Part I

Dynamics of the most recent common ancestor in a critical branching population

Chapter 1

Introduction

Any asexually reproducing population has a unique most recent common ancestor (MRCA), from whom the entire population is descended. In sexually reproducing species, the same is true for each non-recombining piece of DNA. For instance, our "mitochondrial Eve" from whom all modern-day humans inherited their mitochondrial DNA is estimated to have lived around 180,000 years ago [160], while our "Y-chromosomal Adam" is estimated to have lived around 50,000 years ago [290]. There have also been efforts to estimate the time since the MRCA lived (which we will also call the "age of the MRCA") in populations of other organisms, particularly pathogens [294, 298]. These studies, using sophisticated models of demographic history, are focused on estimating the age of the MRCA at a single point in time (the present).

As time progresses into the future, eventually the mitochondrial lineages of all but one of the daughters of the current mitochondrial Eve will die out, at which point the new mitochondrial Eve will have lived somewhat later in time. The age of the MRCA is thus a dynamically evolving *process* that exhibits periods of upwards linear growth separated by downwards jumps.

Recently, [248] and [277] independently investigated the MRCA age process for the diffusion limit of the classical Wright-Fisher model. The Wright-Fisher model is perhaps the most commonly used model in population dynamics: each individual in a fixed size population independently gives birth to an identically distributed random number of individuals (with finite variance), and after the new offspring are produced some are chosen at random to survive so that the total population size remains constant. The diffusion limit arises by letting the population size go to infinity and taking the time between generations

to be proportional to the reciprocal of the population size.

In this work, we investigate the MRCA age process for a parametric family of population models in which the population size varies with time and, by suitable choice of parameters, allows control over the extent to which rare individuals can have large numbers of offspring that survive to maturity. The model for the dynamics of the population size is based on the *critical* $(1+\beta)$-*stable continuous state branching process* for $0 < \beta \le 1$. These processes arise as scaling limits of Galton-Watson branching processes as follows.

Write $Z_t^{(n)}$ for the number of individuals alive in a critical continuous time Galton-Watson branching process with branching rate λ and offspring distribution γ. The distribution γ has mean 1 (and thus, the process is "critical"). Suppose that if W is a random variable with distribution γ, then the random walk with steps distributed as the random variable $(W-1)$ falls into the domain of attraction of a stable process of index $1+\beta \in (1, 2]$. The case $\beta = 1$ corresponds to γ having finite variance and the random walk converging to Brownian motion after rescaling. Set $X_t^{(n)} = n^{-\frac{1}{\beta}} Z_t^{(n)}$ and suppose that $X_0^{(n)} \to x$ as $n \to \infty$. Then, up to a time-rescaling depending on λ and the scaling of the stable process above, the processes $X^{(n)}$ converge to a Markov process X that is a critical $(1+\beta)$-stable continuous state branching process, and whose distribution is determined by the Laplace transform

$$\mathbb{E}\left[e^{-\theta X_t} \,\middle|\, X_0 = x\right] = \exp\left(-\frac{\theta x}{(1+\theta^\beta t)^{1/\beta}}\right). \tag{1.0.1}$$

If $\beta = 1$, this is Feller's critical continuous state branching process [102, 192]. (Note that time here is scaled by a factor of 2 relative to some other authors, so that the generator of our "Feller continuous state branching process" is $x\frac{\partial^2}{\partial x^2}$.)

Let $\tau = \inf\{t > 0 : X_t = 0\}$ denote the *extinction time* of X (it is not hard to show that $X_t = 0$ for all $t \ge \tau$). Taking $\theta \to \infty$ in (1.0.1) gives

$$\mathbb{P}\{\tau > t \mid X_0 = x\} = 1 - \exp\left(-\frac{x}{t^{1/\beta}}\right),$$

so X dies out almost surely. However, it is possible to condition X to live forever in the following sense:

$$\lim_{T \to \infty} \mathbb{E}[f(X_t) \mid X_0 = x, \ \tau > T] = \frac{1}{x}\mathbb{E}[f(X_t)X_t \mid X_0 = x].$$

Thus, if $P_t(x', dx'')$ are the transition probabilities of X, then there is a Markov process Y with transition probabilities

$$Q_t(y', dy'') = \frac{1}{y'}P_t(y, y'')y''.$$

The process Y is the critical $(1+\beta)$-stable continuous state branching process X *conditioned on non-extinction.*

Using our expression (1.0.1) for the Laplace transforms of the unconditioned $(1+\beta)$-stable continuous state branching process, we see that the transition probabilities of the conditioned process Y are characterized by the Laplace transforms

$$
\begin{aligned}
\mathbb{E}^y[\exp(-\theta Y_t)] &= \frac{1}{y}\mathbb{E}^y[\exp(-\theta X_t)X_t] \\
&= \frac{1}{y}\frac{\partial}{\partial\theta}\mathbb{E}^y[\exp(-\theta X_t)] \\
&= \exp\left(-y\theta\left(t\theta^\beta+1\right)^{-1/\beta}\right)\left(t\theta^\beta+1\right)^{-\frac{\beta+1}{\beta}}.
\end{aligned}
\tag{1.0.2}
$$

Moreover, it is possible to start the process Y from the initial state $Y_0 = 0$, and the formula (1.0.2) continues to hold for $y = 0$. The super–process generalization of this construction was considered for $\beta = 1$ in [95, 93, 92] and for general β in [88].

For $\beta = 1$, the conditioned process Y can be described informally as a single "immortal particle" constantly throwing off infinitesimally small masses with each mass then evolving according to the dynamics of the unconditioned process. These infinitesimal masses can be interpreted as the single progenitors of families whose lineage splits from the immortal particle at the birth time of the progenitor and are eventually doomed to extinction. Most such families die immediately but a rare few live for a non-infinitesimal amount of time. More formally, there is a σ-finite measure ν on the space of continuous positive excursion paths

$$
\mathcal{E}^0 := \{u \in C(\mathbb{R}_+, \mathbb{R}_+) : u_0 = 0 \ \& \ \exists \gamma > 0 \text{ s.t. } u_t > 0 \ \Leftrightarrow \ 0 < t < \gamma\}
$$

such that if Π is a Poisson point process on $\mathbb{R}_+ \times \mathcal{E}^0$ with intensity $\lambda \otimes \nu$, where λ is Lebesgue measure, and $(\bar{X}_t)_{t\geq 0}$ is an independent copy of X begun at $\bar{X}_0 = y$, then the process

$$
\left(\bar{X}_t + \sum_{(s,u)\in\Pi} u_{(t-s)\vee 0}\right)_{t\geq 0}
\tag{1.0.3}
$$

has the same distribution as $(Y_t)_{t\geq 0}$ begun at $Y_0 = y$ — see [93]. A point $(s, u) \in \Pi$ corresponds to a family that grows to non-negligible size: the time s records the moment the family splits off from the immortal particle, and the value u_r of the trajectory u gives the size of the family τ units of time after it split off. The family becomes extinct after the period of time $\gamma(u) := \inf\{r > 0 : u(t) = 0, \ \forall t > r\}$. The σ-finite measure ν is Markovian

with transition probabilities the same as those of the unconditioned process X — in other words, ν arises from a family of entrance laws for the semigroup of X. The process $(\bar{X}_t)_{t\geq 0}$ records the population mass due to descendants of individuals other than the immortal particle who are present at time 0.

An analogous description of the conditioned process Y for the case $\beta \in (0,1)$ is presented in [88]. There is again a single immortal lineage, but now families split off from that lineage with a non-infinitesimal initial size, reflecting the heavy-tailed offspring distributions underlying these models. More precisely, a decomposition similar to (1.0.3) holds, but the Poisson point process Π is now on $\mathbb{R}_+ \times \mathcal{E}$ where $\mathcal{E} = \{u \in D(\mathbb{R}_+, \mathbb{R}_+) : \exists \gamma > 0 \text{ s.t. } u_t > 0 \iff 0 \leq t < \gamma\}$, the set of càdlàg paths starting above zero that eventually hit zero. The non-decreasing process $(M_t)_{t\geq 0}$, where

$$M_t := \sum_{(s,u)\in\Pi\cap[0,t]\times\mathcal{E}} u_0.$$

is the total of the initial family sizes that split off from the immortal particle in the time interval $[0,t]$. It is a stable subordinator of index β.

Suppose now that $\beta \in (0,1]$ is arbitrary. Take $Y_0 = 0$, so that $\bar{X}_t \equiv 0$ in the decomposition (1.0.3) and all "individuals" belong to families that split off from the immortal particle at times $s \geq 0$. Extend the definition of $\gamma(u)$ given above for $u \in \mathcal{E}_0$ to $u \in \mathcal{E}$ in the obvious way. The individuals besides the immortal particle alive at time $t > 0$ belong to families that correspond to the subset $\mathcal{A}_t := \{(s,u) \in \Pi : 0 \leq t - s < \gamma(u)\}$ of the random set Π. At time t, the amount of time since the most recent common ancestor of the entire population lived is $A_t := \sup\{t - s : (s,u) \in \mathcal{A}_t\}$. As depicted in Figure 1.1, the MRCA age process $(A_t)_{t\geq 0}$ has saw-tooth sample-paths that drift up with slope 1 until the current oldest family is extinguished, at which time they jump downward to the age of the next-oldest family.

It is not necessary to know the Poisson point process Π in order to construct the MRCA age process $(A_t)_{t\geq 0}$. Clearly, it is enough to know the point process Λ on $\mathbb{R}_+ \times \mathbb{R}_{++}$ given by $\Lambda := \{(s, \gamma(u)) : (s,u) \in \Pi\}$. Indeed, if we define the *left-leaning wedge with apex at* (t,x) to be the set

$$\triangle(t,x) := \{(u,v) \in \mathbb{R}^2 : u < t \ \& \ u + v > x + t\}, \tag{1.0.4}$$

then (see Figure 1.1)

$$A_t = t - \inf\{s : \exists x > 0 \text{ s.t. } (s,x) \in \Lambda \cap \triangle(t,0)\},$$

where A_t is defined to be 0 if $\Lambda \cap \triangle(t, 0)$ is empty.

Note that Λ is a Poisson point process with intensity $\lambda \otimes \mu$, where μ is the push-forward of ν by γ; that is, $\mu((t, \infty)) = \nu(\{u : \gamma(u) > t\})$. We will show in Section 4.1 that $\mu((t, \infty)) = (1 + \beta)/(\beta t)$ for the critical $(1 + \beta)$-stable continuous state branching process.

With these observations in mind, we see that if Λ is now **any** Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_{++}$ with intensity $\lambda \otimes \mu$, where μ is **any** measure on $\mathbb{R}_{++}$ with $\mu(\mathbb{R}_{++}) = \infty$ and $0 < \mu((x, \infty)) < \infty$ for all $x > 0$, then the construction that built $(A_t)_{t \geq 0}$ from the particular point process Λ considered above will still apply, and produce an $\mathbb{R}_+$-valued process with saw-tooth sample paths. We are therefore led to the following general definition.

Definition 1.0.1. let Λ be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_{++}$ with intensity measure $\lambda \otimes \mu$, where λ is Lebesgue measure and μ is a σ-finite measure on $\mathbb{R}_{++}$ with $\mu(\mathbb{R}_{++}) = \infty$ and $\mu((x, \infty)) < \infty$ for all $x > 0$. Define $(A_t)_{t \in \mathbb{R}_+}$ by

$$A_t := t - \inf\{s \geq 0 : \exists x > 0 \text{ s.t. } (s, x) \in \Lambda \cap \triangle(t, 0)\},$$

where $\triangle(t, 0)$ is defined by 1.0.4, and $A_t = 0$ if $\Lambda \cap \triangle(t, 0)$ is empty.

Note that an equivalent definition of A_t is

$$A_t = t - \inf\{s > 0 : |\triangle(t - s, s) \cap \Lambda| \geq 1\},$$

which leads more naturally to the generalization to multiple split times.

We will suppose from now on that we are in this general situation unless we specify otherwise. We will continue to use terminology appropriate for the genealogical setting and refer to $(A_t)_{t \geq 0}$ as the *MRCA age process* and μ as the *lifetime measure*. We will assume for convenience that the measure μ is absolutely continuous, with a density m with respect to Lebesgue measure that is positive Lebesgue almost everywhere. It is straightforward to remove these assumptions.

