

On the age of a randomly picked individual in a linear birth and death process

Fabian Kück^{1,2} and Dominic Schuhmacher^{1,3}

University of Göttingen

February 9, 2017

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 c.d.f. for this distribution. In the critical and supercritical cases we also give rates for the convergence in terms of total variation and other metrics towards the appropriate exponential distribution.

MSC 2010 subject classifications: Primary 60J80; secondary 60E05, 62E20.

Keywords: age distribution, birth and death process, contour process.

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 nice properties we can exploit and much of their theory has been well-understood since at least the 1970s; see [Har50, Bai64, AN72, Jag75].

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 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 been 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 zero, 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 ≥ 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 λ , where all these lifetimes and processes are independent.

¹Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany.

²E-mail: fabian.kueck@mathematik.uni-goettingen.de

³E-mail: 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 $\mathcal{L}(A | Y_T > 0)$ 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 λ , a result which follows immediately from Theorem 1 in [NR71]. It is furthermore well-known that $\mathcal{L}(A | Y_T > 0) \xrightarrow{w} \text{Exp}(\lambda)$ as $T \rightarrow \infty$ in the supercritical case ($\lambda > \mu$). This follows e.g. from [Jag75], 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 the present note, we provide such a formula for the c.d.f. 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 [BPS12]. We also give upper bounds on the closeness of $\mathcal{L}(A | Y_T > 0)$ 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. Firstly, the term “(actual) age distribution” is sometimes encountered in the classic branching process literature from the 1970s (see e.g. [Jag75], beginning of 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

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

However, it seems that no finite time results are available for Y_T^a/Y_T but only limit results, such as [Jag75], Corollary (6.10.5), which implies the above-mentioned convergence towards $\text{Exp}(\lambda)$ in the supercritical case.

Secondly, Theorem 2.5 in [Ger08] shows that the speciation times (times of vertices) in the *reconstructed tree* given there are n individuals at time T are iid and gives an explicit formula for their distribution. 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 in [SKBD13] (more precisely, Theorem 3 in the “supplementary information” of this paper) gives in the special case $m = 1$, $\mu_1 = 0$, $\varrho_1 = 1$ the density of a linear birth and death tree, from which one can see that, at time T , the times since the non-zero birth times (including individuals that have died by time T) are iid truncated exponentially distributed with parameter $\lambda + \mu$.

2 Results

Theorem 2.1 below gives the distribution of the age conditioned on the population size. It contains the pure birth case ($\mu = 0$) considered in Theorem 1 of [NR71] as a special case.

Theorem 2.1. *Let F_{y_T} denote the cumulative distribution function 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*

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

for $t \in [0, T]$ if $\lambda \neq \mu$ and by

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

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

Remark 2.2. 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 c.d.f. F_* given by (3.10) and the last individual has a different c.d.f. F^* given by (3.13). 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 $\frac{y_T-1}{y_T}F_* + \frac{1}{y_T}F^{**}$, where the new c.d.f. 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 2.3. The cumulative distribution function F of the age distribution of $(Y_t)_{t \geq 0}$ at time T is given by

$$F(t) = \left(1 - \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \log \left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu} \right) \right) \left(1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right) + \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \log \left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu} \right) \left(\frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0,t)}(t) + \mathbb{1}_{\{T\}}(t) \right)$$

for $t \in [0, T]$ if $\lambda \neq \mu$ and by

$$F(t) = \left(1 - \frac{\log(1 + \lambda T)}{\lambda T}\right) \left(1 - \frac{e^{-\lambda t}(T-t)}{T}\right) + \frac{\log(1 + \lambda T)}{\lambda T} \left((1 - e^{-\lambda t}(1 + \lambda t)) \mathbb{1}_{\{t < T\}} + \mathbb{1}_{\{T\}}(t) \right)$$

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

The cumulative distribution function of the age distribution is illustrated in Figure 1. Note that Corollary 2.3 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 $\lambda < \mu$.

