fixed number of extant species. *Systematic Biology*, **60**(5), 676–684.

Stadler, T. and Steel, M. (2012). Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *Journal of Theoretical Biology*, **297**, 33–40.

Stadler, T. and Steel, M. (2019). Swapping birth and death: Symmetries and transformations in phylodynamic models. *Systematic Biology*, **68**(5), 852–858.

Stadler, T., Vaughan, T. G., Gavryushkin, A., Guindon, S., Kühnert, D., Leventhal, G. E. and Drummond, A. J. (2015). How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics? *Proceedings of the Royal Society B: Biological Sciences*, **282**(1806), 20150420.

Thompson, E. A. (1975). *Human evolutionary trees*. CUP Archive.

Waugh, W. A. O. (1958). Conditioned Markov processes. *Biometrika*, **45**(1-2), 241–249.

Wiuf, C. (2018). Some properties of the conditioned reconstructed process with Bernoulli sampling. *Theoretical Population Biology*, **122**, 36–45.

Yang, Z. and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution*, **14**(7), 717–724.

A Proofs

A.1 Proof of Theorem 4.1

Proof. In the time-reversed Yule rate 1 process, the density of waiting times between the k-th and $k+1$-th event, $k = 0, \ldots, n-1$, conditional on $T_k = s$ is:

$$f_{W_k}(t|s) = (n-k)e^{-(n-k)((s+t)-s)} = (n-k)e^{-(n-k)t}. \tag{A.1}$$

Using the time transformation (4.1), in units of δ the waiting time is:

$$w = \rho_\delta^{-1}(s+t) - \rho_\delta^{-1}(s) = \log\left(\frac{\psi\lambda' + e^{s+t} - 1}{\psi\lambda' + e^s - 1}\right).$$

Rearranging, this gives:

$$t = \log\big(e^w\left[1 - (1-\psi\lambda')e^{-s}\right] + (1-\psi\lambda')e^{-s}\big),$$

and

$$\frac{dt}{dw} = \frac{e^w(1-(1-\psi\lambda')e^{-s})}{e^w(1-(1-\psi\lambda')e^{-s}) + (1-\psi\lambda')e^{-s}}.$$

Thus, by using a change of variables in (A.1) and writing $\phi = 1 - \psi\lambda'$:

$$\begin{aligned} f^{\psi}_{W_k}(w|s) &= (n-k)e^{w}\big[1-(1-\psi\lambda')e^{-s}\big]\big[e^{w}(1-(1-\psi\lambda')e^{-s})+(1-\psi\lambda')e^{-s}\big]^{-(n-k+1)} \\ &= (n-k)e^{w}\big[1-\phi e^{-s}\big]\big[e^{w}(1-\phi e^{-s})+\phi e^{-s}\big]^{-(n-k+1)}. \end{aligned}$$

Since s is the time of the k-th event in the time-reversed Yule rate 1 process, it has density given by (2.5):

$$f_{T_k}(s) = \binom{n}{k-1}(n-k+1)\big(1-e^{-s}\big)^{k-1}\big(e^{-s}\big)^{n-k+1}.$$

The marginal distribution of W_k is thus

$$\begin{aligned} f^{\psi}_{W_k}(w) &= \int_0^{\infty} f^{\psi}_{W_k}(w|s)\, f_{T_k}(s)ds \\ &= \binom{n}{k-1}(n-k+1)(n-k)\underbrace{\int_0^{\infty} e^{w}\frac{(1-\phi e^{-s})(1-e^{-s})^{k-1}(e^{-s})^{n-k+1}}{\big(e^{w}(1-\phi e^{-s})+\phi e^{-s}\big)^{n-k+1}}ds}_{(*)}. \end{aligned}$$

Integrating using the change of variables $u = e^{-s}$:

$$\begin{aligned} (*) &= e^{w}\int_0^1 \frac{(1-\phi u)(1-u)^{k-1}u^{n-k}}{\big(e^{w}(1-\phi u)+\phi u\big)^{n-k+1}}du \\ &= e^{-(n-k)w}\int_0^1 \frac{(1-\phi u)(1-u)^{k-1}u^{n-k}}{\big(1-\phi u(1-e^{-w})\big)^{n-k+1}}du \\ &= e^{-(n-k)w}\int_0^1 \left[\frac{(1-u)^{k-1}u^{n-k}}{\big(1-\phi u(1-e^{-w})\big)^{n-k+1}} - \phi\frac{(1-u)^{k-1}u^{n-k+1}}{\big(1-\phi u(1-e^{-w})\big)^{n-k+1}}\right]du. \end{aligned}$$

