

# ON THE AGE OF A RANDOMLY PICKED INDIVIDUAL IN A LINEAR BIRTH-AND-DEATH PROCESS

FABIAN KÜCK \* \*\* AND

DOMINIC SCHUHMACHER, \* \*\*\* *University of Göttingen*

## Abstract

We consider the distribution of the age of an individual picked uniformly at random at some fixed time in a linear birth-and-death process. By exploiting a bijection between the birth-and-death tree and a contour process, we derive the cumulative distribution function for this distribution. In the critical and supercritical cases, we also give rates for the convergence in terms of the total variation and other metrics towards the appropriate exponential distribution.

*Keywords:* Age distribution; birth-and-death process; contour process

2010 Mathematics Subject Classification: Primary 60J80

Secondary 60E05; 62E20

## 1. Introduction

Linear birth-and-death processes are the most fundamental processes for modelling the evolution of a population. Being at the intersection of branching processes and general birth-and-death processes, they provide us with many useful properties we can exploit and much of their theory has been well-understood since at least the 1970s; see [1], [3], [5], and [6].

Suppose that  $(Y_t)_{t \geq 0}$  is a linear birth-and-death process with per-capita birth rate  $\lambda > 0$  and per-capita death rate  $\mu \geq 0$ . For simplicity, we set the initial population to 1. We denote the distribution of this process by  $\text{BDP}(\lambda, \mu)$ . To be able to track the ages of individuals, we equip the birth-and-death process with a *phylogeny* in form of a rooted binary tree whose edges have labels in  $\{0, 1\}$  indicating whether the individual ‘on that edge’ has just given birth and was alive before (0) or has itself just been born (1). Drawing time on the vertical axis and starting with a single branch with label 1, departing from the root at time 0 we split off a new branch with label 1 from an existing branch (the existing one subsequently gets label 0) each time a birth occurs and we terminate an existing branch each time a death occurs. The existing branch is chosen in both cases uniformly at random and independently from everything else. We may then recover the lifetime of any individual by a sequence of edges starting with a 1-edge followed forward in time by a number greater than or equal to 0 of 0-edges up to a leaf (death event).

It is straightforward to check that in such a tree each individual has an  $\text{Exp}(\mu)$ -distributed lifetime during which it gives birth according to a homogeneous Poisson process with rate  $\lambda$ , where all these lifetimes and processes are independent.

---

Received 23 February 2017; revision received 7 September 2017.

\* Postal address: Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany.

\*\* Email address: fabian.kueck@mathematik.uni-goettingen.de

\*\*\* Email address: schuhmacher@math.uni-goettingen.de

For some fixed time  $T > 0$ , we pick an individual (i.e. a branch) uniformly at random from all living individuals and denote its age by  $A = A_T$ . For simplicity, we refer to  $\mathcal{L}(A \mid Y_T > 0)$  as the *age distribution at time  $T$* . In the pure birth case, i.e. if  $\mu = 0$ , it is well known that the age distribution is a truncated exponential with parameter  $\lambda$ , a result which follows immediately from Theorem 1 of [7]. It is, furthermore, well known that  $\mathcal{L}(A \mid Y_T > 0) \xrightarrow{w} \text{Exp}(\lambda)$  as  $T \rightarrow \infty$  in the supercritical case ( $\lambda > \mu$ ). This follows, e.g. from [6, Example (6.10.14)].

However, somewhat surprisingly, if  $\mu > 0$ , an exact formula for the distribution of  $A$  at finite  $T$  is nowhere to be found in the literature. In this paper we provide such a formula for the cumulative distribution function (CDF) of  $A$ , both conditionally on the number  $Y_T$  of individuals at time  $T$  and unconditionally (the conditioning on  $Y_T > 0$  being always tacitly implied). Our main proof idea relies on a bijection between Galton–Watson trees in continuous time and exploration processes, recently shown in [2]. We also give upper bounds on the closeness of  $\mathcal{L}(A \mid Y_T > 0)$  and  $\text{Exp}(\lambda)$  in the critical and supercritical cases, as well as convergence rates in various metrics.

We point out some related work where other age distributions have been considered. First, the term ‘(actual) age distribution’ is sometimes encountered in the classic branching process literature from the 1970s (see, e.g. [6, Section 6.11]), referring to the random proportion  $Y_t^a / Y_t$ , where  $Y_t^a$  is the number of individuals alive at time  $t$  with age  $\leq a$ . The relation to the quantity we seek is

$$\mathbb{P}(A \leq a \mid Y_T > 0) = \mathbb{E}(\mathbb{P}(A \leq a \mid Y_T^a, Y_T, \{Y_T > 0\})) = \mathbb{E}\left(\frac{Y_T^a}{Y_T} \mid Y_T > 0\right).$$

However, it seems that no finite-time results are available for  $Y_T^a / Y_T$  but only limit results, such as [6, Corollary (6.10.5)], which implies the above-mentioned convergence towards  $\text{Exp}(\lambda)$  in the supercritical case.

Second, Theorem 2.5 of [4] can be used to show that the speciation times (times of vertices) in the *reconstructed tree* given there are  $n$  individuals at time  $T$  are independent and identically distributed (i.i.d.) and an explicit formula for their distribution can be given. In such a tree, any branches that are not required in order to connect the  $n$  individuals at time  $T$  to the root are omitted, so that the reconstructed tree has exactly  $n$  leaves in total.

Finally, Theorem 1 of [8] (more precisely, Theorem 3 in the ‘supplementary information’ of this paper) can be used in the special case  $m = 1, \mu_1 = 0, \varrho_1 = 1$  to provide the density of a linear birth-and-death tree, from which one can see that, at time  $T$ , the times since the nonzero birth times (including individuals that have died by time  $T$ ) are i.i.d. truncated exponentially distributed with parameter  $\lambda + \mu$ .