The next corollary bounds a first kind of discrepancy between the age distribution and $\text{Exp}(\lambda)$ for $\lambda \geq \mu$. By introducing an explicit minimum over all couplings of $A_* \sim \mathcal{L}(A | Y_T > 0)$ 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 2.4. There exist $A_* \sim \mathcal{L}(A | Y_T > 0)$ and $Z_* \sim \text{Exp}(\lambda)$ on a common probability space such that for all $c > 0$

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

if $\lambda > \mu$ and

$$\mathbb{E} |e^{-cA_*} - e^{-cZ_*}| \leq \frac{c}{(\lambda + c)^2 T} + \frac{\log(1 + \lambda T)}{\lambda T} = O(\log(T)/T)$$

if $\lambda = \mu > 0$.

Remark 2.5. By slightly adapting the proof we can obtain a similar bound for the Wasserstein distance. In particular, we have, as $T \rightarrow \infty$,

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

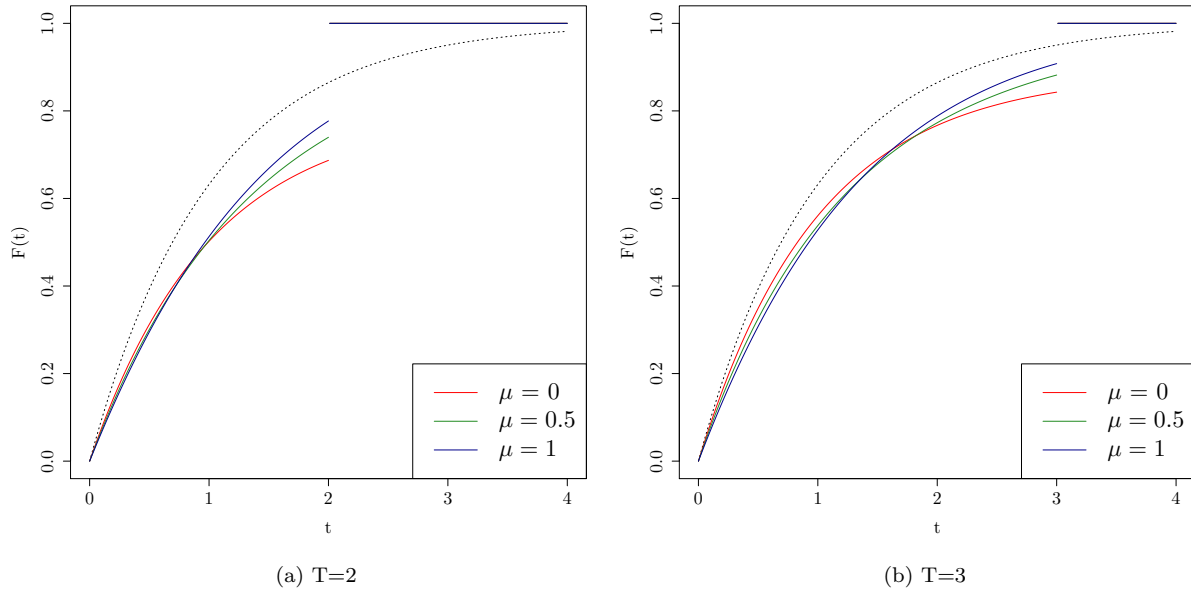


Figure 1: Cumulative distribution functions of the age distribution (solid lines) and the asymptotic age distribution (dotted line) for $\lambda = 1$.

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

Corollary 2.6. *We have, as $T \rightarrow \infty$,*

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

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

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

by (8.15) and (8.46) in [Bai64]. 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 (8.53) in [Bai64]. From this we easily obtain the following result.

Lemma 3.1.

(i) For $\lambda \neq \mu$, we have

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

(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

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

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

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

□

3.1 Proof of Theorem 2.1 and Corollary 2.3

The expression for the c.d.f. in Corollary 2.3 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 in [BPS12] that shows that the distribution of the random binary tree generated by a birth and death process (see introduction) can be obtained by inscribing a binary tree under the following *contour process* (also known as *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, identically exponentially distributed random variables with parameters μ and λ , 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 *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 non-zero local minimum is (W_3, H_{W_3}) if they exist. In