Using the following identity for the ordinary hypergeometric function (Abramowitz and Stegun, 1965, p.558):

$${}_2F_1(a,b,c,x) = \frac{\Gamma(c)}{\Gamma(c-a)\Gamma(a)}\int_0^1 \frac{(1-t)^{c-a-1}t^{a-1}}{(1-xt)^{b}}dt,$$

we obtain

$$\begin{aligned} (*) = e^{-(n-k)w}\frac{(k-1)!(n-k)!}{(n+1)!}\Big[&(n+1){}_2F_1(n-k+1,n-k+1;n+1;\phi(1-e^{-w})) \\ &-\phi(n-k+1){}_2F_1(n-k+1,n-k+2;n+2;\phi(1-e^{-w}))\Big]. \end{aligned}$$

Thus,

$$\begin{aligned} f^{\psi}_{W_k}(w) = \frac{(n-k)}{(n+1)}e^{-(n-k)w}\Big[&(n+1){}_2F_1(n-k+1,n-k+1;n+1;(1-\psi\lambda')(1-e^{-w})) \\ &-(1-\psi\lambda')(n-k+1){}_2F_1(n-k+1,n-k+2;n+2;(1-\psi\lambda')(1-e^{-w}))\Big]. \end{aligned}$$

For the RRP of a critical branching process, Z^{α}_{ψ}, the derivation is very similar. Using instead the time transformation

$$v = \rho_{\alpha}^{-1}(s+t) - \rho_{\alpha}^{-1}(s) = \frac{1}{\psi\lambda}\Big[e^{s+t}-1-e^{s}+1\Big] = \frac{1}{\psi\lambda}e^{s}(e^{t}-1)$$

and following the same steps, we obtain

$$\hat{f}^{\psi}_{W_k}(v) = \frac{(n-k+1)(n-k)}{n+1}\psi\lambda \cdot {}_2F_1(n-k+1, n-k+2; n+2; -\psi\lambda v).$$

□

A.2 Proof of Corollary 4.1

Proof. Substituting $\psi = 0$ into (4.4):

$$f^0_{W_k}(w) = \frac{(n-k)}{(n+1)}e^{-(n-k)w}\Big[(n+1){}_2F_1(n-k+1, n-k+1; n+1; 1-e^{-w}) \\ - (n-k+1){}_2F_1(n-k+1, n-k+2; n+2; (1-e^{-w}))\Big].$$

Identity (15.2.16) of Abramowitz and Stegun (1965, p. 558) gives:

$$ac(1-z)\ {}_2F_1(a+1, b; c; z) = c[a-(c-b)z]\ {}_2F_1(a, b; c; z) + (c-a)(c-b)z\ {}_2F_1(a, b; c+1; z) \qquad (A.2)$$

Substituting $a+1$ instead of a in identity (15.2.20) of Abramowitz and Stegun (1965, p. 558) gives:

$$c(1-z)\ {}_2F_1(a+1, b; c; z) = c\ {}_2F_1(a, b; c; z) - (c-b)z\ {}_2F_1(a+1, b; c+1; z). \qquad (A.3)$$

Multiplying (A.3) by a, equating with (A.2) and simplifying gives:

$$c\ {}_2F_1(a, b; c; z) - a\ {}_2F_1(b, a+1; c+1; z) = (c-a)\ {}_2F_1(a, b; c+1; z).$$

Thus, we obtain

$$f^0_{W_k}(w) = \frac{k(n-k)}{(n+1)}e^{-(n-k)w}{}_2F_1(n-k+1, n-k+1; n+2; 1-e^{-w}).$$