The strong Markov property of the Poisson point processes Λ implies that $(A_t)_{t \geq 0}$ is a time-homogeneous strong Markov process. In particular, there is a family of probability distributions $(\mathbb{P}^x)_{x \in \mathbb{R}_+}$ on the space of $\mathbb{R}_+$-valued càdlàg paths, with $\mathbb{P}^x$ interpreted in the usual way as the "distribution of $(A_t)_{t \geq 0}$ started from $A_0 = x$". More concretely, the probability measure $\mathbb{P}^x$ is the distribution of the process $(A_t^x)_{t \geq 0}$ defined as follows. Let Λ^x be a point process on $[-x, \infty) \times \mathbb{R}_{++}$ that has the distribution of the random point set

$\{(t-x,y):(t,y)\in\Lambda\}\cup\{(-x,Z)\}$, where Z is an independent random variable with values in the interval (x,∞) and distribution

$$\mathbb{P}\{Z\leq z\}=\mu((x,z])/\mu((x,\infty)).$$

Then we can define

$$A_t^x:=t-\inf\{s\geq-x:\exists y>0\text{ s.t. }(s,y)\in\Lambda^x\cap\triangle(t,0)\},\quad t\geq0$$

— see the proof of part (a) of Theorem 2.0.1 below for more detail. From now on, when we speak of the process $(A_t)_{t\geq0}$ we will either be referring to the process constructed as in Definition 1.0.1 from the Poisson process Λ on a probability space equipped with the probability measure $\mathbb{P}$ or the canonical process on the space of càdlàg $\mathbb{R}_+$-valued paths equipped with the family of probability measures $(\mathbb{P}^x)_{x\geq0}$. This should cause no confusion.

Note that A_t, the time back to the most recent common ancestor, is also the age of the oldest split in the genealogical tree of all individuals alive at time t. We can also consider the sequence of ages of the n oldest families to have split from the immortal particle. If we denote $A_t^{(k)}$ as the age of the kth oldest split, then $A_t^{(1)}=A_t$, and $A_t^{(1)},A_t^{(2)},A_t^{(3)},\ldots,A_t^{(n)}$ forms a nonincreasing sequence for each t, which can be defined as follows:

$$A_t^{(k)}=t-\inf\{s:|\triangle(t-s,s)\cap\Lambda|\geq k\},\quad\text{for }k\in\mathbb{N}_{>0}.\qquad(1.0.5)$$

It is straightforward to extend the considerations above to this sequence of split times, showing that for any n, the process $(A_t^{(1)},A_t^{(2)},A_t^{(3)},\ldots,A_t^{(n)})_{t\geq0}$ is a càdlàg strong Markov process on the set of nonincreasing sequences in $\mathbb{R}_{\geq0}$, and that the sequence will be strictly decreasing for all t almost surely.

In Chapter 2 we describe the transition probabilities and some sample path properties of of the split times. We focus on the properties of $(A_t)_{t\geq0}$, since although the results for larger collections of split times are interesting, the proofs add nothing new. We are able to derive explicit formulas for the transition probabilities, formulas for the stationary distribution when it exists and the condition under which it exists, bounds on convergence to stationarity, and necessary and sufficient conditions for whether or not A ever hits zero.

Using these results, in Chapter 4.1 we derive specific formulas for the dynamics of the MRCA process in our motivating example, the critical $(1+\beta)$-stable continuous state branching process. These results are not merely a straightforward substitution of the specific lifetime measure into the general formulas: in this case, the MRCA process does

not have a stationary version, but under a space-time rescaling, it transforms into another, stationary process, which is obtained by the same general "age" construction from a different point process. This observation allows us to derive additional "stationary" properties of A. Turning to larger collections of split times, the same space-time rescaling reveals two interesting continuous-time "fragmentation-coagulation-drift" processes on $[0,1]^k$ for each k, both of which converge to an interesting stationary distribution, in one case the distribution of a size-biased ordering of a Dirichlet$((1 + 1/\beta)/k, \ldots, (1 + 1/\beta)/k)$ random variable, and in the other, the distribution of independent Beta$(1 + 1/\beta, 1)$ random variables.

In Chapter 3 we describe other interesting properties of A. First we describe a duality property of the process under time–reversal. Then, note that the sample paths of $(A_t)_{t \geq 0}$ have local "peaks" immediately before jumps and local "troughs" immediately after. We also consider the discrete-time Markov chain of these, and describe a duality between the jump heights and inter-jump intervals.

We end this introduction by commenting on the connections with previous work. Firstly, we may think of each point $(s, x) \in \Lambda$ as a "job" that enters a queue with infinitely many servers at time s and takes x amount of time to complete. We thus have a classical $M/G/\infty$ queue [286], except we are assuming that the arrival rate is infinite. With this interpretation, the quantity A_t is how long the oldest job at time t has been in the queue. Many properties of such $M/G/\infty$ queues with infinite arrival rates have been studied (see, for example, [84]), but the age of the oldest job does not appear to have been studied in this context.

Secondly, as discussed in Section 4.2, the normalized split times $(A_t^{(1)}/t, A_t^{(2)}/t, \ldots)$ divide $[0,1]$ into a random partition whose marginal distribution at any time is the GEM$(1 + 1/\beta)$ distribution. There are several other similar models in the literature, often with connections to the GEM distribution. For instance, Donnelly & Tavaré in [68] study the partition induced by mutations in an infinite-alleles model, ordered by age of appearance, which has the GEM(θ) distribution, where θ is the rescaled mutation rate. They also note that this ordering has the same distribution as a size-biased ordering of the partition, a fact further discussed by Donnelly in [66], and which will also be true in our setting. Donnelly & Tavaré do not discuss the dynamics of the partition induced by mutations as time moves forwards, but such a process could be constructed. Tavaré [287] and Joyce & Tavaré [171] also studied the partition structure of a Yule process with immigration – which is related to our process in the case $\beta = 1$ – and showed that it had a stick-breaking construction. The

Yule process with immigration can be constructed as a Poisson process of families arising in time, similar to the processes we study. Karlin & McGregor in [178] also studied the partition structure induced by models of mutation of this form, in a somewhat more general context.

For each k, the dynamics of the intervals between the first k normalized split times has two components. Each interval drifts smaller, and furthermore at some rate each two adjacent intervals will merge, while the last interval splits in two. The marginal distribution of these intervals at any time is that of a size-biased ordering of a Dirichlet$((1 + 1/\beta)/k, \ldots, (1+1/\beta)/k)$ random variable, and is simliar in character to others models studied in the literature. There is a substantial literature on "coagulation-fragmentation" models (most of which do not include a "drift" component) and we provide only a few pointers to the relevant research. For instance, Durrett, Granovsky, & Gueron [79] studied a large class of reversible coagulation-fragmentation processes, and give a comprehensive review of the history of such models in the literature. More recently, Pitman [251] and Diaconis, Meyer-Wolf, Zeitouni, & Zerner [225, 65], each studied coagulation-fragmentation processes whose equilibrium distrubtion was given by the Poisson-Dirichlet distribution. In 2008, Bertoin [17] showed that the two-parameter Poisson-Dirichlet distribution introduced in [244] is the unique reversible distribution for a natural class of *exchangeable fragmentation-coalescence* models, which were introduced by Berestycki in [11].

Furthermore, note that the process $(A_t)_{t \geq 0}$ is an example of a *piecewise deterministic Markov process*: it consists of deterministic flows punctuated by random jumps. Such processes were introduced in [59] and studied further in [60] (see also [165], where the nomenclature *jumping Markov processes* is used). The general properties of such processes have been studied further in, for example, [76, 55, 50].

Lastly, piecewise deterministic Markov processes like $(A_t)_{t \geq 0}$ that have periods of linear increase interspersed with random jumps have been used to model many phenomena, such as stress in an earthquake zone [28], congestion in a data transmission network [77], and growth-collapse [29]. They also have appeared in the study of the additive coalescent [96] and $\mathbb{R}$-tree-valued Markov processes [97]. The scaling limit of the length of loop-erased random walk on the complete graph (see Pittel [259] and Alappattu & Pitman [3]) is another such process with saw-tooth sample paths, and futhermore arises from a Poisson process. However, it does not seem possible to find an MRCA age process whose distribution is that of the length of the LERW.
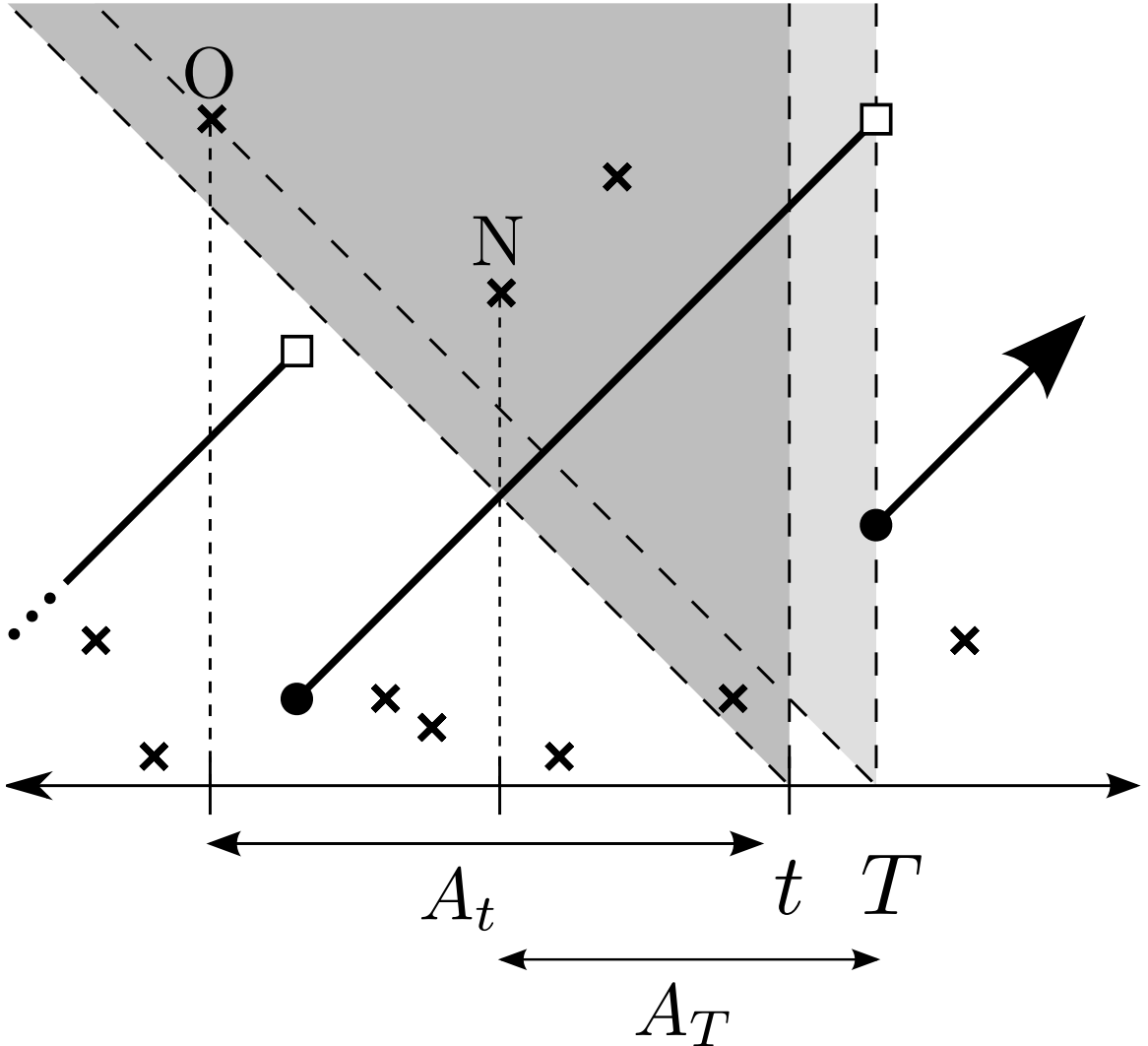
Figure 1.1: In this diagram, the points of Λ are marked "x", the sample path of the process A is drawn with a solid line, the oldest extant family at time t is represented by the point in Λ labeled "O", and the left-leaning wedge $\triangle(t, 0)$ is the darkly shaded region with apex at $(t, 0)$. At time T, the family represented by O will die out and the family represented by N will become the oldest extant family, as we can see by the fact that O is on the boundary of (the lightly shaded) $\triangle(T, 0)$. If the coordinates of O are (u, y) and the coordinates of N are (v, z), then the size of the jump that A makes at T is $A_T - A_{T-} = u - v = (T - v) - (T - u) = (T - v) - y$.