## 2. Results

From Theorem 2.1 below we state the distribution of the age conditioned on the population size. It contains the pure birth case ( $\mu = 0$ ) considered in Theorem 1 of [7] as a special case.

**Theorem 2.1.** *Let  $F_{y_T}$  denote the CDF of the age of an individual picked uniformly at random at time  $T$  given  $Y_T = y_T$  for some  $y_T > 0$ . Then  $F_{y_T}$  is given by*

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} \left( 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right) + \frac{1}{y_T} \left( \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbf{1}_{[0,t)}(t) + \mathbf{1}_{\{t=T\}} \right) \text{ for } t \in [0, T] \text{ if } \lambda \neq \mu$$

and by

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} \left( 1 - \frac{e^{-\lambda t}(T - t)}{T} \right) + \frac{1}{y_T} ((1 - e^{-\lambda t}(1 + \lambda t)) \mathbf{1}_{[0,t)}(t) + \mathbf{1}_{\{t=T\}})$$

for  $t \in [0, T]$  if  $\lambda = \mu > 0$ .

**Remark 2.1.** The proof of Theorem 2.1 yields a stronger result. Conditionally on  $Y_T = y_T$  and enumerating the individuals alive at time  $T$  in a specific way, which depends on the birth-and-death tree of the process, their ages are independent, where the first  $y_T - 1$  individuals have CDF  $F_*$  given by (3.7) below and the last individual has a different CDF  $F^*$  given by (3.9) below. Consequently, if we sample two different individuals uniformly from all individuals alive at time  $T$ , assuming  $y_T \geq 2$ , their ages are *dependent* draws from the above age distribution  $F_{y_T}$ .

The enumeration is such that the starting individual, if it is still alive at time  $T$ , always comes last. If we condition on the event that the starting individual has died by time  $T$ , the ages are still dependent, but come from a different mixture distribution  $((y_T - 1)/y_T)F_* + (1/y_T)F^{**}$ , where the new CDF  $F^{**}$  can be easily computed based on the proof of Theorem 2.1. On the other hand, if we condition on the event that the starting individual survives but sample only from all other individuals alive at time  $T$ , we obtain *independent* draws from  $F_*$ .

A simple computation yields the following unconditional age distribution.

**Corollary 2.1.** *The CDF  $F$  of the age distribution of  $(Y_t)_{t \geq 0}$  at time  $T$  is given by*

$$F(t) = \left( 1 - \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \lambda} \log \left( \frac{\lambda e^{(\lambda - \mu)T} - \mu}{\lambda - \mu} \right) \right) \left( 1 - \frac{e^{-\lambda t} - e^{-(\lambda - \mu)T} e^{-\mu t}}{1 - e^{-(\lambda - \mu)T}} \right) + \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \lambda} \log \left( \frac{\lambda e^{(\lambda - \mu)T} - \mu}{\lambda - \mu} \right) \times \left( \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbf{1}_{[0,t)}(t) + \mathbf{1}_{\{T\}}(t) \right) \text{ for } t \in [0, T] \text{ if } \lambda \neq \mu$$

and by

$$F(t) = \left( 1 - \frac{\log(1 + \lambda T)}{\lambda T} \right) \left( 1 - \frac{e^{-\lambda t}(T - t)}{T} \right) + \frac{\log(1 + \lambda T)}{\lambda T} ((1 - e^{-\lambda t}(1 + \lambda t)) \mathbf{1}_{\{t < T\}} + \mathbf{1}_{\{T\}}(t)) \text{ for } t \in [0, T] \text{ if } \lambda = \mu > 0.$$

The CDF of the age distribution is illustrated in Figure 1. Note that Corollary 2.1 immediately implies that the age distribution converges weakly to  $\text{Exp}(\lambda)$  if  $\lambda \geq \mu$  and to a mixture of  $\text{Exp}(\lambda)$  and  $\text{Exp}(\mu)$  if  $\lambda < \mu$ .

In the next corollary we bound the first discrepancy between the age distribution and  $\text{Exp}(\lambda)$  for  $\lambda \geq \mu$ . By introducing an explicit minimum over all couplings of  $A_* \sim \mathcal{L}(A \mid Y_T > 0)$  and  $Z \sim \text{Exp}(\lambda)$  and/or by integrating (or taking a supremum) over  $c > 0$ , this result can be easily turned into an upper bound for various metrics and pseudometrics.

**Corollary 2.2.** *There exist  $A_* \sim \mathcal{L}(A \mid Y_T > 0)$  and  $Z_* \sim \text{Exp}(\lambda)$  on a common probability space such that, for all  $c > 0$ ,*

$$\begin{aligned} \mathbb{E}|e^{-cA_*} - e^{-cZ_*}| &\leq \frac{\lambda}{\lambda + c} \frac{1}{e^{(\lambda - \mu)T} - 1} + \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \lambda} \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) \\ &= O(T e^{-(\lambda - \mu)T}) \text{ if } \lambda > \mu \end{aligned}$$