□

A.3 Proof of Corollary 4.2

Proof. By integrating the density in (4.6):

$$\begin{aligned}
F^0_{W_k}(w) &= \frac{k(n-k)}{n+1}\int_0^w e^{-ku}{}_2F_1(k+1, k+1; n+2; 1-e^{-u})du \\
&= \frac{k(n-k)}{n+1}\int_0^w e^{-u}e^{-(k-1)u}{}_2F_1(k+1, k+1; n+2; 1-e^{-u})du \\
&= \frac{k(n-k)}{n+1}\int_0^{1-e^{-w}} (1-z)^{k-1}{}_2F_1(k+1, k+1; n+2; z)dz \\
&= \frac{k(n-k)}{n+1}\Big[-\frac{(1-z)^k(n+1)}{k(n-k)}\ {}_2F_1(k, k+1; n+1; z)\Big]_0^{1-e^{-w}} \\
&= 1 - e^{-kw}{}_2F_1(k, k+1; n+1; 1-e^{-w}),
\end{aligned}$$

having used the substitution $z = 1 - e^{-u}$, and the identity (Erdélyi et al., 1953, p.102, eq. (25) with $n = 1$)

$$\int^z (1-x)^{a-2}{}_2F_1(a,b,c,x)\,dx = \frac{c-1}{(a-1)(b-c+1)}(1-z)^{a-1}{}_2F_1(a-1,b,c-1,z).$$

□

A.4 Proof of Proposition 4.1

Proof. Noting that

$$e^{-kw} = \exp\left(-\frac{k(n-k)}{n}w\right)\cdot\exp\left(-\frac{k^2}{n}w\right),$$

we have:

$$\begin{aligned}\left|\widetilde{F}^0_{W_k}(w) - F^0_{W_k}(w)\right| &= \left|1 - e^{-kw}{}_2F_1(k,k+1;n+1;1-e^{-w}) - 1 + \exp\left(-\frac{k(n-k)}{n}w\right)\right| \\ &= \exp\left(-\frac{k(n-k)}{n}w\right)\cdot\Bigg|\underbrace{\exp\left(-\frac{k^2}{n}w\right){}_2F_1(k,k+1;n+1;1-e^{-w})}_{=:\,h(w)} - 1\Bigg|. \qquad \text{(A.4)}\end{aligned}$$

We need to obtain an upper bound on the maximum of this distance. The first exponential term decays rapidly to 0, while $h(0) = 1$ and h initially increases; the global maximum of h occurs near $w = 0$, where $h(w) - 1 \geq 0$. We first obtain an upper bound on $h(w) - 1$, and then use this to obtain an upper bound on (A.4). Using the mean value theorem (or, equivalently, Taylor's theorem to first order):

$$h(w) = h(0) + wh'(c) = 1 + wh'(c)$$

for some $c \in (0, w)$, with

$$\begin{aligned}h'(c) = &-\frac{k^2}{n}\exp\left(-\frac{k^2}{n}c\right){}_2F_1(k,k+1;n+1;1-e^{-c}) \\ &+ \exp\left(-\frac{k^2+n}{n}c\right).\frac{k(k+1)}{n+1}{}_2F_1(k+1,k+2;n+2;1-e^{-c}).\end{aligned}$$

Differentiating again and considering the sign of the second derivative, we find that $h''(0) < 0$, so h' has a maximum at $c = 0$; h' has no other extrema before it reaches 0. We have:

$$h'(0) = -\frac{k^2}{n} + \frac{k(k+1)}{n+1} = \frac{k(n-k)}{n(n+1)},$$

so an upper bound on $h(w) - 1$ is given by

$$h(w) - 1 \leq \frac{k(n-k)}{n(n+1)}w.$$

Substituting this into (A.4):

$$\begin{aligned}\left|\widetilde{F}^0_{W_k}(w) - F^0_{W_k}(w)\right| &\leq \exp\left(-\frac{k(n-k)}{n}w\right) \cdot \big(h(w) - 1\big) \\ &\leq \exp\left(-\frac{k(n-k)}{n}w\right) \cdot \frac{k(n-k)}{n(n+1)}w. \end{aligned} \tag{A.5}$$

This attains the maximum at $\hat{w} = \frac{n}{k(n-k)}$. Substituting this into (A.5), we obtain the bound:

$$\left|\widetilde{F}^0_{W_k}(w) - F^0_{W_k}(w)\right| \leq \frac{1}{e(n+1)} < \frac{1}{n}.$$

The approximation error is thus bounded by $\frac{1}{n}$. □

A.5 Proof of Proposition 4.2

Proof. Wiuf (2018, Appendix F) derives an expression for the expectation of time to origin, which in our notation is:

$$\mathbb{E}(T_n) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{i=1}^{n}\frac{1}{i} - \sum_{i=1}^{n}\frac{1}{i}\frac{1}{\left(1-\frac{1}{\psi\lambda'}\right)^{n-i}} - \frac{1}{(1-\frac{1}{\psi\lambda'})^n}\log\left(\frac{1}{\psi\lambda'}\right). \tag{A.6}$$