Chapter 2

General lifetime measures

In this chapter, we present explicit formulas for many properties associated with the MRCA process and larger collections of split times, defined in terms of a Poisson point process as described in the introduction. First we present results about the MRCA process $(A_t)_{t \geq 0}$, defined in 1.0.1, and follow this with more general results about larger collections of split times.

Theorem 2.0.1. *(a) The transition probabilities of the time-homogeneous Markov process A have an absolutely continuous part*

$$\mathbb{P}^x\{A_t \in dy\} = \frac{\mu((x, x+t])}{\mu((x, \infty))} \exp\left(-\int_y^{x+t} \mu((z, \infty))\, dz\right) \mu((y, \infty))\, dy,$$

for $y < x + t$ and a single atom

$$\mathbb{P}^x\{A_t = x + t\} = \frac{\mu((x+t, \infty))}{\mu((x, \infty))}.$$

(b) The total rate at which the process A jumps from state $x > 0$ is

$$\frac{m(x)}{\mu((x, \infty))},$$

and when the process jumps from state $x > 0$, the distribution of the state to which it jumps is absolutely continuous with density

$$y \mapsto \exp\left(-\int_y^x \mu((z, \infty))\, dz\right) \mu((y, \infty)), \quad 0 < y < x.$$

(c) *The probability* $\mathbb{P}^0\{\exists t > 0 : A_t = 0\}$ *that the process* A *returns to the state zero is positive if and only if*

$$\int_0^1 \exp\left(\int_x^1 \mu((y, \infty))\, dy\right) dx < \infty.$$

(d) *If*

$$\int_1^\infty \exp\left(-\int_1^x \mu((y, \infty))\, dy\right) dx = \infty,$$

then for each $x > 0$ *the set* $\{t \geq 0 : A_t = x\}$ *is* $\mathbb{P}^x$-*almost surely unbounded. Otherwise,* $\lim_{t \to \infty} A_t = \infty$, $\mathbb{P}^x$-*almost surely, for all* $x \geq 0$.

(e) *A stationary distribution* π *exists for the process* A *if and only if*

$$\int_1^\infty \mu((z, \infty))\, dz < \infty,$$

in which case it is unique, and

$$\pi(dx) = \mu((x, \infty)) \exp\left(-\int_x^\infty \mu((z, \infty))\, dz\right) dx.$$

(f) *If* A *has a stationary distribution* π, *then*

$$d_{TV}(\mathbb{P}^x\{A_t \in \cdot\}, \pi) \leq 1 - \exp\left(-\int_{t+x}^\infty \mu((y, \infty))\, dy\right) \times \frac{\mu([x, x+t))}{\mu([x, \infty))}.$$

where d_{TV} *denotes the total variation distance. In particular, the distribution of* A_t *under* $\mathbb{P}^x$ *converges to* π *in total variation as* $t \to \infty$.

Recall the definition of $A_t^{(k)}$ from 1.0.5, as the age of the kth oldest family to have split from the immortal particle. Let $0 < x_1 < \cdots < x_k < t$, and let $(y_1, \ldots, y_k)$ be such that $0 < y_1 < \cdots < y_k < t+s$ and for some $m \leq k$, $\{y_1 - s, y_2 - s, \ldots, y_m - s\} \subset \{x_1, x_2, \ldots, x_k\}$. So, if $(A_t^{(1)}, \ldots, A_t^{(k)}) = (x_1, \ldots, x_k)$ and $(A_{t+s}^{(1)}, \ldots, A_{t+s}^{(k)}) = (y_1, \ldots, y_k)$, then m of the k oldest families alive at time t are still alive at time $t + s$, while the remainder have been replaced by younger families. The following theorem gives the probability of this occurrence, and therefore the transition probabilities of $(A_t^{(1)}, \ldots, A_t^{(k)})$.

Theorem 2.0.2. *Suppose that* $0 < x_1 < \cdots < x_k < t$. *Let* $i_1 < \cdots < i_m$ *be distinct integers in* $\{1, \ldots, k\}$, *let* $\ell_1 < \ldots < \ell_{k-m}$ *be such that* $\{1, \ldots, k\} = \{i_1, \ldots, i_m\} \cup \{\ell_1, \ldots, \ell_{k-m}\}$,

and let $x_{i_m} + s < y_1 < y_2 < \cdots < y_{k-m} < t + s$. *Then the joint transition probabilities have density with respect to the appropriate Lebesgue measure*

$$\mathbb{P}\{A_{t+s}^{(j)} = x_{i_j} + s, A_{t+s}^{(m+r)} = y_r, \ 1 \leq j \leq m, \ 1 \leq r \leq k - m \mid A_t^{(\ell)} = x_\ell, \ 1 \leq \ell \leq k\}$$

$$= \left(\prod_{j=1}^{m} \frac{\mu((x_{i_j} + s, \infty))}{\mu((x_{i_j}, \infty))} \right) \left(\prod_{j=1}^{k-m} \frac{\mu((x_{\ell_j}, x_{\ell_j} + s))}{\mu((x_{\ell_j}, \infty))} \mu((y_j, \infty)) \right)$$

$$\times \ \exp\left\{ -\int_{y_{k-m}}^{x_k+s} \mu((u,\infty)) du \right\}$$

The process $(A_t^{(1)}, A_t^{(2)}, \ldots, A_t^{(k)})_{t \geq 0}$ *has a stationary distribution if and only if* $A_t = A_t^{(1)}$ *does, in which case it is unique and is given by*

$$\pi(dx_1, \ldots, dx_k) = \prod_{j=1}^{k} \mu((x_j, \infty)) dx_j \exp\left(-\int_{x_k}^{\infty} \mu((z, \infty)) dz \right) \qquad (2.0.1)$$

for $x_1 > x_2 > \cdots x_k > 0$.

Proof of Theorem 2.0.1. (a) Suppose that $(A_t)_{t \geq 0}$ is constructed from Λ as in Definition 1.0.1. For $s \geq x$ consider the conditional distribution of A_{s+t} given $A_s = x$. The condition $A_s = x$ is equivalent to the requirements that there is a point $(s - x, Z)$ in Λ for some $Z > x$ and that furthermore there are no points of Λ in the left-leaning wedge $\triangle(s - x, x)$. The conditional probability of the event $\{Z > z\}$ is $\mu((z, \infty))/\mu((x, \infty))$ for $z \geq x$. If $Z > x + t$, then $A_{s+t} = x + t$. Otherwise, $A_{s+t} < x + t$. The second claim of part (a) follows immediately.

Now consider $\mathbb{P}\{A_{s+t} \in dy \mid A_s = x\}$ for $y < x + t$. This case is depicted in Figure 2.1. By construction, $A_{s+t} = y$ if and only if there is a point $(s + t - y, W) \in \Lambda$ for some $W > y$, and there are no points of Λ in $\triangle(s + t - y, y)$. From above, the condition $A_s = x$ requires there to be no points of Λ in the wedge $\triangle(s - x, x)$ (the lightly shaded region in Figure 2.1). Therefore, if $A_s = x$, then $A_{s+t} = y$ if and only if $Z \leq x + t$, there are no points of Λ in the darkly shaded region of Figure 2.1), and there is a point of Λ of the form $(s + t - y, w)$ with $w > y$.

Now, the conditional probability that $Z \leq x + t$ is

$$\frac{\mu((x, x + t])}{\mu((x, \infty))}.$$

The probability that no points of Λ are in the darkly shaded region is

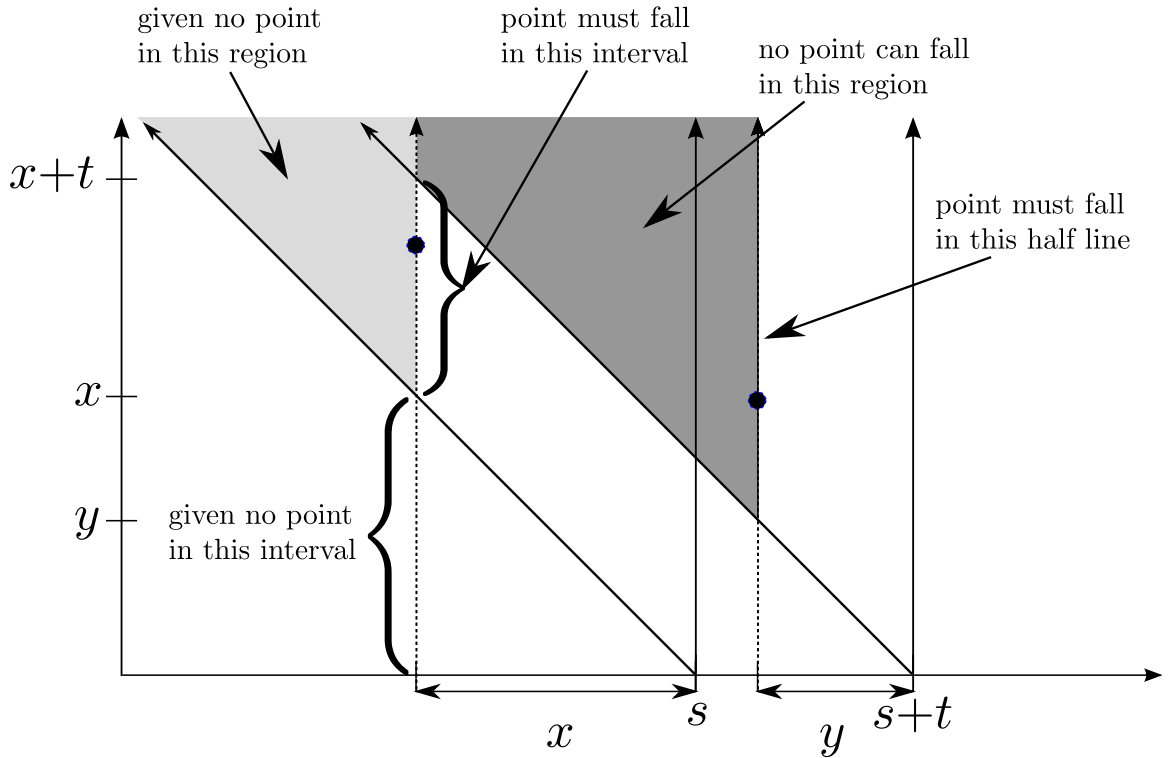$$\exp\left(-\int_{y}^{x+t} \mu((u, \infty)) \, du \right).$$

Figure 2.1: Computing $\mathbb{P}\{A_{s+t} \in dy \mid A_s = x\}$.

The probability that Λ has a point in the infinitesimal region $[s+t-y, s+t-y+dy] \times (y, \infty)$ is $\mu((y, \infty))\, dy$. Multiplying these three probabilities together gives the first claim of part (a).

(b) Both claims follow readily by differentiating the formulae in (a) at $t = 0$. It is also possible to argue directly from the representation in terms of Λ.

(c) Suppose that $(A_t)_{t \geq 0}$ is constructed from Λ as in Definition 1.0.1. Note that

$$\{t > 0 : A_t = 0\} = \mathbb{R}_{++} \setminus \bigcup_{(t,x) \in \Lambda} (t, t+x).$$

The question of when such a "Poisson cut out" random set is almost surely empty was asked in [220], and the necessary and sufficient condition presented in part (c) is simply the one found in [275] (see also [115]).

(d) For $x > 0$, the random set $\{t \geq 0 : A_t = x\}$ is a discrete regenerative set under $\mathbb{P}^x$ that is not almost surely equal to $\{0\}$. Hence this set is a renewal process with an inter-arrival

time distribution that possibly places some mass at infinity, in which case the number of arrivals is almost surely finite with a geometric distribution and the set is almost surely bounded.

Suppose that the set $\{t \geq 0 : A_t = x\}$ is $\mathbb{P}^x$-almost surely unbounded for some $x > 0$. Let $0 = T_0 < T_1 < \ldots$ be the successive visits to x. It is clear that $\mathbb{P}^x\{\exists t \in [0, T_1] : A_t = y\} > 0$ for any choice of $y > 0$ and hence, by the strong Markov property, the set $\{t \geq 0 : A_t = y\}$ is also $\mathbb{P}^x$-almost surely unbounded. Another application of the strong Markov property establishes that the set $\{t \geq 0 : A_t = y\}$ is $\mathbb{P}^y$-almost surely unbounded. Thus, the set $\{t \geq 0 : A_t = x\}$ is either unbounded $\mathbb{P}^x$-almost surely for all $x > 0$ or bounded $\mathbb{P}^x$-almost surely for all $x > 0$.