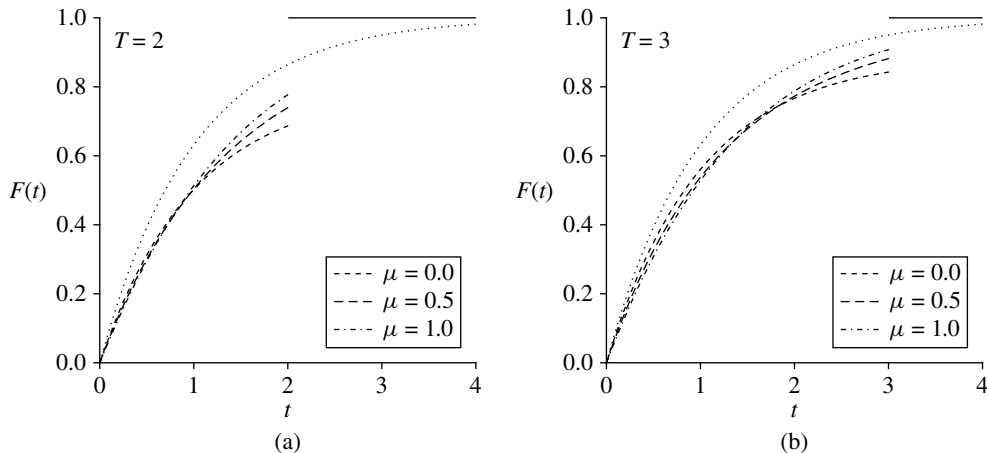


FIGURE 1: CDFs of the age distribution (*dashed lines*) and the asymptotic age distribution (*dotted line*) for  $\lambda = 1$ .

and

$$\mathbb{E}|e^{-cA_*} - e^{-cZ_*}| \leq \frac{c}{(\lambda + c)^2 T} + \frac{\log(1 + \lambda T)}{\lambda T} = O\left(\frac{\log(T)}{T}\right) \text{ if } \lambda = \mu > 0.$$

**Remark 2.2.** By slightly adapting the proof we can obtain a similar bound for the Wasserstein distance. In particular, we have, as  $T \rightarrow \infty$ ,

$$\begin{aligned} \mathcal{W}_1(\mathcal{L}(A \mid Y_T > 0), \text{Exp}(\lambda)) &= \mathbb{E}\left(\min_{A_* \sim \mathcal{L}(A \mid Y_T > 0), Z_* \sim \text{Exp}(\lambda)} |A_* - Z_*|\right) \\ &= \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu, \\ O\left(\frac{\log(T)}{T}\right) & \text{if } \lambda = \mu > 0. \end{cases} \end{aligned}$$

To give a complete picture, we also provide a convergence rate in terms of the total variation metric for  $\lambda \geq \mu$ .

**Corollary 2.3.** *We have, as  $T \rightarrow \infty$ ,*

$$d_{TV}(\mathcal{L}(A \mid Y_T > 0), \text{Exp}(\lambda)) = \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu, \\ O\left(\frac{\log(T)}{T}\right) & \text{if } \lambda = \mu > 0. \end{cases}$$

### 3. Proofs

First recall that the probabilities  $p_n(t)$  that  $Y_t = n$  are given as follows. For  $\lambda \neq \mu$ ,

$$\begin{aligned} p_0(t) &= \mu \tilde{p}(t), \\ p_n(t) &= (1 - \mu \tilde{p}(t))(1 - \lambda \tilde{p}(t))(\lambda \tilde{p}(t))^{n-1}, \quad n \in \mathbb{N}, \end{aligned} \tag{3.1}$$

where

$$\tilde{p}(t) := \frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu} = \frac{1}{\lambda} \frac{1 - e^{-(\lambda-\mu)t}}{1 - (\mu/\lambda)e^{-(\lambda-\mu)t}},$$

by Equations (8.15) and (8.46) of [3]. For  $\lambda = \mu > 0$ , these probabilities are given by

$$\begin{aligned}
 p_0(t) &= \frac{\lambda t}{1 + \lambda t}, \\
 p_n(t) &= \frac{(\lambda t)^{n-1}}{(1 + \lambda t)^{n+1}}, \quad n \in \mathbb{N},
 \end{aligned}
 \tag{3.2}$$

by Equation (8.53) of [3]. From this we easily obtain the following result.

**Lemma 3.1.** (i) For  $\lambda \neq \mu$ , we have

$$\begin{aligned}
 \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) &= \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \log\left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu}\right) \\
 &\leq \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right).
 \end{aligned}$$

(ii) For  $\lambda = \mu > 0$ , we have

$$\mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) = \frac{\log(1 + \lambda T)}{\lambda T}.$$

*Proof.* (i) For  $\lambda \neq \mu$ , we obtain

$$\begin{aligned}
 \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) &= \sum_{n=1}^{\infty} \frac{1}{n} \frac{p_n(T)}{1 - p_0(T)} \\
 &= \sum_{n=1}^{\infty} \frac{1}{n} \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} \left(\frac{\lambda e^{(\lambda-\mu)T} - \lambda}{\lambda e^{(\lambda-\mu)T} - \mu}\right)^{n-1} \\
 &= \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \log\left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu}\right) \\
 &\leq \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right).
 \end{aligned}$$

(ii) Analogously, we obtain, for  $\lambda = \mu > 0$ ,

$$\mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) = \sum_{n=1}^{\infty} \frac{1}{n} \frac{p_n(T)}{1 - p_0(T)} = \sum_{n=1}^{\infty} \frac{1}{n} \frac{(\lambda T)^{n-1}}{(1 + \lambda T)^n} = \frac{\log(1 + \lambda T)}{\lambda T}. \quad \square$$

*Proof of Theorem 2.1 and Corollary 2.1.* The expression for the CDF in Corollary 2.1 follows directly from conditioning on  $Y_T$  and applying Theorem 2.1 and Lemma 3.1.

Our proof of Theorem 2.1 relies on a recent result of [2] that shows that the distribution of the random binary tree generated by a birth-and-death process (see the introduction of that paper) can be obtained by inscribing a binary tree under the following *contour process* (also known as the *exploration process*).