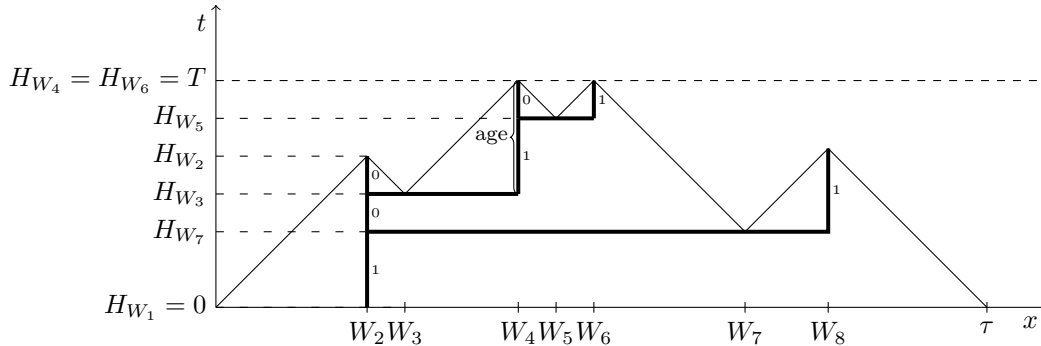


Figure 2: The contour process deflected at T and the inscribed left-aligned tree representing a linear birth and death process up to time T .

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 non-zero local extrema; see Figure 2.

Theorem 3.1 in [BPS12] implies that the tree obtained by this procedure has the same distribution $Q_{\lambda, \mu, T}$ as the (left-aligned) tree obtained from a BDP(λ, μ)-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 passed 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 we may now 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 point 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 now that 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. 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 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.

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 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

$$\xi_1 = \inf\{k \in \mathbb{N} : (H_{W_k}, (-1)^k) = (T, 1)\} \text{ and } \xi_{l+1} = \inf\{k > \xi_l : (H_{W_k}, (-1)^k) = (T, 1)\}$$

for all $l \in \mathbb{N}$, where $\inf \emptyset = \infty$. Note that $\max\{l : \xi_l < \infty\} = Y_T$. Moreover, we define the “horizontally

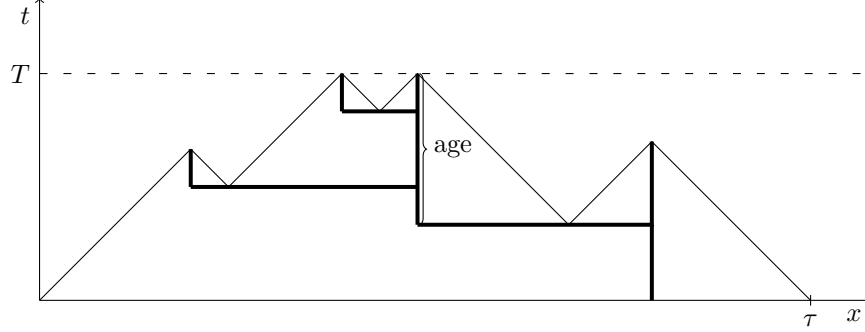


Figure 3: The deflected contour process and the inscribed right-aligned tree.

mirrored” excursions

$$E^{(l)} = (E_k^{(l)})_{0 \leq k \leq \xi_{l+1} - \xi_l} = (T - H_{W_k})_{\xi_l \leq k \leq \xi_{l+1}}$$

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 $l \in \mathbb{N}$ that $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$ 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- σ -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- σ -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 $l-1$ “horizontally mirrored” excursions $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$ all have the same distribution under this conditioning because $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$ 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 $\{Y_T = y_T\} = \{\xi_{y_T} < \infty, \xi_{y_T+1} = \infty\}$ we obtain 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 $Y_T = y_T \geq 1$ depends 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 $E_1^{(y_T)}$. We determine their conditional distributions given $Y_T = y_T$ by considering a contour process $(\tilde{H}_x)_{x \geq 0} \sim P_{\mu, \lambda, T}$ corresponding in [BPS12] to the left-aligned tree of a BDP(μ, λ)-process (note the order of the parameters!) killed at T by Theorem 3.1. 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 non-zero time where 0 or T is hit. So until this happens, they have the same distribution.