Recall that away from the set $\{t \geq 0 : A_t = 0\}$ the sample paths of A are piecewise linear with slope 1. It follows from the coarea formula (see, for example, Section 3.8 of [232]) that

$$\int_0^\infty f(A_t)\, dt = \int_0^\infty f(y)\#\{t > 0 : A_t = y\}\, dy$$

for a Borel function $f : \mathbb{R}_+ \to \mathbb{R}_+$. Hence, by Fubini's theorem

$$\int_0^\infty f(y)\mathbb{E}^x\left[\#\{t > 0 : A_t = y\}\right]\, dy = \int_0^\infty f(y)\int_0^\infty \frac{\mathbb{P}^x\{A_t \in dy\}}{dy}\, dt\, dy$$

for any $x > 0$. It follows from the continuity of the transition probability densities that

$$\mathbb{E}^x\left[\#\{t > 0 : A_t = y\}\right] = \int_0^\infty \frac{\mathbb{P}^x\{A_t \in dy\}}{dy}\, dt$$

for all $x, y > 0$ and, in particular, that the expected number of returns to $x > 0$ under $\mathbb{P}^x$ is

$$\int_0^\infty \frac{\mathbb{P}^x\{A_t \in dx\}}{dx}\, dt.$$

Using the expression from part (a) and the argument above, this quantity is infinite, and hence the number of visits is $\mathbb{P}^x$-almost surely infinite, if and only if

$$\int_1^\infty \exp\left(-\int_1^u \mu((y, \infty))\, dy\right) du = \infty.$$

If the set $\{t \geq 0 : A_t = x\}$ is $\mathbb{P}^x$-almost surely bounded for all $x > 0$, then, by an argument similar to the above, the set $\{t \geq 0 : A_t = y\}$ is $\mathbb{P}^x$-almost surely bounded for all $x, y > 0$. It follows that, for all $x > 0$, $\mathbb{P}^x$-almost surely all of the sets $\{t \geq 0 : A_t = y\}$ are finite. This implies that $\lim_{t\to\infty} A_t$ exists $\mathbb{P}^x$-almost surely and takes values in the set $\{0, \infty\}$. However, it is clear from the Poisson process construction that 0 does not occur as a limit with positive probability.

(e) Suppose there exists a probability measure π on $\mathbb{R}_+$ such that

$$\int_{\mathbb{R}_+} \mathbb{P}\{A_t \in dy \mid A_0 = x\} \, \pi(dx) = \pi(dy), \quad y \in \mathbb{R}_+.$$

Taking $t \to \infty$ in part (a) gives

$$
\begin{aligned}
\pi(dy) &= \lim_{t\to\infty} \int_{\mathbb{R}_+} \pi(dx) \frac{\mu((x, x+t])}{\mu((x,\infty))} \exp\left(-\int_y^{x+t} \mu((u,\infty)) \, du\right) \mu((y,\infty)) \, dy \\
&= \begin{cases} 0, & \text{if } \int_y^\infty \mu((u,\infty)) \, du = \infty, \\ \exp\left(-\int_y^\infty \mu((u,\infty)) \, du\right) \mu((y,\infty)) \, dy, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Therefore, a stationary probability distribution exists if and only if $\int_y^\infty \mu((u,\infty)) \, du < \infty$, and if a stationary distribution exists, then it is unique.

(f) It will be useful to begin with a concrete construction of a stationary version of the process A in terms of a Poisson point process. Suppose that $\int_x^\infty \mu((u,\infty)) \, du < \infty$ for all $x > 0$, so that a stationary distribution exists. Let $\Lambda^\leftrightarrow$ be a Poisson point process on $\mathbb{R} \times \mathbb{R}_{++}$ with intensity measure $\lambda \otimes \mu$. Define $(A_t^\leftrightarrow)_{t \in \mathbb{R}}$ by

$$A_t^\leftrightarrow := t - \inf\{s : \exists x > 0 \text{ s.t. } (s, x) \in \Lambda_\leftrightarrow \cap \triangle(t, 0)\}.$$

The condition on μ ensures that almost surely any wedge $\triangle(t, x)$ with $x > 0$ will contain only finitely many points of $\Lambda^\leftrightarrow$, and so $(A_t^\leftrightarrow)_{t \in \mathbb{R}}$ is well-defined. The process $(A_t^\leftrightarrow)_{t \in \mathbb{R}}$ is stationary and Markovian, with the same transition probabilities as $(A_t)_{t \geq 0}$.

Recall the construction of the process A^x started at x for $x > 0$ that was described preceding the statement of Theorem 2.0.1. Construct the point process Λ^x that appears there by setting $\Lambda^x := \{(t, y) \in \Lambda^\leftrightarrow : t > -x\} \cup \{(-x, Z)\}$, where Z is an independent random variable with values in the interval (x, ∞) and distribution $\mathbb{P}\{Z > z\} = \mu((z,\infty))/\mu((x,\infty))$.

By construction, $A_t^x = A_t^\leftrightarrow$ for all $t \geq T$, where T is the death time of all families alive at time $-x$ in either process:

$$T := \inf\{t > 0 : Z \leq t + x \text{ and } \Lambda^\leftrightarrow \cap \triangle(-x, t + x) = \emptyset\}.$$

Thus,

$$
\begin{aligned}
d_{TV}(\mathbb{P}^x\{A_t \in \cdot\}, \pi) &\leq \mathbb{P}\{A_t^x \neq A_t^\leftrightarrow\} \\
&\leq \mathbb{P}\{T > t\} \\
&= 1 - \mathbb{P}\{Z \leq t + x\} \, \mathbb{P}\{\Lambda^\leftrightarrow \cap \triangle(-x, t + x) = \emptyset\},
\end{aligned}
$$

and part (f) follows.

□

Proof of Theorem 2.0.2. The transition probabilities for larger collection of split times can be seen in the same way, by extending Figure 2.1 and the associated Poisson calculations to more than one point. It is also clear that the proof of Theorem 2.0.1 parts (e) and (f) apply to the case of more than one split time.

□

Chapter 3

Duality, time-reversal, and the peak and trough chains

Suppose in this chapter that $\int_x^\infty \mu((y,\infty))\,dy < \infty$ for all $x > 0$, so that, by part (e) of Theorem 2.0.1, the process A has a stationary distribution. Let $(A_t^{\leftrightarrow})_{t\in\mathbb{R}}$ be the stationary Markov process with the transition probabilities of A that was constructed from the Poisson point process $\Lambda^{\leftrightarrow}$ in the proof of part (f) of Theorem 2.0.1.

In the first section we describe an interesting duality between $A^{\leftrightarrow}$ and its time reversal, which exchanges the roles of the jump heights and the inter-jump intervals. In the following section we find explicit transition probabilities for the discrete-time Markov chains that give the successive trough depths and peak heights of the sample path of $A^{\leftrightarrow}$.

3.1 Duality and time-reversal

Define the *dual process* $(\widehat{A}_t^{\leftrightarrow})_{t\in\mathbb{R}}$ by $\widehat{A}_t^{\leftrightarrow} := \inf\{s > 0 : \triangle(t,s) \cap \Lambda^{\leftrightarrow} = \emptyset\}$. See Figure 3.1. Thus, $\widehat{A}_t^{\leftrightarrow}$ is the amount of time that must elapse after t until all families alive at time t have died out, or, equivalently, until the MRCA for the population lived at some time after t. The càdlàg $\mathbb{R}_+$-valued process $(\widehat{A}_t^{\leftrightarrow})_{t\in\mathbb{R}}$ has saw-tooth sample-paths that drift down with slope -1 between upward jumps.

Proposition 3.1.1. *The dual process $(\widehat{A}_t^{\leftrightarrow})_{t\in\mathbb{R}}$ has the same distribution as the time-reversed process $(\bar{A}_t^{\leftrightarrow})_{t\in\mathbb{R}}$, where $\bar{A}_t := \lim_{u\downarrow t} A_{-u}^{\leftrightarrow}$.*

Proof. Define the bijections ϕ and σ of $\mathbb{R}\times\mathbb{R}_{++}$ by $\phi(t,x) := (t+x, x)$ and $\sigma(t,x) = (-t, x)$.
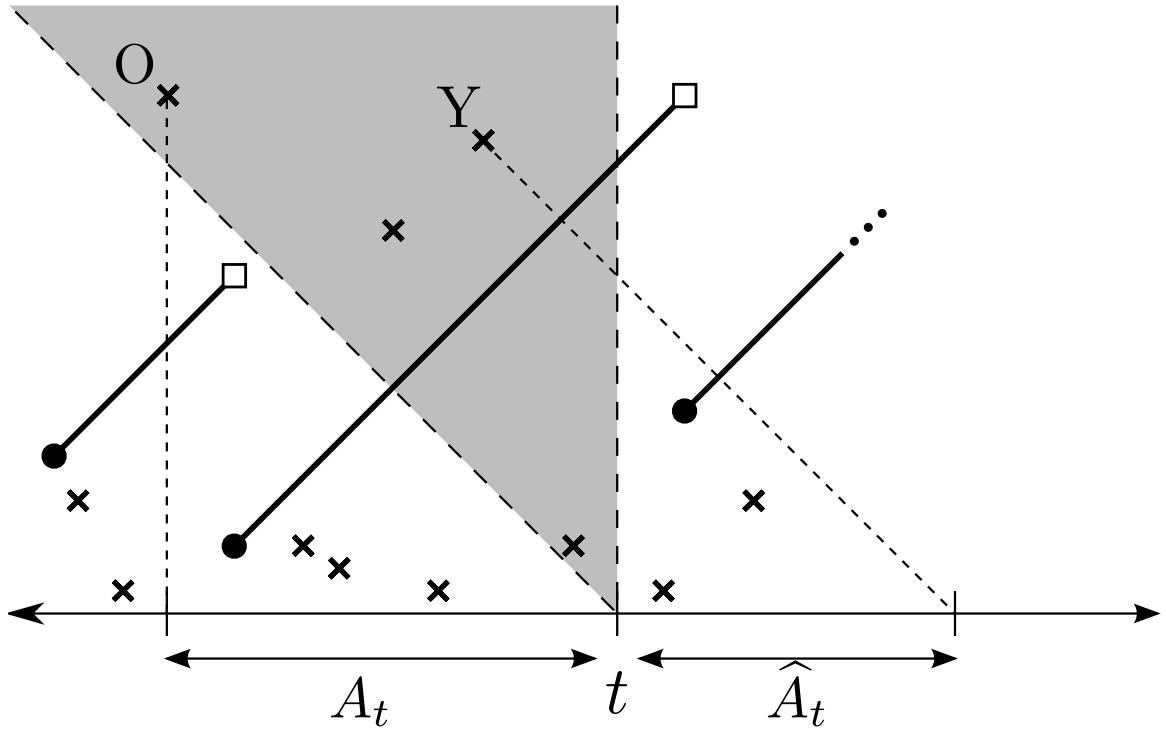
Figure 3.1: The process $A_t^{\leftrightarrow}$ and the dual process $\widehat{A}_t^{\leftrightarrow}$. The points in the shaded area represent the families alive at time t, and the solid line is the sample path of $A_t^{\leftrightarrow}$. The point marked "O" is the oldest living family at time t; the point marked "Y" is the family extant at time t that will live the longest into the future.