Let  $(U_k)_{k \in \mathbb{N}_0}$  and  $(V_k)_{k \in \mathbb{N}_0}$  be two mutually independent sequences of independent and identically exponentially distributed random variables with parameters  $\mu$  and  $\lambda$ , respectively. Starting at  $(0, 0)$ , alternately add straight lines of slope  $+1$  and  $-1$  whose random heights are governed by the  $U$ - and  $V$ -variables, respectively. More precisely, for any  $k \in \mathbb{N}_0$ , given that

after  $2k$  steps of this procedure we are at position  $(x, y)$ , we add a straight line from  $(x, y)$  to  $(x + \min(U_k, T - y), y + \min(U_k, T - y))$ ; given that after  $2k + 1$  steps of this procedure we are at position  $(x, y)$ , we add a straight line from  $(x, y)$  to  $(x + \min(V_k, y), y - \min(V_k, y))$ . We stop the procedure when reaching a point  $(\tau, 0)$  with  $\tau > 0$ . The random function  $H = (H_x)_{0 \leq x \leq \tau}$  induced by this graph is what we call a *contour process* deflected at  $T$ . We denote its distribution by  $P_{\lambda, \mu, T}$ .

We inscribe a *left-aligned tree* under (the graph of) the contour process as follows. We denote by  $0 = W_1 < \dots < W_{2B_T+1} = \tau$  the points at which it has its local extrema, where  $B_T$  is the number of individuals born up to time 0 (including the one that existed at time 0). Note that  $(W_2, H_{W_2})$  is the leftmost local maximum of  $H$ . We draw a (vertical) line from  $(W_2, 0)$  to  $(W_2, H_{W_2})$ . The second leftmost local maximum is  $(W_4, H_{W_4})$  and the leftmost nonzero local minimum is  $(W_3, H_{W_3})$  if they exist. In this case we add a (vertical) line from  $(W_4, H_{W_3})$  to  $(W_4, H_{W_4})$  and a (horizontal) line from  $(W_4, H_{W_3})$  connecting the vertical line horizontally to the rest of the tree. We continue this procedure until we have explored all nonzero local extrema; see Figure 2.

Theorem 3.1 of [2] implies that the left-aligned tree obtained by this procedure has the same distribution  $Q_{\lambda, \mu, T}$  as the (left-aligned) tree obtained from a  $BDP(\lambda, \mu)$ -process (always assumed to be started with a single individual) killed at time  $T$ . This is essentially because each  $\text{Exp}(\mu)$ -distributed up-step in the contour process corresponds to the lifetime of an individual and each  $\text{Exp}(\lambda)$ -distributed down-step corresponds to the time at the death of an individual that has died since the last birth along its ancestry line (not including births of individuals belonging to this line). Note that horizontal edge lengths in the tree have no meaning but are simply adjusted to comply with the lines of the contour process.

The advantage of drawing the tree in a left-aligned fashion is that it allows us to always label the left branch by 0 and the right branch by 1 in order to obtain the birth and death times of each individual directly as lower and upper end points of vertical lines (which may be interrupted by further vertices). In particular, we may read off the ages of the individuals alive at time  $T$  as the total lengths of vertical lines that reach  $T$ ; see Figure 2.

The key observation is that now we may apply the inscription procedure described above also from right to left, i.e. we inscribe a *right-aligned tree* under the contour process by going through the locations  $W_1, W_2, \dots, W_{2B_T+1}$  of the local extrema in the reverse order, replacing each occurrence of the word ‘leftmost’ by ‘rightmost’ in the above description; see Figure 3. By interpreting vertical lines as before, we then obtain another tree that still has the correct distribution  $Q_{\lambda, \mu, T}$  in terms of birth and death times and phylogeny, but is now simply drawn

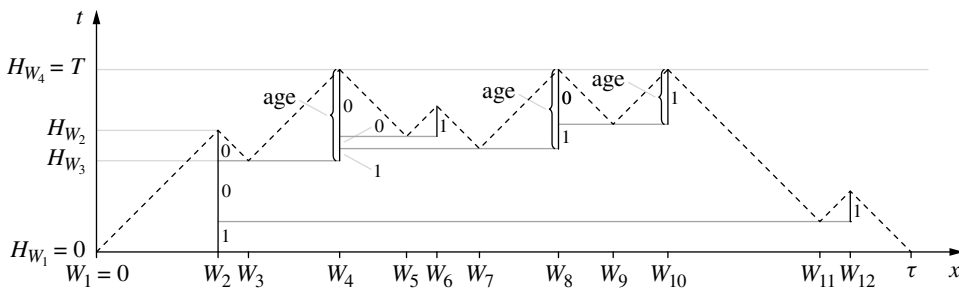


FIGURE 2: The contour process deflected at  $T$  and the inscribed left-aligned tree representing a linear birth-and-death process up to time  $T$ .

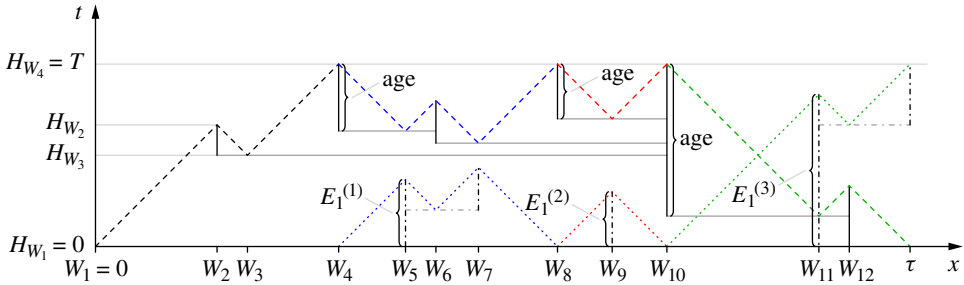


FIGURE 3: The deflected contour process, the inscribed *right-aligned* tree and the left-aligned trees corresponding to the mirrored excursions  $E^{(1)}$ ,  $E^{(2)}$ , and  $E^{(3)}$ .

in a right-aligned way. This is because applying the inscription procedure from right to left rather than left to right maintains the birth and death times exactly and maps the phylogeny bijectively onto another (typically different) phylogeny; since by construction the phylogeny of a linear birth-and-death process is uniformly distributed, it remains so after this mapping.