We therefore obtain for $y_T \geq 2$ that

$$\mathcal{L}(E_1^{(1)} \mid Y_T = y_T) = \mathcal{L}(E_1^{(1)} \mid E^{(1)} \text{ returns to 0 before reaching } T) \quad (3.3)$$

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

$$\mathcal{L}(E_1^{(y_T)} \mid Y_T = y_T) = \mathcal{L}(E_1^{(y_T)} \mid E^{(y_T)} \text{ reaches } T \text{ before returning to 0}) \quad (3.4)$$

is the distribution of the lifetime (up to time T) of the starting individual in a BDP(μ, λ)-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 c.d.f.s 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 c.d.f.

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} F_*(t) + \frac{1}{y_T} F^*(t) \quad (3.5)$$

for all $t \geq 0$.

By Bayes' Theorem, F_* has density

$$f_*(t) \propto \lambda e^{-\lambda t} \mathbb{P}(Z_T = 0 | L_1 = t), \quad (3.6)$$

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 μ . By conditioning on the number of offspring and their birth times and plugging in their extinction probabilities as given by Equation (3.1), we obtain

$$\begin{aligned} \mathbb{P}(Z_T = 0 | L_1 = t) &= \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} e^{-\mu t} \frac{1}{t^k} \int_0^t \cdots \int_0^t \frac{\lambda}{\mu} \frac{1 - e^{(\lambda-\mu)(T-t_1)}}{1 - \frac{\lambda}{\mu} e^{(\lambda-\mu)(T-t_1)}} \cdots \frac{\lambda}{\mu} \frac{1 - e^{(\lambda-\mu)(T-t_k)}}{1 - \frac{\lambda}{\mu} e^{(\lambda-\mu)(T-t_k)}} dt_1 \cdots dt_k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} \left(\int_0^t \lambda \frac{e^{-(\lambda-\mu)(T-u)} - 1}{e^{-(\lambda-\mu)(T-u)} - \frac{\lambda}{\mu}} du \right)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} \left(\log(\lambda e^{(\lambda-\mu)(T-t)} - \mu) + \lambda t - \log(\lambda e^{(\lambda-\mu)T} - \mu) \right)^k \\ &= \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t} \end{aligned} \quad (3.7)$$

for $t \in [0, T)$ if $\lambda \neq \mu$. If $\lambda = \mu > 0$, the same argument using the extinction probabilities from Equation (3.2) yields

$$\begin{aligned} \mathbb{P}(Z_T = 0 | L_1 = t) &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \frac{1}{t^k} \int_0^t \cdots \int_0^t \frac{\lambda(T-t_1)}{1 + \lambda(T-t_1)} \cdots \frac{\lambda(T-t_k)}{1 + \lambda(T-t_k)} dt_1 \cdots dt_k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} \left(\int_0^t \lambda \frac{\lambda(T-u)}{1 + \lambda(T-u)} du \right)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} (\log(1 + \lambda(T-t)) - \log(1 + \lambda T) + \lambda t)^k \\ &= \frac{1 + \lambda(T-t)}{1 + \lambda T}. \end{aligned} \quad (3.8)$$

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

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

By integration we obtain that the c.d.f. F_* for $t \in [0, T]$ is given by

$$F_*(t) = \begin{cases} 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} & \text{if } \lambda \neq \mu, \\ 1 - \frac{e^{-\lambda t} (T-t)}{T} & \text{if } \lambda = \mu > 0. \end{cases} \quad (3.10)$$

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 measure $\text{Leb}_{[0, T]} + \delta_T$ given by

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

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

$$f^*(t) \propto \tilde{f}(t) \mathbb{P}(Z_T > 0 | L_1 = t) \quad (3.11)$$

for $t \in [0, T]$. By (3.7) and (3.8) we have

$$\mathbb{P}(Z_T > 0 | L_1 = t) = \begin{cases} 1 - \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t} = \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} & \text{if } \lambda \neq \mu, \\ 1 - \frac{1 + \lambda(T-t)}{1 + \lambda T} = \frac{\lambda t}{1 + \lambda T} & \text{if } \lambda = \mu > 0, \end{cases}$$