In the usual manner, we may also think of φ and σ as mapping subsets of $\mathbb{R} \times \mathbb{R}_{++}$ to other subsets of $\mathbb{R} \times \mathbb{R}_{++}$. Note that φ maps left-leaning wedges to right-leaning wedges, and σ maps right-leaning wedges to left-leaning wedges. Thus, the composition $\sigma \circ \varphi$ maps left-leaning wedges to left-leaning wedges. More precisely,

$$
\begin{aligned}
\sigma \circ \varphi(\triangle(t,x)) &= \{\sigma \circ \varphi(s,y) : (s,y) \in \triangle(t,x)\} \\
&= \{(-(s+y), y) : s < t \ \& \ s+y > t+x\} \\
&= \{(u,v) : u < -(t+x) \ \& \ u+v > -t = -(t+x)+x\} \\
&= \triangle(\sigma \circ \varphi(t,x)),
\end{aligned}
\tag{3.1.1}
$$

where we stress that $\triangle(\sigma \circ \varphi(t,x))$ is the left-leaning wedge with its apex at the point $\sigma \circ \varphi(t,x) \in \mathbb{R} \times \mathbb{R}_{++}$. Define a map Γ that takes a subset of $\mathbb{R} \times \mathbb{R}_{++}$ and returns another such subset by

$$
\Gamma(S) := \{(t,x) \in S : \triangle(t,x) \cap S = \emptyset\}.
$$

The points of $\Gamma(\Lambda^{\leftrightarrow})$ correspond precisely to those families that at some time will be the oldest surviving family in the population, and determine the jumps of both the MRCA process and the dual process. As can be seen with the help of Figure 3.2, the linear segments of the paths of the dual process $(\widehat{A}_t^{\leftrightarrow})_{t\in\mathbb{R}}$ each begin at a point in $\Gamma(\Lambda^{\leftrightarrow})$ and descend with slope -1, whereas the linear segments of the paths of the MRCA process $(A_t^{\leftrightarrow})_{t\in\mathbb{R}}$ ascend with slope $+1$ to points in $\phi \circ \Gamma(\Lambda^{\leftrightarrow})$.

This implies that the path of the time-reversed process $(\bar{A}_t)_{t\in\mathbb{R}}$ begin at points in $\sigma \circ \phi \circ \Gamma(\Lambda^{\leftrightarrow})$ and descend with slope -1, and so the points of $\sigma \circ \phi \circ \Gamma(\Lambda^{\leftrightarrow})$ determine the path of $(\bar{A}_t)_{t\in\mathbb{R}}$ in the same way that the points of $\Gamma(\Lambda^{\leftrightarrow})$ determine the path of $(\widehat{A}_t)_{t\in\mathbb{R}}$. Therefore, all we need to do is show that $\sigma \circ \phi \circ \Gamma(\Lambda^{\leftrightarrow}) \stackrel{d}{=} \Gamma(\Lambda^{\leftrightarrow})$.

To see this, first note that for an arbitrary point set $S \subset \mathbb{R} \times \mathbb{R}_{++}$,

$$
\begin{aligned}
\sigma \circ \phi \circ \Gamma(S) &= \{\sigma \circ \phi(t,x) : (t,x) \in S \text{ and } \triangle(t,x) \cap S = \emptyset\} \\
&= \{(t,x) : \phi^{-1} \circ \sigma^{-1}(t,x) \in S \text{ and } \triangle(\phi^{-1}\circ\sigma^{-1}(t,x)) \cap S = \emptyset\} \\
&= \{(t,x) : \phi^{-1}\circ\sigma^{-1}(t,x) \in S \text{ and } \phi^{-1}\circ\sigma^{-1}(\triangle(t,x)) \cap S = \emptyset\} \\
&= \{(t,x) : (t,x) \in \sigma\circ\phi(S) \text{ and } \triangle(t,x) \cap \sigma\circ\phi(S) = \emptyset\} \\
&= \Gamma \circ \sigma \circ \phi(S)
\end{aligned}
$$

where the third equality follows from identity (3.1.1). Thus, $\sigma \circ \phi \circ \Gamma(\Lambda^{\leftrightarrow}) = \Gamma \circ \sigma \circ \phi(\Lambda^{\leftrightarrow})$. However, both maps ϕ and σ preserve the measure $\lambda \otimes \mu$, and hence $\sigma \circ \phi(\Lambda^{\leftrightarrow}) \stackrel{d}{=} \Lambda^{\leftrightarrow}$, so $\sigma \circ \phi \circ \Gamma(\Lambda^{\leftrightarrow}) \stackrel{d}{=} \Gamma(\Lambda^{\leftrightarrow})$.

$\square$

Remark 3.1.1. There is an interesting connection between the jump sizes and the inter-jump intervals, stemming from the observation that $A^{\leftrightarrow}$ and $\widehat{A}^{\leftrightarrow}$ have the same sequences of "trough" and "peak" heights, while the roles of the jump sizes and inter-jump intervals for the two are exchanged. The following situation is depicted in Figure 3.2. To explain the connection, suppose that $T \in \{t \in \mathbb{R} : A_{t-}^{\leftrightarrow} \neq A_t^{\leftrightarrow}\}$ is a jump time for the process $A^{\leftrightarrow}$. Let $T^{<} := \sup\{t < T : A_{t-}^{\leftrightarrow} \neq A_t^{\leftrightarrow}\}$ and $T^{>} := \inf\{t > T : A_{t-}^{\leftrightarrow} \neq A_t^{\leftrightarrow}\}$ be the jump times on either side of T. Put $L := A_{T-}^{\leftrightarrow}$ and $R := A_T^{\leftrightarrow}$, and define $L^{<}$, $R^{<}$, $L^{<}$, and $R^{>}$ as the analogous values and left limits of $A^{\leftrightarrow}$ at the times $T^{<}$ and $T^{>}$. Write $\Delta := L - R$ for the size of the jump at time T and $H := T - T^{<}$ for the length of the time interval since the previous jump. Observe that $T - L$ is a jump time for the dual process $\widehat{A}^{\leftrightarrow}$, with

$\widehat{A}^{\leftrightarrow}_{(T-L)-} = R^<$ and $\widehat{A}^{\leftrightarrow}_{(T-L)} = L$. Moreover,

$$H = L - R^< = \widehat{A}^{\leftrightarrow}_{(T-L)} - \widehat{A}^{\leftrightarrow}_{(T-L)-}$$

and

$$\Delta = (T^> - L^>) - (T - L) = \inf\{t > T - L : \widehat{A}^{\leftrightarrow}_{t-} \neq \widehat{A}^{\leftrightarrow}_t\} - (T - L).$$

Note that the map $T \mapsto T - L$ sets up a monotone bijection between the jump times of the process $A^{\leftrightarrow}$ and those of the process $\widehat{A}^{\leftrightarrow}$. It thus follows from Proposition 3.1.1 that the point processes

$$\left\{(T, T - T^<, A^{\leftrightarrow}_{T-} - A^{\leftrightarrow}_T) : A^{\leftrightarrow}_{T-} \neq A^{\leftrightarrow}_T\right\}$$

and

$$\left\{(T, A^{\leftrightarrow}_{T-} - A^{\leftrightarrow}_T, T - T^<) : A^{\leftrightarrow}_{T-} \neq A^{\leftrightarrow}_T\right\}$$

have the same distribution.

3.2 The peak and trough jump chains

Now we turn to the properties of the sequences of peak heights and trough depths in a saw-tooth sample path of $A^{\leftrightarrow}$.

For $t \in \mathbb{R}$, denote by $J_t := \inf\{u > 0 : A^{\leftrightarrow}_u \neq A^{\leftrightarrow}_{u-}\}$ the next jump time of $A^{\leftrightarrow}$ after time t. Define an increasing sequence of random times $0 < T_0 < T_1 < \cdots$ by $T_0 := J_0$ and $T_{n+1} := J_{T_n}$ for $n \geq 0$. Put $L_n := A^{\leftrightarrow}_{T_n-}$ and $R_n := A^{\leftrightarrow}_{T_n}$. Thus, the sequences $(L_n)_{n=0}^\infty$ and $(R_n)_{n=0}^\infty$ record, respectively, the "peaks" and the "troughs" of the path of $A^{\leftrightarrow}$ that occur between the times 0 and $\sup_n T_n$.

The next result can be proved along the same lines as part (a) of Theorem 2.0.1, and we leave the proof to the reader.

Proposition 3.2.1. *The sequence* $(L_0, R_0, L_1, R_1, \ldots)$ *is Markovian with the following transition probabilities:*

$$\mathbb{P}\{R_n \in dy \mid L_n = x\} = \mu((y, \infty)) \exp\left(-\int_y^x \mu((u, \infty))\, du\right) dy, \quad 0 < y \leq x,$$

and

$$\mathbb{P}\{L_{n+1} \in dz \mid R_n = y\} = \frac{m(z)}{\mu((y, \infty))}\, dz, \quad z > y.$$

In particular, the sequence of pairs $((L_n, R_n))_{n=0}^\infty$ *is a time-homogeneous Markov chain.*

Now we may compute the transition probabilities of the peak and trough chains. By Proposition 3.2.1,

$$
\begin{aligned}
\mathbb{P}\{L_{n+1} &\in dz \mid L_n = x\}/dz \\
&= \int_0^x \exp\left(-\int_y^x \mu((u,\infty))\,du\right) \mu((y,\infty)) \frac{m(z)\mathbf{1}_{y\leq z}}{\mu((y,\infty))}\,dy \\
&= m(z) \int_0^{x\wedge z} \exp\left(-\int_y^x \mu((u,\infty))\,du\right)\,dy,
\end{aligned}
\tag{3.2.1}
$$

and

$$
\begin{aligned}
\mathbb{P}\{R_{n+1} &\in dz \mid R_n = x\}/dz \\
&= \int_x^\infty \frac{m(y)}{\mu((x,\infty))} \mu((z,\infty)) \exp\left(-\int_z^y \mu((u,\infty))\,du\right) \mathbf{1}_{y>z}\,dy \\
&= \mu((z,\infty)) \int_{x\vee z}^\infty \frac{m(y)}{\mu((x,\infty))} \exp\left(-\int_z^y \mu((u,\infty))\,du\right)\,dy.
\end{aligned}
\tag{3.2.2}
$$

It follows from (3.2.1) that the peak chain $(L_n)_{n=0}^\infty$ is λ-*irreducible*, where λ is Lebesgue measure on $\mathbb{R}_{++}$. That is, if A is a Borel subset of $\mathbb{R}_{++}$ with $\lambda(A) > 0$, then, for any $x \in \mathbb{R}_{++}$, there is positive probability that the peak chain begun at x will hit A at some positive time — see Ch. 4 of [228] for more about this notion of irreducibility. It follows that the peak chain is either *recurrent*, in the sense that

$$
\sum_{n=0}^\infty \mathbb{P}\{L_n \in A \mid L_0 = x\} = \infty
$$

for all $x \in \mathbb{R}_{++}$ and all Borel subsets of $A \subseteq \mathbb{R}_{++}$ with $\lambda(A) > 0$, or it is *transient*, in the sense that there is a countable collection of Borel sets $(A_j)_{j=1}^\infty$ and finite constants $(M_j)_{j=1}^\infty$ such that $\bigcup_{j=1}^\infty A_j = \mathbb{R}_{++}$ and

$$
\sup_{x\in\mathbb{R}_{++}} \sum_{n=0}^\infty \mathbb{P}\{L_n \in A_j \mid L_0 = x\} \leq M_j
$$

— see Theorem 8.0.1 of [228].

The peak chain is *strong Feller*; that is, the function

$$
x \mapsto \mathbb{E}[f(L_{n+1}) \mid L_n = x]
$$

is continuous for any bounded Borel function f. Also, because the support of λ is all of $\mathbb{R}_{++}$, if the peak chain is recurrent, then each point x of $\mathbb{R}_{++}$ is *topologically recurrent* in the sense that

$$
\sum_{n=0}^\infty \mathbb{P}\{L_n \in U \mid L_0 = x\} = \infty
$$

for every open neighborhood U of x. Hence, by Theorem 9.3.6 of [228], if the peak chain is recurrent, then it is *Harris recurrent*, which means that given any Borel set A with $\lambda(A) > 0$, the chain visits A infinitely often almost surely starting from any x. Moreover, the chain is recurrent (equivalently, Harris recurrent) if and only if it is *non-evanescent*; that is, started from any x there is zero probability that the chain converges to 0 or ∞ — see Theorem 9.2.2 of [228].

If the peak chain is recurrent (equivalently, Harris recurrent or non-evanescent), then it has an invariant measure that is unique up to constant multiples — see Theorem 10.4.4 of [228]. If the invariant measure has finite mass, so that it can be normalized to be a probability measure, then the chain is said to be *positive*, otherwise the chain is said to be *null*.

Conversely, if the peak chain has an invariant probability measure, then it is recurrent (equivalently, Harris recurrent or non-evanescent) — see Proposition 10.1.1 of [228].

All of the remarks we have just made for the peak chain apply equally to the trough chain $(R_n)_{n=0}^\infty$. Recall that we are in the situation when A has a stationary version, so the transience or recurrence of L and R depends on their behavior near zero.

Proposition 3.2.2. *Consider the two Markov chains* $(L_n)_{n=0}^\infty$ *and* $(R_n)_{n=0}^\infty$.