In the *right-aligned* tree, the ages of the individuals alive at time  $T$  correspond now to the down-steps from the local maxima at height  $T$ ; see Figure 3.

By the above construction of the contour process, the process  $(H_{W_k}, (-1)^k)_{k \in \mathbb{N}}$  is a Markov chain on  $[0, T] \times \{-1, 1\}$  with initial value  $(0, -1)$  at time 1. The second component simply keeps track of whether the contour process is at a local minimum (if it is  $-1$ ) or a local maximum (if it is  $1$ ).

We define the sequence  $(\xi_k)_{k \in \mathbb{N}}$  of hitting times in  $(T, 1)$  recursively by

$$\begin{aligned} \xi_1 &= \inf\{k \in \mathbb{N} : (H_{W_k}, (-1)^k) = (T, 1)\}, \\ \xi_{l+1} &= \inf\{k > \xi_l : (H_{W_k}, (-1)^k) = (T, 1)\} \quad \text{for all } l \in \mathbb{N}, \end{aligned}$$

where  $\inf \emptyset = \infty$ . Note that  $\max\{l : \xi_l < \infty\} = Y_T$ . Moreover, we define the ‘vertically’ mirrored (i.e. mirrored at a horizontal axis) excursions

$$E^{(l)} = (E_k^{(l)})_{0 \leq k \leq \xi_{l+1} - \xi_l} = (T - H_{W_k})_{\xi_l \leq k \leq \xi_{l+1}} \quad \text{for } l \in \mathbb{N},$$

where we set  $\infty - \infty = 0$ .

Since  $\mathbb{P}((H_{W_{\xi_l}}, (-1)^{\xi_l}) = (T, 1) \mid \xi_l < \infty) = 1$ , the strong Markov property and the fact that the second component of the process is deterministic imply that  $(H_{W_k})_{0 \leq k \leq \xi_l}$  and  $(H_{W_k})_{\xi_l \leq k}$  are independent given  $\xi_l < \infty$  and that  $(H_{W_k})_{\xi_l \leq k}$  has the same distribution given  $\xi_l < \infty$  for any  $l \in \mathbb{N}$ . As a consequence, we have, for any  $l \in \mathbb{N}$ ,  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $(E^{(l)}, (\xi_r)_{r \geq l+1})$  given  $\xi_l < \infty$  and, hence, that  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $E^{(l)}$  given  $\xi_l < \infty$  and given any sub- $\sigma$ -algebra of  $\sigma(\xi_r; r \geq l+1)$ . This implies that  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}, E^{(l)}$  are independent given  $\xi_l < \infty$  and given any sub- $\sigma$ -algebra of  $\sigma(\xi_r; r \geq l+1)$  for any  $l \in \mathbb{N}$ . Note that, for any  $l \in \mathbb{N}$ , the first  $l-1$  mirrored excursions  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  all have the same distribution under this conditioning since  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  all have the same distribution given  $\xi_l < \infty$  and  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $(\xi_r)_{r \geq l+1}$  given  $\xi_l < \infty$ .

Since  $\{Y_T = y_T\} = \{\xi_{y_T} < \infty, \xi_{y_T+1} = \infty\}$ , it follows that, given  $Y_T = y_T$ , the processes  $E^{(1)}, E^{(2)}, \dots, E^{(y_T-1)}, E^{(y_T)}$  are independent and  $E^{(1)}, E^{(2)}, \dots, E^{(y_T-1)}$  are identically distributed. By the strong Markov property used above, it is seen that neither the distribution

of  $E^{(1)}$  given  $Y_T = y_T \geq 2$  nor the distribution of  $E^{(y_T)}$  given  $Y_T = y_T \geq 1$  depend on the concrete value  $y_T$ .

On  $\{Y_T = y_T\}$ , the ages of the individuals alive at time  $T$  are just  $E_1^{(1)}, \dots, E_1^{(y_T-1)}$  and  $E_1^{(y_T)}$  (see Figure 3). In order to determine their conditional distributions given  $Y_T = y_T$ , we show that the mirrored excursions  $E^{(1)}, E^{(2)}, \dots, E^{(y_T)}$  correspond to independent copies of a contour processes  $(\tilde{H}_x)_{x \geq 0} \sim P_{\mu, \lambda, T}$  corresponding to the left-aligned tree of a BDP( $\mu, \lambda$ )-process (note the order of the parameters) killed at time  $T$  now. Let the positions  $(\tilde{W}_k)_{k \in \mathbb{N}}$  of the local extrema of  $(\tilde{H}_x)_{x \geq 0}$  be defined analogously to  $(W_k)_{k \in \mathbb{N}}$ . Both  $(\tilde{H}_{\tilde{W}_{k+1}})_{k \in \mathbb{N}_0}$  and  $(E_k^{(l)})_{k \in \mathbb{N}_0}$  for arbitrary  $l \in \{1, \dots, y_T\}$  start in 0 and alternate between independently adding  $\text{Exp}(\lambda)$  and subtracting  $\text{Exp}(\mu)$  random variables up to the first nonzero time where 0 or  $T$  is hit. So until this happens, they have the same distribution. Note that the contour process  $(\tilde{H}_x)_{x \geq 0}$  and, thus, also  $E^{(1)}, E^{(2)}, \dots, E^{(y_T)}$  correspond to the *left-aligned* tree of a BDP( $\mu, \lambda$ )-process killed at time  $T$  by Theorem 3.1 of [2].