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

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

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

$$f^*(t) = \begin{cases} \lambda \mu \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu} \mathbb{1}_{[0, t)}(t) + \frac{\lambda e^{-\mu T} - \mu e^{-\lambda T}}{\lambda - \mu} \mathbb{1}_{\{T\}}(t) & \text{if } \lambda \neq \mu, \\ \lambda^2 t e^{-\lambda t} \mathbb{1}_{[0, t)}(t) + (1 + \lambda T) e^{-\lambda T} \mathbb{1}_{\{T\}}(t) & \text{if } \lambda = \mu > 0. \end{cases} \quad (3.12)$$

By integration we obtain that the c.d.f. F^* for $t \in [0, T]$ is given by

$$F^*(t) = \begin{cases} \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0, t)}(t) + \mathbb{1}_{\{T\}}(t) & \text{if } \lambda \neq \mu, \\ (1 - e^{-\lambda t}(1 + \lambda t)) \mathbb{1}_{[0, t)}(t) + \mathbb{1}_{\{T\}}(t) & \text{if } \lambda = \mu > 0. \end{cases} \quad (3.13)$$

Plugging (3.10) and (3.13) into Equation (3.5) yields the statement of Theorem 2.1. \square

3.2 Proof of Corollary 2.4

It is enough to construct A_* and Z_* on $\{Y_T > 0\}$ that have the right distributions given $Y_T > 0$ and are such that $\mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| | Y_T > 0)$ satisfies the required bound. We construct them explicitly as a quantile coupling using 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 c.d.f. F by $F^{-1}(u) = \inf\{t \in \mathbb{R}; F(t) \geq u\}$, set

$$\begin{aligned} A_* &:= \mathbb{1}_{\{J_T < Y_T\}} F_*^{-1}(U) + \mathbb{1}_{\{J_T = Y_T\}} (F^*)^{-1}(U), \\ Z_* &:= F_\infty^{-1}(U), \end{aligned}$$

where F_∞ denotes the c.d.f. of $\text{Exp}(\lambda)$. By Equation (3.5) and the independence of U from (Y_T, J_T) we obtain $\mathcal{L}(A_* | Y_T > 0) = \mathcal{L}(A | Y_T > 0)$ and $\mathcal{L}(Z_* | Y_T > 0) = \text{Exp}(\lambda)$ as required.

Since $\lambda \geq \mu$, we may verify directly from (3.10), 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_∞ is continuous and strictly increasing on \mathbb{R}_+ ,

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

We may thus compute

$$\begin{aligned} & \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T > 0) \\ &= \mathbb{E}\left(\frac{Y_T - 1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T < Y_T) + \frac{1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T = Y_T) \mid Y_T > 0\right) \\ &\leq \mathbb{E}(e^{-cF_*^{-1}(U)} - e^{-cF_\infty^{-1}(U)}) + \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right). \end{aligned} \quad (3.14)$$

Note that we could employ a more sophisticated argument, taking care of the sign of $e^{-c(F_*)^{-1}(U)} - e^{-cF_\infty^{-1}(U)}$ for the second inner conditional expectation in line 2 of (3.14). 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

$$\mathbb{E}(e^{-cF_\infty^{-1}(U)}) = \frac{\lambda}{\lambda + c}. \quad (3.15)$$

For $F_*^{-1}(U)$ we use the density from (3.9) and obtain for $\lambda > \mu$

$$\begin{aligned} \mathbb{E}(e^{-cF_*^{-1}(U)}) &= \frac{1}{1 - e^{-(\lambda - \mu)T}} \int_0^T (\lambda e^{-(\lambda + c)t} - \mu e^{-(\lambda - \mu)T} e^{-(\mu + c)t}) dt \\ &= \frac{1}{1 - e^{-(\lambda - \mu)T}} \left(\frac{\lambda}{\lambda + c} (1 - e^{-(\lambda + c)T}) - \frac{\mu}{\mu + c} e^{-(\lambda - \mu)T} (1 - e^{-(\mu + c)T}) \right) \\ &\leq \frac{\lambda}{\lambda + c} \frac{e^{(\lambda - \mu)T}}{e^{(\lambda - \mu)T} - 1}, \end{aligned} \quad (3.16)$$

and for $\lambda = \mu > 0$

$$\mathbb{E}(e^{-cF_*^{-1}(U)}) = \frac{1}{T} \int_0^T e^{-(\lambda + c)t} (1 + \lambda(T - t)) dt = \frac{\lambda}{\lambda + c} + \frac{c(1 - e^{-(\lambda + c)T})}{(\lambda + c)^2 T} \leq \frac{\lambda}{\lambda + c} + \frac{c}{(\lambda + c)^2 T}. \quad (3.17)$$