The third term is:

$$\sum_{i=1}^{n}\frac{1}{i}\frac{1}{\left(1-\frac{1}{\psi\lambda'}\right)^{n-i}} = \frac{1}{n} + \frac{1}{n-1}\frac{1}{1-\frac{1}{\psi\lambda'}} + \frac{1}{n-2}\left(\frac{1}{1-\frac{1}{\psi\lambda'}}\right)^2 + \ldots = \frac{1}{n} + \mathcal{O}(\psi).$$

The fourth term in (A.6) is:

$$\begin{aligned}\frac{1}{(1-\frac{1}{\psi\lambda'})^n}\log\left(\frac{1}{\psi\lambda'}\right) &= (-\psi\lambda')^n(1-\psi\lambda')^{-n}\log\left(\frac{1}{\psi\lambda'}\right) \\ &= -(-\psi\lambda')^n[1+\mathcal{O}(\psi\lambda')]\log(\psi\lambda') \\ &= \mathcal{O}((\psi\lambda')^n\log(\psi\lambda)),\end{aligned}$$

which is $\mathcal{O}(\psi)$ for $n > 1$. In the limit $\psi \to 0$, we thus have

$$\mathbb{E}(T_n) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{i=1}^{n-1}\frac{1}{i} + \mathcal{O}(\psi).$$

□

A.6 Proof of Proposition 4.3

Proof. From Proposition 4.2, the expectation of time to origin for a sample of size n is:

$$\mathbb{E}_n(T_n) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} + \mathcal{O}(\psi),$$

which also implies that, for a sample of size $n-1$,

$$\mathbb{E}_{n-1}(T_{n-1}) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-2}\frac{1}{j} + \mathcal{O}(\psi).$$

We proceed by induction on the event number k, to show that

$$\mathbb{E}_n(T_k) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \sum_{j=k}^{n-1}\frac{n}{j(n-j)} + \mathcal{O}(\psi). \tag{A.7}$$

This holds for event number $k = n-1$, as using (4.13):

$$\begin{aligned}
\mathbb{E}_n(T_{n-1}) &= n\mathbb{E}_{n-1}(T_{n-1}) - (n-1)\mathbb{E}_n(T_n) \\
&= n\left(\log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \frac{1}{n-1}\right) - (n-1)\left(\log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j}\right) + \mathcal{O}(\psi) \\
&= \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \frac{n}{n-1} + \mathcal{O}(\psi).
\end{aligned}$$

Suppose that (A.7) holds for some $k = n-i$, $i \in \{1, \ldots, n-1\}$:

$$\mathbb{E}_n(T_{n-i}) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \sum_{j=1}^{i}\frac{n}{j(n-j)} + \mathcal{O}(\psi),$$

and so, equivalently, for $n-1$ lineages:

$$\mathbb{E}_{n-1}(T_{n-i-1}) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-2}\frac{1}{j} - \sum_{j=1}^{i}\frac{n-1}{j(n-j-1)} + \mathcal{O}(\psi).$$

Then:

$$\begin{aligned}
\mathbb{E}_n(T_{n-i-1}) &= \frac{n}{i+1}\mathbb{E}_{n-1}(T_{n-i-1}) - \frac{n-i-1}{i+1}\mathbb{E}_n(T_{n-i}) \\
&= \underbrace{\log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} + \mathcal{O}(\psi)}_{(*)} - \frac{n}{i+1}\left[\frac{1}{n-1} + \sum_{j=1}^{i}\frac{n-1}{j(n-j-1)} - (n-i-1)\sum_{j=1}^{i}\frac{1}{j(n-j)}\right] \\
&= (*) - \frac{n}{i+1}\left[\frac{1}{n-1} + \sum_{j=1}^{i}\left(\frac{1}{j} + \frac{1}{n-j-1}\right) - \frac{(n-i-1)}{n}\sum_{j=1}^{i}\left(\frac{1}{j} + \frac{1}{n-j}\right)\right] \\
&= (*) - \frac{1}{i+1}\left[\frac{n}{n-1} + (i+1)\sum_{j=1}^{i}\frac{1}{j} + n\sum_{j=1}^{i}\frac{1}{n-j-1} - (n-i-1)\sum_{j=1}^{i}\frac{1}{n-j}\right] \\
&= (*) - \frac{1}{i+1}\left[(i+1)\sum_{j=1}^{i}\frac{1}{j} + n\sum_{j=2}^{i}\frac{1}{n-j} + \frac{n}{n-i-1} - (n-i-1)\sum_{j=2}^{i}\frac{1}{n-j} + \frac{i+1}{n-1}\right] \\
&= (*) - \frac{1}{i+1}\left[(i+1)\sum_{j=1}^{i}\frac{1}{j} + (i+1)\sum_{j=1}^{i}\frac{1}{n-j} + \frac{(n-i-1)+(i+1)}{n-i-1}\right] \\
&= (*) - \frac{1}{i+1}\left[(i+1)\sum_{j=1}^{i+1}\frac{1}{j} + (i+1)\sum_{j=1}^{i+1}\frac{1}{n-j}\right] \\
&= \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \sum_{j=1}^{i+1}\frac{n}{j(n-j)} + \mathcal{O}(\psi).
\end{aligned}$$