We therefore obtain, for  $y_T \geq 2$ ,

$$\mathcal{L}(E_1^{(1)} \mid Y_T = y_T) = \mathcal{L}(E_1^{(1)} \mid E^{(1)} \text{ returns to 0 before reaching } T)$$

is the distribution of the lifetime of the starting individual in a BDP( $\mu, \lambda$ )-process conditioned on extinction of the process by time  $T$ . By the same argument, for  $y_T \geq 1$ , we see that

$$\mathcal{L}(E_1^{(y_T)} \mid Y_T = y_T) = \mathcal{L}(E_1^{(y_T)} \mid E^{(y_T)} \text{ reaches } T \text{ before returning to 0})$$

is the distribution of the lifetime (up to time  $T$ ) of the starting individual in a BDP( $\mu, \lambda$ )-process conditioned on survival of the process until time  $T$ .

Let  $(Z_t)_{t \geq 0} \sim \text{BDP}(\mu, \lambda)$  be equipped with a phylogeny as described in the introduction. Denote by  $L_1$  the lifetime of the starting individual and let  $F_*$  and  $F^*$  be the CDFs of  $L_1$  given  $Z_T = 0$  and of  $\min(L_1, T)$  given  $Z_T > 0$ , respectively. We then obtain, from the above, that the age distribution given  $Y_T = y_T > 0$  has CDF

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} F_*(t) + \frac{1}{y_T} F^*(t) \quad \text{for all } t \geq 0. \tag{3.3}$$

By Bayes' theorem,  $F_*$  has density

$$f_*(t) \propto \lambda e^{-\lambda t} \mathbb{P}(Z_T = 0 \mid L_1 = t) \quad \text{for } t \in [0, T).$$

Given  $L_1 = t$ , the birth times of the offspring of the starting individual form a Poisson process with rate  $\mu$ . By conditioning on the number of offspring and their birth times and substituting in their extinction probabilities as given by (3.1), we obtain

$$\begin{aligned} &\mathbb{P}(Z_T = 0 \mid L_1 = t) \\ &= \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} e^{-\mu t} \frac{1}{t^k} \\ &\quad \times \int_0^t \cdots \int_0^t \frac{\lambda}{\mu} \frac{1 - e^{-(\lambda-\mu)(T-t_1)}}{1 - (\lambda/\mu)e^{-(\lambda-\mu)(T-t_1)}} \cdots \frac{\lambda}{\mu} \frac{1 - e^{-(\lambda-\mu)(T-t_k)}}{1 - (\lambda/\mu)e^{-(\lambda-\mu)(T-t_k)}} dt_1 \cdots dt_k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} \left( \int_0^t \lambda \frac{e^{-(\lambda-\mu)(T-u)} - 1}{e^{-(\lambda-\mu)(T-u)} - \lambda/\mu} du \right)^k \end{aligned}$$



$$\begin{aligned}
 &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} (\log(\lambda e^{(\lambda-\mu)(T-t)} - \mu) + \lambda t - \log(\lambda e^{(\lambda-\mu)T} - \mu))^k \\
 &= \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t} \quad \text{for } t \in [0, T] \text{ if } \lambda \neq \mu.
 \end{aligned} \tag{3.4}$$

If  $\lambda = \mu > 0$ , the same argument using the extinction probabilities from (3.2) yields

$$\begin{aligned}
 &\mathbb{P}(Z_T = 0 \mid L_1 = t) \\
 &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \frac{1}{t^k} \int_0^t \dots \int_0^t \frac{\lambda(T-t_1)}{1+\lambda(T-t_1)} \dots \frac{\lambda(T-t_k)}{1+\lambda(T-t_k)} dt_1 \dots dt_k \\
 &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} \left( \int_0^t \lambda \frac{\lambda(T-u)}{1+\lambda(T-u)} du \right)^k \\
 &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} (\log(1+\lambda(T-t)) - \log(1+\lambda T) + \lambda t)^k \\
 &= \frac{1+\lambda(T-t)}{1+\lambda T} \quad \text{for } t \in [0, T].
 \end{aligned} \tag{3.5}$$

We may then compute the normalizing constant and obtain, for  $t \in [0, T]$ ,

$$f_*(t) = \begin{cases} \frac{\lambda e^{(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-\mu t}}{e^{(\lambda-\mu)T} - 1} = \frac{\lambda e^{-\lambda t} - \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} & \text{if } \lambda \neq \mu, \\ \frac{e^{-\lambda t} (1 + \lambda(T-t))}{T} & \text{if } \lambda = \mu > 0. \end{cases} \tag{3.6}$$

By integration, it follows that the CDF  $F_*$  for  $t \in [0, T]$  is given by

$$F_*(t) = \begin{cases} 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} & \text{if } \lambda \neq \mu, \\ 1 - \frac{e^{-\lambda t} (T-t)}{T} & \text{if } \lambda = \mu > 0. \end{cases} \tag{3.7}$$

It remains to derive  $F^*$ , which we do in a similar way. A slight notational complication arises from the fact that  $F^*$  has a discontinuity at  $T$ . We note that  $\min(L_1, T)$  has a density  $\tilde{f}$  with respect to the Lebesgue measure  $\text{Leb}_{[0, T]} + \delta_T$  given by

$$\tilde{f}(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0, t)}(t) + e^{-\lambda T} \mathbf{1}_{\{T\}}(t).$$

Thus, by Bayes' theorem,  $F^*$  has density  $f^*$  with respect to  $\text{Leb}_{[0, T]} + \delta_T$  satisfying

$$f^*(t) \propto \tilde{f}(t) \mathbb{P}(Z_T > 0 \mid L_1 = t) \quad \text{for } t \in [0, T].$$