Combining (3.15)–(3.17) to bound the first summand on the right hand side of (3.14) and employing Lemma 3.1 for the second summand, we obtain the required bounds. \square

3.3 Proof of Corollary 2.6

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 (3.5) implies that

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

where f_* and f^* are given in (3.9) and (3.12), respectively, and

$$c_T = \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) = \begin{cases} O(Te^{-(\lambda - \mu)T}) & \text{if } \lambda > \mu, \\ O(\log(T)/T) & \text{if } \lambda = \mu. \end{cases} \quad (3.18)$$

by Lemma 3.1.

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

$$\begin{aligned}
d_{TV}(\mathcal{L}(A | Y_T > 0), \text{Exp}(\lambda)) &= \frac{1}{2} \int_0^\infty |f(t) - \lambda e^{-\lambda t} \mathbb{1}_{[0, \infty) \setminus \{T\}}| \nu_T(dt) \\
&\leq \frac{1}{2} \left((1 - c_T) \int_0^\infty |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^\infty |f^*(t) - \lambda e^{-\lambda t}| dt + c_T f^*(T) \right) \\
&\leq \frac{1}{2} \left(\int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt + O(Te^{-\lambda T}) \right).
\end{aligned} \tag{3.19}$$

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

$$\begin{aligned}
\int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt &= \int_0^T \left| \frac{\lambda e^{-(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right| dt \\
&= \frac{e^{-(\lambda-\mu)T}}{1 - e^{-(\lambda-\mu)T}} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt = O(e^{-(\lambda-\mu)T}).
\end{aligned} \tag{3.20}$$

and for the second integral on the right-hand side

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \frac{\lambda}{\lambda - \mu} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt = O(1). \tag{3.21}$$

In conclusion, plugging (3.20), (3.21) and (3.18) into Inequality (3.19), we have that

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

If $\lambda = \mu > 0$, we obtain for the first integral on the right-hand side of Inequality (3.19)

$$\int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt = \frac{1}{T} \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O(1/T). \tag{3.22}$$

and for the second integral

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \lambda \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O(1). \tag{3.23}$$

Thus, plugging (3.22), (3.23) and (3.18) into Inequality (3.19) yields

$$d_{TV}(\mathcal{L}(A | Y_T > 0), \text{Exp}(\lambda)) = O(\log(T)/T).$$

□

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

References

- [AN72] Krishna B Athreya and Peter E Ney. *Branching processes*. Springer, New York, 1972.
- [Bai64] Norman TJ Bailey. *The elements of stochastic processes with applications to the natural sciences*. Wiley, New York, 1964.

- [BPS12] Mamadou Ba, Etienne Pardoux, and Ahmadou B Sow. Binary trees, exploration processes, and an extended Ray-Knight theorem. *Journal of Applied Probability*, 49(1):210–225, 2012.
- [Ger08] Tanja Gernhard. The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4):769–778, 2008.
- [Har50] Theodore E Harris. Some mathematical models for branching processes. Technical report, No. P-152, RAND Corporation, Santa Monica, 1950.
- [Har02] Theodore E Harris. *The Theory of Branching Processes*. Dover Publications, Mineola, New York, 2002.
- [Jag75] Peter Jagers. *Branching processes with biological applications*. Wiley, New York, 1975.
- [NR71] Marcel F Neuts and Sidney I Resnick. On the times of births in a linear birthprocess. *Journal of the Australian Mathematical Society*, 12(04):473–475, 1971.
- [SKBD13] Tanja Stadler, Denise Kühnert, Sebastian Bonhoeffer, and Alexei J Drummond. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, 110(1):228–233, 2013.