(a) *Both chains are transient if and only if*

$$\int_0^1 \exp\left(\int_x^1 \mu((y,\infty))\,dy\right) dx < \infty.$$

(b) *Both chains are positive recurrent if and only if*

$$\int_0^1 m(x) \exp\left(-\int_x^1 \mu((y,\infty))\,dy\right) dx < \infty.$$

(c) *Both chains are null recurrent if and only if both*

$$\int_0^1 \exp\left(\int_x^1 \mu((y,\infty))\,dy\right) dx = \infty$$

and

$$\int_0^1 m(x) \exp\left(-\int_x^1 \mu((y,\infty))\,dy\right) dx = \infty.$$

Proof. Consider the set $\mathcal{Z} := \{t \in \mathbb{R} : A_t^{\leftrightarrow} = 0\}$. It follows from part (c) of Theorem 2.0.1 that $\mathbb{P}\{\mathcal{Z} \neq \emptyset\} > 0$ if and only if

$$\int_0^1 \exp\left(\int_x^1 \mu(y, \infty)\,dy\right) dx < \infty. \tag{3.2.3}$$

By the stationarity of $(A_t^{\leftrightarrow})_{t \in \mathbb{R}}$ and the nature of its sample paths, it is clear that for $x > 0$ the set $\{t \in \mathbb{R} : A_t^{\leftrightarrow} = x\}$ is unbounded above and below almost surely (this also follows from part (d) of Theorem 2.0.1). It follows from a simple renewal argument that if (3.2.3) holds, then $\mathcal{Z}$ is unbounded above and below almost surely.

Because the paths of $(A_t^{\leftrightarrow})_{t \in \mathbb{R}}$ increase with slope 1 in the intervals $[T_n, T_{n+1})$, it follows that if (3.2.3) condition holds, then $\lim_{n \to \infty} T_n = \inf\{t > 0 : A_t^{\leftrightarrow} = 0\} < \infty$ almost surely and $\lim_{n \to \infty} L_n = \lim_{n \to \infty} R_n = 0$ almost surely. In this case, both chains are evanescent, and hence transient.

On the other hand, if (3.2.3) does not hold, then $\lim_{n \to \infty} T_n = \infty$. Moreover, the set $\{t \in \mathbb{R} : A_t^{\leftrightarrow} = x\}$ is almost surely unbounded above and below for any $x > 0$, as we observed above. If we split the path of $(A_t^{\leftrightarrow})_{t \in \mathbb{R}}$ into excursions away from x, then each excursion interval will contain only finitely many jumps almost surely and, because the excursions are independent and identically distributed, it cannot be the case that L_n or R_n converges to 0 or ∞ with positive probability. Thus, both chains are non-evanescent and hence recurrent.

It is clear from (3.2.1) that the kernel giving the transition densities of the peak chain $(L_n)_{n=0}^\infty$ is self-adjoint with respect to the measure having density

$$p(x) = m(x) \exp\left(-\int_x^\infty \mu((u, \infty))\,du\right),$$

with respect to Lebesgue measure, and so this measure is invariant for the peak chain. Clearly, $\int_0^\infty p(x)\,dx < \infty$ if and only if the condition in part (b) holds, in which case the peak chain is positive recurrent. Otherwise, the peak chain is either null recurrent or transient, and so part (a) shows that the peak chain is null recurrent if the two conditions in part (c) hold.

Similarly, it is clear from (3.2.2) that the kernel giving the transition densities of the trough chain $(R_n)_{n=0}^\infty$ is self-adjoint with respect to the measure having density

$$q(x) = \mu((x, \infty))^2 \exp\left(-\int_x^\infty \mu((u, \infty))\,du\right)$$

with respect to Lebesgue measure, and so this measure is invariant for the trough chain. An integration by parts shows that $\int_0^\infty q(x)\,dx < \infty$ if and only if the condition in part (b) holds, and so the trough chain is positive if and only if the peak chain is positive. Alternatively, we can simply observe from Proposition 3.2.1 that integrating the conditional probability kernel of R_n given L_n against an invariant probability measure for the peak chain gives an invariant measure for the trough chain, and integrating the conditional probability kernel of L_{n+1} given R_n against an invariant probability measure for the trough chain chain gives an invariant measure for the peak chain, so that one chain is positive recurrent if and only if the other is. $\qquad\square$

Remark 3.2.1. If $m(x) = \alpha x^{-2}$ for $x \in (0,1]$, then both the peak and trough chains are

1. transient $\Leftrightarrow 0 < \alpha < 1$,

2. null recurrent $\Leftrightarrow \alpha = 1$,

3. positive recurrent $\Leftrightarrow \alpha > 1$.

Remark 3.2.2. It follows from parts (b) and (e) of Theorem 2.0.1 that the stationary point process $\{t \in \mathbb{R} : A_{t-}^{\leftrightarrow} \neq A_t^{\leftrightarrow}\}$ has intensity

$$\rho := \int m(x) \exp\left(-\int_x^\infty \mu((u,\infty))\,du\right)\,dx,$$

and so the peak and trough chains are positive recurrent if and only if ρ is finite. Suppose that ρ is finite and consider the point process

$$\Xi := \{(t, A_{t-}^{\leftrightarrow}, A_t^{\leftrightarrow}) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ : A_{t-}^{\leftrightarrow} \neq A_t^{\leftrightarrow}\}.$$

The companion *Palm* point process Υ has its distribution defined by

$$\mathbb{P}\{\Upsilon \in \cdot\} = \rho^{-1}\mathbb{E}\left[\sum_{\{n:0 \leq T_n \leq 1\}} \mathbf{1}\{\theta_{T_n}\Xi \in \cdot\}\right],$$

where $\theta_s A = \{(t-s,\ell,r) : (t,\ell,r) \in A\}$ for $A \subset \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$. Enumerate the points of Υ as $((\tilde{T}_n, \tilde{L}_n, \tilde{R}_n))_{n \in \mathbb{Z}}$, where $\ldots < \tilde{T}_{-1} < \tilde{T}_0 = 0 < \tilde{T}_1 < \ldots$. A fundamental result of Palm theory for stationary point processes says that the random sequence $((\tilde{T}_n - \tilde{T}_{n-1}, \tilde{L}_n, \tilde{R}_n))_{n \in \mathbb{Z}}$ is stationary and that the distribution of the point process Ξ may be reconstructed from the distribution of this sequence – see, for example, Theorem 12.3.II of [58] or [174]. It is clear that the stationary random sequences $(\tilde{L}_n)_{n=0}^\infty$ and $(\tilde{R}_n)_{n=0}^\infty$ have the same distribution as the peak and trough chains started in their respective stationary distributions.
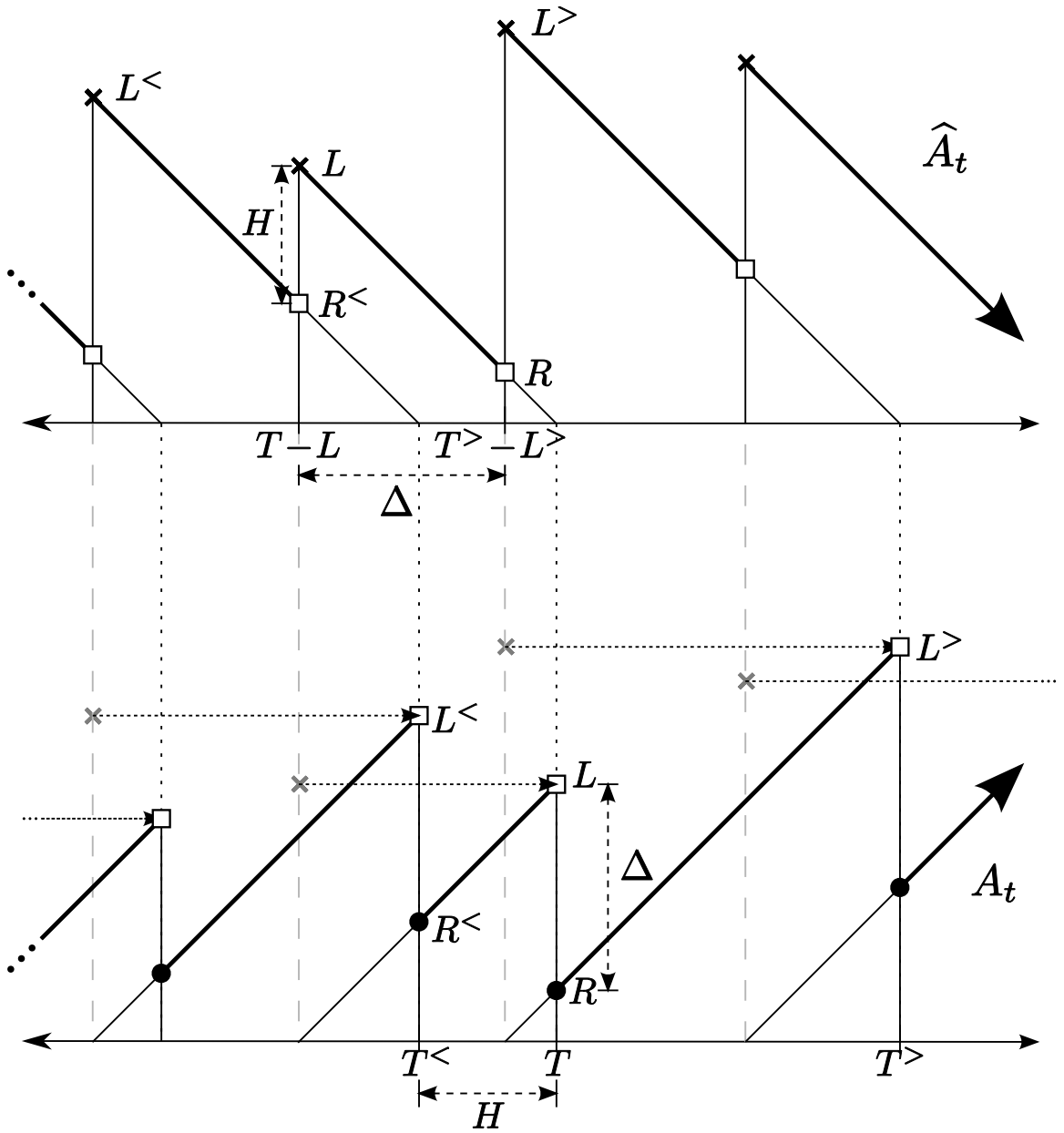
Figure 3.2: $A_t^{\leftrightarrow}$, $\widehat{A}_t^{\leftrightarrow}$, and the map ϕ. This shows the (coupled) sample paths of both $A_t^{\leftrightarrow}$ and $\widehat{A}_t^{\leftrightarrow}$ for the same set of family sizes (the "x"s). $\widehat{A}_t^{\leftrightarrow}$ begins at points in $\Gamma(\Lambda^{\leftrightarrow})$ and descends; $A_t^{\leftrightarrow}$ ascends to points in $\phi(\Gamma(\Lambda^{\leftrightarrow}))$. The mapping ϕ is shown by the horizontal dotted lines in the lower picture.

Chapter 4

Genealogical properties of the $(1+\beta)$-stable CSBP

In this chapter we specialize to the motivating example of a critical $(1+\beta)$-stable continuous state branching process conditioned to live forever. Recall that the unconditioned continuous state branching process has Laplace transform (1.0.1) and the conditioned process has Laplace transform (1.0.2). For $\beta = 1$, the unconditioned process has generator $x\frac{\partial^2}{\partial x^2}$ with this choice of time scale.

In the first section we describe properties of the MRCA process. In the second section we extend these results to larger collections of split times, and along the way describe some interesting "fragmentation-coagulation-drift" processes.

4.1 The $(1+\beta)$-stable MRCA process

Applying Theorem 2.0.1 in a straightforward manner gives parts (a) to (d) of the following result. Part (e) follows from an observation that a space-time rescaling of this MRCA age process is a time-homogeneous Markov process that arises from another Poisson process by the general MRCA age construction of Definition 1.0.1.