By (3.4) and (3.5), we have

$$\begin{aligned}
 &\mathbb{P}(Z_T > 0 \mid L_1 = t) \\
 &= \begin{cases} 1 - \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t} = \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} & \text{if } \lambda \neq \mu, \\ 1 - \frac{1 + \lambda(T-t)}{1 + \lambda T} = \frac{\lambda t}{1 + \lambda T} & \text{if } \lambda = \mu > 0, \end{cases} \quad \text{for } t \in [0, T].
 \end{aligned}$$

Therefore, for any  $t \in [0, T]$ ,

$$f^*(t) \propto \begin{cases} \lambda e^{-\lambda t} \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} \mathbf{1}_{[0,t)}(t) + e^{-\lambda T} \mathbf{1}_{\{T\}}(t) & \text{if } \lambda \neq \mu, \\ \lambda e^{-\lambda t} \frac{\lambda t}{1 + \lambda T} \mathbf{1}_{[0,t)}(t) + e^{-\lambda T} \mathbf{1}_{\{T\}}(t) & \text{if } \lambda = \mu > 0. \end{cases}$$

Computing the normalizing constant, we obtain, for  $t \in [0, T]$ ,

$$f^*(t) = \begin{cases} \lambda \mu \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu} \mathbf{1}_{[0,t)}(t) + \frac{\lambda e^{-\mu T} - \mu e^{-\lambda T}}{\lambda - \mu} \mathbf{1}_{\{T\}}(t) & \text{if } \lambda \neq \mu, \\ \lambda^2 t e^{-\lambda t} \mathbf{1}_{[0,t)}(t) + (1 + \lambda T) e^{-\lambda T} \mathbf{1}_{\{T\}}(t) & \text{if } \lambda = \mu > 0. \end{cases} \tag{3.8}$$

By integration, it follows that the CDF  $F^*$  for  $t \in [0, T]$  is given by

$$F^*(t) = \begin{cases} \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbf{1}_{[0,t)}(t) + \mathbf{1}_{\{T\}}(t) & \text{if } \lambda \neq \mu, \\ (1 - e^{-\lambda t} (1 + \lambda t)) \mathbf{1}_{[0,t)}(t) + \mathbf{1}_{\{T\}}(t) & \text{if } \lambda = \mu > 0. \end{cases} \tag{3.9}$$

Substituting (3.7) and (3.9) into (3.3) yields the statement of Theorem 2.1. □

*Proof of Corollary 2.2.* It is enough to construct  $A_*$  and  $Z_*$  on  $\{Y_T > 0\}$  that have the correct distributions given  $Y_T > 0$  and are such that  $\mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T > 0)$  satisfies the required bound. We construct them explicitly as a quantile coupling using the notation from the proof of Theorem 2.1. Let  $J_T$  be a random variable that is uniformly distributed on  $\{1, \dots, y_T\}$  given  $Y_T = y_T > 0$ , independent from everything else, and let  $U$  be uniformly distributed on  $[0, 1]$  and also independent from everything else. Defining the generalized inverse of a CDF  $F$  by  $F^{-1}(u) = \inf\{t \in \mathbb{R}; F(t) \geq u\}$ , set

$$A_* := \mathbf{1}_{\{J_T < Y_T\}} F_*^{-1}(U) + \mathbf{1}_{\{J_T = Y_T\}} (F^*)^{-1}(U), \quad Z_* := F_\infty^{-1}(U),$$

where  $F_\infty$  denotes the CDF of  $\text{Exp}(\lambda)$ . By (3.3) and the independence of  $U$  from  $(Y_T, J_T)$ , we obtain  $\mathcal{L}(A_* \mid Y_T > 0) = \mathcal{L}(A \mid Y_T > 0)$  and  $\mathcal{L}(Z_* \mid Y_T > 0) = \text{Exp}(\lambda)$ , as required.

Since  $\lambda \geq \mu$ , we may verify directly from (3.7), noting that  $F_*(t) = 1$  for all  $t \geq T$ , that

$$F_*(t) > 1 - e^{-\lambda t} = F_\infty(t) \quad \text{for all } t > 0.$$

Therefore, since  $F_\infty$  is continuous and strictly increasing on  $\mathbb{R}_+$ ,

$$F_*^{-1}(U) \leq F_\infty^{-1}(U).$$

Thus, we may compute

$$\begin{aligned} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T > 0) &= \mathbb{E}\left(\frac{Y_T - 1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T < Y_T) \right. \\ &\quad \left. + \frac{1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T = Y_T) \mid Y_T > 0\right) \\ &\leq \mathbb{E}(e^{-cF_*^{-1}(U)} - e^{-cF_\infty^{-1}(U)}) + \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right). \end{aligned} \tag{3.10}$$

Note that we could employ a more sophisticated argument, and in doing so take care of the sign of  $e^{-c(F^*)^{-1}(U)} - e^{-cF_\infty^{-1}(U)}$  for the (inner) conditional expectation in the second line of (3.10). But since  $F^*(t) \not\rightarrow F_\infty(t)$  as  $T \rightarrow \infty$  for any  $t > 0$ , we would not gain anything in terms of the convergence rates; in particular, the factors  $T$  and  $\log(T)$  in the orders of the upper bound if  $\mu < \lambda$  and  $\mu = \lambda$ , respectively, cannot be removed.