Thus,

$$\mathbb{E}_n(T_k) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \sum_{j=1}^{n-k}\frac{n}{j(n-j)} + \mathcal{O}(\psi) = \log\left(\frac{1}{\psi\lambda'}\right) + \sum_{j=1}^{n-1}\frac{1}{j} - \sum_{j=k}^{n-1}\frac{n}{j(n-j)} + \mathcal{O}(\psi),$$

and so

$$\mathbb{E}(W_k) = \mathbb{E}(T_{k+1}) - \mathbb{E}(T_k) = \frac{n}{k(n-k)} + \mathcal{O}(\psi).$$

□

B Summary of RRPs

RRP	Y	Z_ψ^α	X_1^β	X_ψ^γ	X_ψ^δ
Time variable	t	$\alpha = \frac{1}{\lambda\psi}(e^t - 1)$ $t = \log(1 + \psi\lambda\alpha)$	$\beta = \frac{1}{\lambda-\mu}\log\Big(1 + \frac{\lambda-\mu}{\lambda}(e^t-1)\Big)$ $t = \log\Big(1 + \frac{\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\beta} - 1\Big)\Big)$	$\gamma = \frac{1}{\lambda-\mu}\log\Big(1 + \frac{1}{\psi}(e^{(\lambda-\mu)\beta} - 1)\Big)$ $\beta = \frac{1}{\lambda-\mu}\log\Big(1 + \psi(e^{(\lambda-\mu)\gamma} - 1)\Big)$	$\delta = (\lambda-\mu)\gamma$ $\gamma = \frac{1}{\lambda-\mu}\delta$
Corresponding complete process	Yule(1)	CBP(λ, ψ)	BDP$(\lambda, \mu, 1)$	BDP(λ, μ, ψ)	BDP(λ', μ', ψ) with $\lambda' = \frac{\lambda}{\lambda-\mu}, \mu' = \frac{\mu}{\lambda-\mu}$
m (death rate of the RRP, per lineage)	1	$\frac{\psi\lambda}{1+\psi\lambda\alpha}$	$\frac{\lambda e^{(\lambda-\mu)\beta}}{1 + \frac{\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\beta} - 1)}$	$\frac{\psi\lambda e^{(\lambda-\mu)\gamma}}{1 + \frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)\gamma} - 1)}$	$\frac{\psi\lambda' e^{\delta}}{1 + \psi\lambda'(e^{\delta} - 1)}$
$\rho = \int m$	t	$\log(1 + \psi\lambda\alpha)$	$\log\Big(1 + \frac{\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\beta} - 1\Big)\Big)$	$\log\Big(1 + \frac{\psi\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\gamma} - 1\Big)\Big)$	$\log\Big(1 + \psi\lambda'\Big(e^{\delta} - 1\Big)\Big)$
$e^{-\rho}$	e^{-t}	$\frac{1}{1+\psi\lambda\alpha}$	$\frac{1}{1 + \frac{\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\beta} - 1\Big)}$	$\frac{1}{1 + \frac{\psi\lambda}{\lambda-\mu}\Big(e^{(\lambda-\mu)\gamma} - 1\Big)}$	$\frac{1}{1 + \psi\lambda'\Big(e^{\delta} - 1\Big)}$