Corollary 4.1.1. *Suppose that A is the MRCA age process associated with the critical $(1+\beta)$-stable continuous state branching process.*

(a) *The transition probabilities of the process A have an absolutely continuous part*

$$\mathbb{P}^x \left\{ A_t \in dy \right\} = \frac{(1+\beta)ty^{1/\beta}}{\beta(x+t)^{2+1/\beta}} \, dy \,, \qquad 0 < y < x+t,$$

and a single atom

$$\mathbb{P}^x \left\{ A_t = x+t \right\} = \frac{x}{x+t}.$$

(b) *The total rate at which the process A jumps from the state $x > 0$ is $1/x$, and when it jumps from state $x > 0$, the distribution of the state to which it jumps is absolutely continuous with density*

$$(1+1/\beta)\frac{y^{1/\beta}}{x^{1+1/\beta}}, \qquad 0 < y < x.$$

(c) *The probability $\mathbb{P}^0\{\exists t > 0 : A_t = 0\}$ that the process A returns to the state zero is 0.*

(d) *For each $x \geq 0$, $\lim_{t\to\infty} A_t = \infty$, $\mathbb{P}^x$-almost surely.*

(e) *The process*

$$\left(e^{-t} A_{e^t} \right)_{t \in \mathbb{R}}$$

indexed by the whole real line is a time-homogeneous Markov process under $\mathbb{P}^x$ for any $x \geq 0$, and it is stationary when $x = 0$. Moreover, A_t/t converges in distribution to the $\mathrm{Beta}(1+1/\beta, 1)$ distribution as $t \to \infty$ under $\mathbb{P}^x$ for any $x \geq 0$, and A_t/t has the $\mathrm{Beta}(1+1/\beta, 1)$ distribution for all $t > 0$ when $x = 0$.

As mentioned in the introduction, the set of points $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_{++}$, where t is the time that a family splits from the immortal lineage and x is its total lifetime, is a Poisson point process with intensity measure $\lambda \otimes \mu$ for some σ-finite measure μ. This lemma describes the measure μ in this case:

Lemma 4.1.1. *The lifetime measure μ associated with the critical $(1+\beta)$-branching process conditioned on non-extinction is given by*

$$\mu((x, \infty)) = \frac{1+\beta}{\beta x}, \quad x > 0.$$

Proof. As we remarked in the introduction

$$\mathbb{P}\{X_t > 0 \mid X_0 = x\} = 1 - \exp\left(-\frac{x}{t^{1/\beta}}\right). \tag{4.1.1}$$

First consider the case of $\beta = 1$. Recall from the introduction that if Π is a Poisson point process on $\mathbb{R}_+ \times \mathcal{E}^0$ with intensity $\lambda \otimes \nu$, then

$$\left(\sum_{(s,u) \in \Pi} u_{(t-s) \vee 0} \right)_{t \geq 0} \tag{4.1.2}$$

has the same distribution as the conditioned process $(Y_t)_{t \geq 0}$ with $Y_0 = 0$, and recall that μ is the push-forward of ν by the total lifetime function γ. Also,

$$\left(\sum_{(s,u) \in \Pi : s \leq y/2} u_t \right)_{t \geq 0} \tag{4.1.3}$$

has the same distribution as the unconditioned process $(X_t)_{t \geq 0}$ with $X_0 = y$ — see [93]. The factor of 2 differs from [93] and arises from our choice of time scale. Therefore,

$$\begin{aligned} \mathbb{P}\{X_t > 0 \mid X_0 = y\} &= \mathbb{P}\{\exists (s,u) \in \Pi \ : \ s \leq y/2 \text{ and } \gamma(u) > t\} \\ &= 1 - \exp\left(-y\,\mu((t,\infty))/2\right), \end{aligned}$$

and comparing with (4.1.1) gives $\mu((t,\infty)) = 2/t$.

Now take $\beta \in (0,1)$. It is shown in [88] that the mass thrown off the immortal lineage is determined by the jumps of a stable subordinator: if M_s is the amount of mass thrown off during the time interval $[0,s]$, then

$$\begin{aligned} \mathbb{E}\left[e^{-\theta M_s} \right] &= \exp\left(-s \frac{1+\beta}{\beta} \theta^\beta \right) \\ &= \exp\left(-s \int_0^\infty \left(1 - e^{-\theta x}\right) \nu(dx) \right), \end{aligned}$$

where $\nu(dx) = \frac{1+\beta}{\Gamma(1-\beta)} x^{-(1+\beta)} dx$ is the Lévy measure of the subordinator.

Since the jump size of the subordinator corresponds to the initial size of the new family, the lifetime measure μ is given by

$$\mu((t,\infty)) = \int_0^\infty \mathbb{P}\{X_t > 0 \mid X_0 = x\}\,\nu(dx).$$

and so, from the above and an integration by parts,

$$\begin{aligned} \mu((t,\infty)) &= \int_0^\infty \left(1 - e^{-x/t^{1/\beta}}\right) \nu(dx) \\ &= \frac{1+\beta}{\beta t}. \end{aligned}$$

$\square$

Proof of Corollary 4.1.1

Parts (a) to (d) follow immediately from Theorem 2.0.1. Part (e) will also follow from parts (e) and (f) of Theorem 2.0.1 after the following time and space change.

Define a new time parameter u by $t = e^u$. If the MRCA at time t lived at time $t - x$ on the original scale, then on the new time scale she lived at time $u - y$, where $t - x = e^{u-y}$. Solving for y, the MRCA age process in the new time scale is the process $(B_t)_{t \geq 0}$ given by $B_u = -\log(1 - e^{-u} A_{e^u})$. The process $(B_t)_{t \geq 0}$ is obtained by applying the construction (1.0.1) to the point process given by

$$\left\{ \left(\log s, \log\left(1 + \frac{x}{s}\right) \right), (s, x) \in \Lambda \right\},$$

which is a Poisson point process on $\mathbb{R} \times \mathbb{R}_+$ with intensity measure $\lambda \otimes \rho$, where

$$\rho((y, \infty)) = \frac{1 + \beta}{\beta(e^y - 1)}, \quad y \in \mathbb{R}_{++}.$$

Note that, in general, a time and space change of the Poisson process Λ gives a new Poisson point process, but the resulting intensity measure will not typically be of the form $\lambda \otimes \kappa$ for some measure κ: it is a special feature of μ and the transformation that the product measure structure is maintained in this case.

It is straightforward to check parts (e) and (f) of Theorem 2.0.1 that $(B_t)_{t \geq 0}$ has the stationary distribution

$$\pi(dx) = \frac{1 + \beta}{\beta} e^{-x} (1 - e^{-x})^{\frac{1}{\beta}} \, dx,$$

and that the distribution of B_t converges to π in total variation as $t \to \infty$. Part (e) of the corollary then follows from the observation that $\frac{A_t}{t} = 1 - e^{-B_{\log(t)}}$ and an elementary change of variables. $\square$

4.2 The sequence of split times

The following is a straightforward application of Theorem 2.0.2, applied to both A and its space-time transformation B, as in the proof of Corollary 4.1.1.

Corollary 4.2.1. *If $A_t^{(1)}, \ldots, A_t^{(k)}$ are the oldest k splits in the genealogy of a critical $1 + \beta$-stable branching process, then the transition probabilities have the following form. Let $0 < x_1 < \cdots < x_k < t$, let $i_1 < \cdots < i_m$ be distinct integers in $\{1, \ldots, k\}$, let $\ell_1 < \ldots < \ell_{k-m}$*

be such that $\{1, \ldots, k\} = \{i_1, \ldots, i_m\} \cup \{\ell_1, \ldots, \ell_{k-m}\}$, *and let* $x_{i_m} + s < y_1 < y_2 < \cdots < y_{k-m} < t + s$. *Then*

$$\mathbb{P}\{A_{t+s}^{(j)} = x_{i_j} + s, A_{t+s}^{(m+r)} = y_r, \ 1 \le j \le m, \ 1 \le r \le k - m \mid A_t^{(\ell)} = x_\ell, \ 1 \le \ell \le k\}$$

$$= ((1 + 1/\beta)s)^{(k-m)} \left(\frac{y_{k-m}}{x_k + s}\right)^{1+1/\beta} \prod_{j=1}^{m} \frac{x_{i_j}}{x_{i_j} + s} \prod_{j=1}^{k-m} \frac{1}{y_j(x_{\ell_j} + s)} \qquad (4.2.1)$$

and if $A_0^{(1)} = 0$, *then for all* $t \ge 0$,

$$\mathbb{P}\left\{\frac{A_t^{(1)}}{t} \in da_1, \ldots, \frac{A_t^{(k)}}{t} \in da_k\right\} = (1 + \frac{1}{\beta})^k a_k^{1+1/\beta} \prod_{i=1}^{k} \frac{da_i}{a_i} \qquad (4.2.2)$$

for $1 > a_1 > \cdots > a_k > 0$.

We can further interpret the values of $A_t^{(1)}/t, A_t^{(2)}/t, \ldots$ as follows. We know that if $A_0^{(1)} = 0$, then $A_t^{(1)}/t$ has the Beta$(1 + 1/\beta, 1)$ distribution. By the branching property, we therefore know that $A_t^{(2)}/A_t^{(1)}$ also has the Beta$(1 + 1/\beta, 1)$ distribution, independently of $A_t^{(1)}$, and so forth recursively. Therefore, if we let $U_1, U_2, \ldots$ be a sequence of i.i.d. Beta$(1 + 1/\beta, 1)$ random variables, then we have the following *stick-breaking* construction:

$$(A_t^{(1)}/t, A_t^{(2)}/t, A_t^{(3)}/t, \ldots) \overset{d}{=} (1 - U_1, (1 - U_1)(1 - U_2), (1 - U_1)(1 - U_2)(1 - U_3), \ldots).$$

By this representation, for any time $t > 0$, if we define $W_0 = 1 - A_t^{(1)}$ and $W_k = (A_t^{(k+1)} - A_t^{(k)})/t$ for $k \ge 1$, then $(W_0, W_1, W_2, \ldots)$ has the GEM$(1 + 1/\beta)$ distribution. In other words, the birth times of the families extant at any time t divide the interval $[0, t]$ into a partition whose distribution is a size-biased ordering of a Poisson-Dirichlet$(1 + 1/\beta)$ sequence [67, 253].

We now more fully describe how this partition evolves in time. Note that by the above, $e^{-t} A_{e^t}^{(k)}$ is a time-homogeneous Markov process. For each k, define functions $w = w^k$ and $u = u^k$ that each map $\{x = (x_1, \ldots, x_k) : 1 > x_1 > x_2 > \cdots > x_k > 0\}$ onto $(0, 1)^k$ by

$$w_0(x) = 1 - x_1,$$
$$w_i(x) = x_i - x_{i+1} \quad \text{for } 1 \le i < k,$$
$$w_k(s) = x_k,$$
$$u_1(x) = x_1,$$
$$\text{and} \quad u_i(x) = x_i/x_{i-1} \quad \text{for } i \ge 2.$$
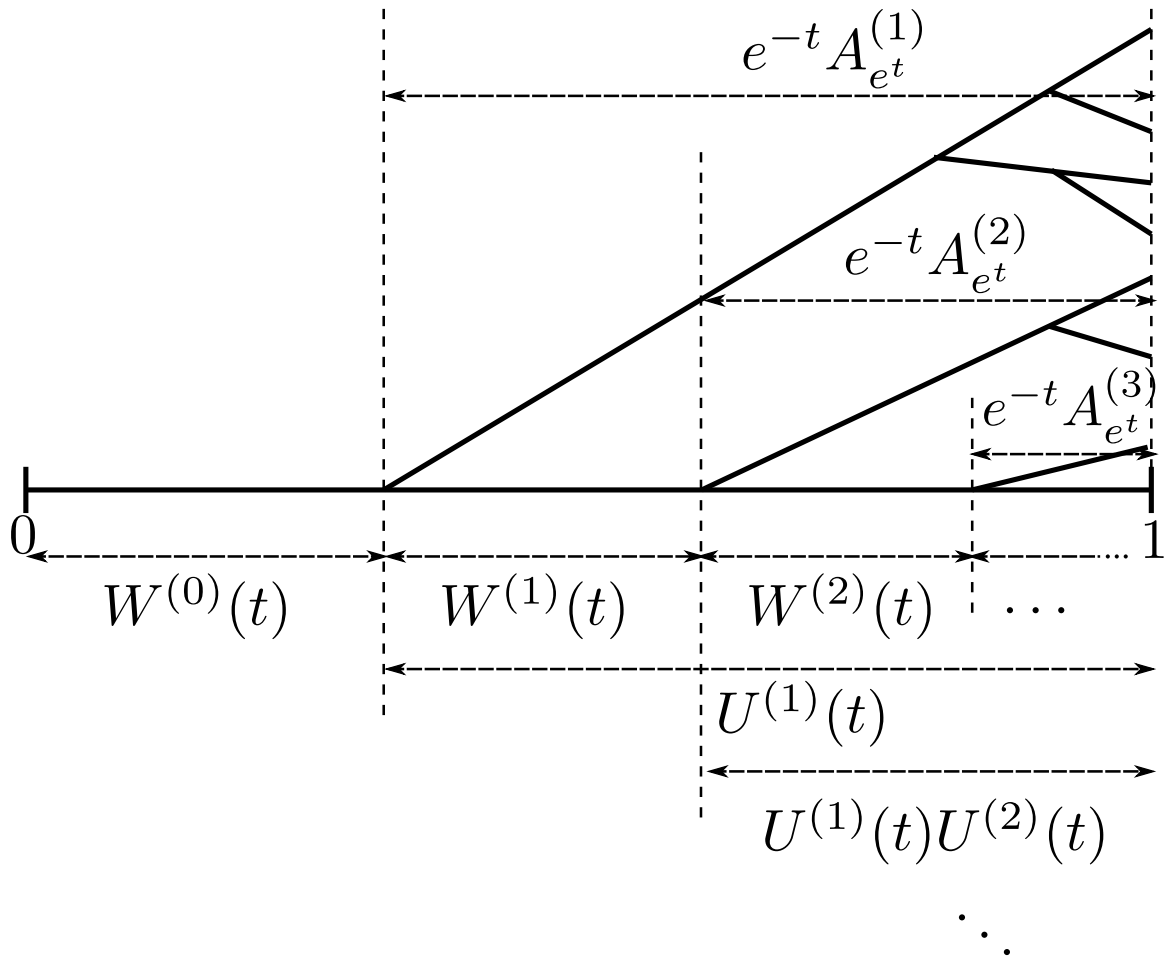
Figure 4.1:

Applying these maps to the process of split times, we get new processes, which are depicted in Figure 4.1, and defined by

$$W_k^{(i)}(t) = w_i(e^{-t}A_{e^t}^{(1)}, e^{-t}A_{e^t}^{(2)}, \ldots, e^{-t}A_{e^t}^{(k)}) \tag{4.2.3}$$

$$U_k^{(i)}(t) = u_i(e^{-t}A_{e^t}^{(1)}, e^{-t}A_{e^t}^{(2)}, \ldots, e^{-t}A_{e^t}^{(k)}). \tag{4.2.4}$$

If we think of the "zeroth split time" as having occurred at $t = 0$, $W^{(i)}$ is the gap between the ith and the $(i+1)^{\text{st}}$ oldest split times and $W_k^{(k)}$ is the remaining amount of time, each rescaled by the total amount of time that has passed. Similarly, $U^{(i)}$ is the ratio of successive split times. Our description of the dynamics of the split times yields the following results.

Theorem 4.2.1. *(I) For each $k \geq 1$, $W_k = (W_k^{(0)}(t), \ldots, W_k^{(k)}(t))_{t \in \mathbb{R}}$ defined in (4.2.3) is a time-homogeneous càdlàg Markov process. The stationary distribution of W_k is a size-biased ordering of a* Dirichlet$(\alpha, \ldots, \alpha)$ *random variable, where $\alpha = (1 + \beta)/\beta k$, and if $A_0^{(1)} = 0$, then the marginal distribution of $(W_k^{(0)}(t), \ldots, W_k^{(k)}(t))$ at each time t has this distribution. The dynamics of W_k can be described as follows:*

(a) *Each $W_k^{(i)}$ moves continuously with drift $-W_k^{(i)}(t)$, except*

(b) *for each $1 \leq i \leq k$, at rate $\left(1 - \sum_{j=0}^{i-1} W_k^{(j)}(t)\right)^{-1}$, let V be an independent* Beta$(1 + 1/\beta, 1)$ *random variable, and let*

$$W_k^{(j)}(t) = \begin{cases} W_k^{(j)}(t-) & \text{if } 1 \leq j < i - 1, \\ W_k^{(i-1)}(t-) + W_k^{(i)}(t-) & \text{if } j = i - 1, \\ W_k^{(j+1)}(t-) & \text{if } i \leq j < k - 1, \\ (1 - V)W_k^{(k)}(t-) & \text{if } j = k - 1, \\ V W_k^{(k)}(t-) & \text{if } j = k. \end{cases} \tag{4.2.5}$$

(II) *For each $k \geq 1$, the process $U_k = (U_k^{(1)}(t), \ldots, U_k^{(k)}(t))_{t \in \mathbb{R}}$ defined in (4.2.4) is a time-homogeneous càdlàg Markov process. The stationary distribution of U_k is that of a sequence of k independent* Beta$(1 + 1/\beta, 1)$ *random variables, and if $A_0^{(1)} = 0$, then the marginal distribution of $(U_k^{(1)}(t), \ldots, U_k^{(k)}(t))$ at each time t has this distribution. The dynamics of U_k can be described as follows:*

(a) *Each $U_k^{(i)}$ moves continuously with drift $(1 - U_k^{(i)}(t))/\prod_{j=1}^{i-1} U_k^{(j)}(t)$, except*

(b) *for each $1 \leq i \leq k$, at rate $\left(\prod_{j=1}^{i} U_k^{(j)}(t)\right)^{-1}$, let V be an independent* Beta$(1 + 1/\beta, 1)$ *random variable, and let*

$$U_k^{(j)}(t) = \begin{cases} U_k^{(j)}(t-) & \text{if } 1 \leq j < i - 1 \\ U_k^{(i-1)}(t-)U_k^{(i)}(t-) & \text{if } j = i - 1 \\ U_k^{(j+1)}(t-) & \text{if } i \leq j < k \\ (1 - V) & \text{if } j = k. \end{cases} \tag{4.2.6}$$

Note that both W and U are "fragmentation-coagulation-drift" processes, in the following sense. W_k describes a partition of $[0, 1]$ in which the breaks between adjacent blocks drift leftwards, and at some rate, two adjacent blocks (additively) coalesce and the

right-most block (additively) splits in two. U_k describes a sequence of Beta$(1 + 1/\beta, 1)$ random variables that each drift larger, and at some rate, adjacent values (multiplicatively) coalesce and the last value (multiplicatively) splits in two.

Proof. Define $X^{(i)}(t) = e^{-t}A^{(i)}_{e^t}$. We know from the proof of Corollary 4.1.1 that if we define B_t by $X_t = 1 - e^{-B_t}$, then B is defined by the "age" construction using the lifetime measure $\rho((b, \infty)) = (1 + 1/\beta)\frac{e^{-b}}{1-e^{-b}}$. The density of ρ is $r(b) = (1 + 1/\beta)\frac{e^{-b}}{(1-e^{-b})^2}$. We defined $W = w(X)$ and $U = u(X)$, and both functions w and u are invertible, so W and U are also time-homogeneous Markov processes.

Since each $B^{(i)}$ drifts larger at constant rate 1, the following calculation shows that each $W^{(i)}$ drifts smaller at rate $-W^{(i)}$:

$$\partial_t \left(e^{-B^{(i+1)}(0)-t} - e^{-B^{(i)}(0)-t} \right)\Big|_{t=0} = -\left(e^{-B^{(i+1)}(0)} - e^{-B^{(i)}(0)} \right)$$
$$= -W^{(i)}(0)$$

Similarly, we may compute the drift of $U^{(i)}$:

$$\partial_t \left(\frac{1 - e^{-B^{(i)}(0)-t}}{1 - e^{-B^{(i-1)}(0)-t}} \right)\Big|_{t=0} = \frac{(1 - e^{-B^{(i-1)}(0)}) - (1 - e^{-B^{(i)}(0)})}{(1 - e^{-B^{(i-1)}(0)})^2}$$
$$= \frac{X^{(i-1)}(0) - X^{(i)}(0)}{X^{(i-1)}(0)^2}$$
$$= \left(1 - U^{(i)}(0) \right) / \prod_{j=1}^{i-1} U^{(j)}(0).$$

As for the jumps, the family corresponding to $B^{(i)}$, when $B^{(i)} = b$, is extinguished at rate

$$r(b)/\rho((b, \infty)) = 1/(1 - e^{-b}),$$

so when $X^{(i)} = x$, it is extinguished at rate $1/x$. Substitution of the definitions of U and W give the death rates of those processes.

When $B^{(i)}$ dies, the remaining families $B^{(i+1)}, \ldots, B^{(k)}$ are renumbered, and if $B^{(k)}(t-) = b$, then the new k^{th}-oldest family has density on $0 < a < b$:

$$f_B(b, a) = \rho((a, \infty)) \exp\left(-\int_a^b \rho((z, \infty))dz \right) = (1 + 1/\beta)\frac{e^{-a}(1 - e^{-a})^{1/\beta}}{(1 - e^{-b})^{1+1/\beta}}.$$

To get the density of the new k^{th} point of X, we need to change variables to $x = 1 - e^{-b}$ and $y = 1 - e^{-a}$, which gives

$$f_X(y, x) = (1 + 1/\beta)(1 - y)y^{1/\beta}x^{-(1+1/\beta)}\frac{1}{1 - y}$$
$$= (1 + 1/\beta)\left(\frac{y}{x}\right)^{1/\beta}\frac{1}{x} \quad \text{for } 0 < y < x.$$

Therefore, the ratio $X^{(k)}(t-)/X^{(k)}(t) = y/x$ has the Beta$(1 + 1/\beta, 1)$ distribution. Under the mappings w and u, this ratio corresponds to the random variable V used in the theorem to define the dynamics of W and U.

The identity of the stationary distributions follows from the fact that a size-biased ordering of a Dirichlet$((1 + \beta)/k\beta, \cdots, (1 + \beta)/k\beta)$ distribution can be represented as a product of independent Beta$(1 + 1/\beta, 1)$ random variables, a fact which can be found, for instance, in [67].

$\square$

It is natural to wish to take the $k \to \infty$ limit in Theorem 4.2.1. Suppose we wish to do this in Part (I), thus obtaining a family of dynamics on $[0, 1]^\infty$ that converges in distribution to the GEM$(1 + 1/\beta)$ distribution. Such a process would have each block drifting smaller at rate proportional to their size, and adjacent blocks coalescing at some other rate while the creation of new blocks with the random variable V would disappear to infinity. However, as was pointed out to us by Jim Pitman, since β only enters the dynamics in the choice of how the final value is split, and so β *does not appear* in the description of the infinite-particle dynamics, namely, each process shares the same limiting dynamics. Therefore, the naive idea that the infinite-particle dynamics would always converge in distribution to the GEM$(1 + 1/\beta)$ distribution cannot be true. We intend to investigate this further in future work.

Finally, we end with a note on the connection of these random partitions to another natural partition obtainable from the branching process. In this case, at $t = 0$ (in our new time scale, corresponding to $t = 1$ in the old scale) we obtain the partition $W^{(0)}(0) + W^{(1)}(0) + W^{(2)}(0) + \cdots = 1$ from the Poisson process on $[0, 1]$ whose points are $A^{(1)}(1), A^{(2)}(1), \ldots$, namely, the x-coordinates of the points lying in the left-leaning wedge $\triangle(1, 0)$. The "blocks" of the partition are the intervals between the x-coordinates of the points.

It is natural to ask if this partition is in some way related to another natural partition related to the branching process. If we label each individual in the population at some time according to the order in which their family split from the immortal lineage, and denote by P_i the proprtions of the ith label in the population, then the P_i are obtained by renormalizing a Poisson process as follows. If for each i we denote by $Z^{(i)}$ the mass of the ith family at time 1, then $P_i = Z_i / \sum_j Z_j$. Given $A^{(i)}(1)$, $Z^{(i)}$ has the distribution of a zero-dimensional squared Bessel process begun at zero and conditioned to be alive at time $A^{(i)}(1)$. Therefore, the points $Z^{(i)}$, $i \geq 1$, form a Poisson process on $(0, \infty)$ whose intensity measure can in principle be determined from ρ and the transition probabilities of the zero-dimensional Bessel process.

Partitions derived from Poisson processes in this way have been studied in generality by Kingman [190], and the relation of these to stick-breaking constructions was investigated by Perman, Pitman, and Yor in [244], in which there are formulas that could be used to find the size-biased distribution of P.

However, there does not seem to be an immediate connection between these two partitions, or a way of defining W so that the methods of [244] apply. Such a connection would be interesting.