It is now a matter of computing Laplace transforms. Since  $F_\infty^{-1}(U) \sim \text{Exp}(\lambda)$ , we have

$$\mathbb{E}(e^{-cF_\infty^{-1}(U)}) = \frac{\lambda}{\lambda + c}. \tag{3.11}$$

For  $F_*^{-1}(U)$ , we use the density from (3.6) and obtain, for  $\lambda > \mu$ ,

$$\begin{aligned} \mathbb{E}(e^{-cF_*^{-1}(U)}) &= \frac{1}{1 - e^{-(\lambda-\mu)T}} \int_0^T (\lambda e^{-(\lambda+c)t} - \mu e^{-(\lambda-\mu)T} e^{-(\mu+c)t}) dt \\ &= \frac{1}{1 - e^{-(\lambda-\mu)T}} \left( \frac{\lambda}{\lambda + c} (1 - e^{-(\lambda+c)T}) - \frac{\mu}{\mu + c} e^{-(\lambda-\mu)T} (1 - e^{-(\mu+c)T}) \right) \\ &\leq \frac{\lambda}{\lambda + c} \frac{e^{(\lambda-\mu)T}}{e^{(\lambda-\mu)T} - 1}, \end{aligned} \tag{3.12}$$

and, for  $\lambda = \mu > 0$ ,

$$\begin{aligned} \mathbb{E}(e^{-cF_*^{-1}(U)}) &= \frac{1}{T} \int_0^T e^{-(\lambda+c)t} (1 + \lambda(T - t)) dt \\ &= \frac{\lambda}{\lambda + c} + \frac{c(1 - e^{-(\lambda+c)T})}{(\lambda + c)^2 T} \\ &\leq \frac{\lambda}{\lambda + c} + \frac{c}{(\lambda + c)^2 T}. \end{aligned} \tag{3.13}$$

Combining (3.11)–(3.13) to bound the first summand on the right-hand side of (3.10) and employing Lemma 3.1 for the second summand, we obtain the required bounds.  $\square$

*Proof of Corollary 2.3.* Let  $f$  denote the density of  $\mathcal{L}(A \mid Y_T > 0)$  with respect to  $\nu_T = \text{Leb}_{[0,\infty)\setminus\{T\}} + \delta_T$ . Equation (3.3) implies that

$$f(t) = (1 - c_T) f_*(t) \mathbf{1}_{[0,T)}(t) + c_T f^*(t),$$

where  $f_*$  and  $f^*$  are given by (3.6) and (3.8), respectively, and

$$c_T = \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) = \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu, \\ O\left(\frac{\log(T)}{T}\right) & \text{if } \lambda = \mu, \end{cases} \tag{3.14}$$

by Lemma 3.1.

Noting that  $f_*(t) = f^*(t) = 0$  for  $t > T$ , we have

$$\begin{aligned} d_{\text{TV}}(\mathcal{L}(A \mid Y_T > 0), \text{Exp}(\lambda)) &= \frac{1}{2} \int_0^\infty |f(t) - \lambda e^{-\lambda t} \mathbf{1}_{[0,\infty)\setminus\{T\}}| \nu_T(dt) \\ &\leq \frac{1}{2} \left( (1 - c_T) \int_0^\infty |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^\infty |f^*(t) - \lambda e^{-\lambda t}| dt + c_T f^*(T) \right) \\ &\leq \frac{1}{2} \left( \int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt + O(Te^{-\lambda T}) \right). \end{aligned} \tag{3.15}$$

If  $\lambda > \mu$ , we obtain, for the first integral on the right-hand side,

$$\begin{aligned} \int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt &= \int_0^T \left| \frac{\lambda e^{-(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right| dt \\ &= \frac{e^{-(\lambda-\mu)T}}{1 - e^{-(\lambda-\mu)T}} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt \\ &= O(e^{-(\lambda-\mu)T}) \end{aligned} \tag{3.16}$$

and, for the second integral on the right-hand side,

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \frac{\lambda}{\lambda - \mu} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt = O(1). \tag{3.17}$$

In conclusion, substituting (3.16), (3.17), and (3.14) into (3.15), we have

$$d_{TV}(\mathcal{L}(A \mid Y_T > 0), \text{Exp}(\lambda)) = O(Te^{-(\lambda-\mu)T}).$$

If  $\lambda = \mu > 0$ , we obtain, for the first integral on the right-hand side of (3.15),

$$\int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt = \frac{1}{T} \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O\left(\frac{1}{T}\right) \tag{3.18}$$

and, for the second integral,

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \lambda \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O(1). \tag{3.19}$$

Thus, substituting (3.18), (3.19), and (3.14) into (3.15) yields

$$d_{TV}(\mathcal{L}(A \mid Y_T > 0), \text{Exp}(\lambda)) = O\left(\frac{\log(T)}{T}\right). \quad \square$$

### Acknowledgement

We would like to thank Thomas Rippl for drawing our attention to the connection between birth and death and contour processes.

### References

- [1] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.
- [2] BA, M., PARDOUX, E. AND SOW, A. B. (2012). Binary trees, exploration processes, and an extended Ray–Knight theorem. *J. Appl. Prob.* **49**, 210–225.
- [3] BAILEY, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences*. John Wiley, New York.
- [4] GERNHARD, T. (2008). The conditioned reconstructed process. *J. Theoret. Biol.* **253**, 769–778.
- [5] HARRIS, T. E. (1951). Some mathematical models for branching processes. In *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, Berkeley, CA, pp. 305–328
- [6] JAGERS, P. (1975). *Branching Processes with Biological Applications*. John Wiley, London.
- [7] NEUTS, M. F. AND RESNICK, S. I. (1971). On the times of births in a linear birthprocess. *J. Austral. Math. Soc.* **12**, 473–475.
- [8] STADLER, T., KÜHNERT, D., BONHOEFFER, S. AND DRUMMOND, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Nat. Acad. Sci.* **110**, 228–233.