The Annals of Probability
2024, Vol. 52, No. 2, 387–433
https://doi.org/10.1214/23-AOP1664

UNIVERSALITY CLASSES FOR THE COALESCENT STRUCTURE OF HEAVY-TAILED GALTON–WATSON TREES

BY SIMON HARRIS[1,a], SAMUEL G. G. JOHNSTON[2,b] AND JUAN CARLOS PARDO[3,c]

[1]*Department of Mathematics, University of Auckland,* [a]*simon.harris@auckland.ac.nz*
[2]*Department of Mathematics, King's College London,* [b]*samuel.g.johnston@kcl.ac.uk*
[3]*Centro de Investigación en Matemáticas,* [c]*jcpardo@cimat.mx*

Consider a population evolving as a critical continuous-time Galton–Watson (GW) tree. Conditional on the population surviving until a large time T, sample k individuals uniformly at random (without replacement) from amongst those alive at time T. What is the genealogy of this sample of individuals? In cases where the offspring distribution has finite variance, the probabilistic properties of the joint ancestry of these k particles are well understood, as seen in (*Ann. Appl. Probab.* **30** (2020) 1368–1414; *Electron. J. Probab.* **24** (2019) 1–35). In the present article, we study the joint ancestry of a sample of k particles under the following regime: the offspring distribution has mean 1 (critical) and the tails of the offspring distribution are *heavy* in that $\alpha \in (1, 2]$ is the supremum over indices β such that the βth moment is finite. We show that for each α, after rescaling time by $1/T$, there is a universal stochastic process describing the joint coalescent structure of the k distinct particles. The special case $\alpha = 2$ generalises the known case of sampling from critical GW trees with finite variance where only pairwise mergers are observed and the genealogical tree is, roughly speaking, some kind of mixture of time-changed Kingman coalescents. The cases $\alpha \in (1, 2)$ introduce new universal limiting partition-valued stochastic processes with interesting probabilistic structures, which, in particular, have representations connected to the Lauricella function and the Dirichlet distribution and whose coalescent structures exhibit multiple-mergers of family lines. Moreover, in the case $\alpha \in (1, 2)$, we show that the coalescent events of the ancestry of the k particles are associated with birth events that produce giant numbers of offspring of the same order of magnitude as the entire population size, and we compute the joint law of the ancestry together with the sizes of these giant births.

1. Introduction.

1.1. *Continuous-time Galton–Watson trees and their coalescent processes.* Let $r > 0$, and let $\overline{p} = (p_i)_{i \in \mathbb{N}_0}$ be a probability mass function on the nonnegative integers. Consider a continuous-time Galton–Watson tree with branching rate r and offspring distribution $\overline{p}$, where we start from a single initial particle at time zero. The initial particle has an exponential lifetime with parameter r (i.e., expected length $1/r$) and upon death is replaced by a random number L of offspring particles, where $\mathbb{P}(L = i) = p_i$. Similarly, each offspring particle independently repeats the behaviour of their parent and so on for all subsequent generations: each particle dies at rate r and upon death is replaced by a random number of offspring distributed like $\overline{p}$. In this process we write Z_t for the number of particles alive at time t.

Continuous-time Galton–Watson trees are endowed with a natural notion of genealogy: each particle living at some time t had a unique ancestor particle living at each earlier time

Received February 2023; revised September 2023.
MSC2020 subject classifications. Primary 60J80, 60G09.
Key words and phrases. Galton–Watson tree, coalescent process, genealogy, spines, regularly varying functions.

$s < t$. It is then natural to ask questions about the shared genealogy of different particles alive in the population alive at a certain time. Specifically, conditioning on the event $\{Z_T \geq k\}$ that there are at least k particles alive at a time $T > 0$, consider picking k particles uniformly at random without replacement from the population alive at time T. Label these k sampled particles with the integers 1 through k. Recalling some standard terminology, a collection of disjoint nonempty subsets of $\{1, \dots, k\}$ whose union is $\{1, \dots, k\}$ is known as a set partition of $\{1, \dots, k\}$. We may associate with our sample of k labelled particles a stochastic process $\pi^{(k,T)} := (\pi_t^{(k,T)})_{t \in [0,T]}$ taking values in the collection of set partitions of $\{1, \dots, k\}$ by declaring

$$\begin{aligned} &i \text{ and } j \text{ in the same block of } \pi_t^{(k,T)} \\ &\quad \iff \quad i \text{ and } j \text{ are descended from the same time } t \text{ ancestor,} \end{aligned} \tag{1}$$

where, more precisely, in (1) we mean that the time T particle labelled with $i \in \{1, \dots, k\}$ and the time T particle labelled with $j \in \{1, \dots, k\}$ share the same unique ancestor in the time t population. This set partition process construction is also seen in, for example, [8] and [21].

Since the entire process begins with a single particle at time 0, it follows that each of the k particles share the same initial ancestor, and accordingly, $\pi_0^{(k,T)} = \{\{1, \dots, k\}\}$, that is, $\pi_0^{(k,T)}$ is the partition of $\{1, \dots, k\}$ into a single block. Conversely, since we choose uniformly without replacement, each of the particles are distinct at time T, hence $\pi_T^{(k,T)} = \{\{1\}, \dots, \{k\}\}$ is the partition of $\{1, \dots, k\}$ into singletons. More generally, as t increases across $[0, T]$, the stochastic process $\pi^{(k,T)}$ takes a range of values in the partitions of $\{1, \dots, k\}$ with the property that the constituent blocks of the process break apart as time passes. With this picture in mind, we define the *split times*

$$\tau_1 < \cdots < \tau_m$$

to be the times of discontinuity of $\pi^{(k,T)}$. That is, at each time τ_i, a block of $\pi_{\tau_i-}^{(k,T)}$ breaks into several smaller blocks in $\pi_{\tau_i}^{(k,T)}$. We note that $\pi^{(k,T)}$ is almost surely right continuous.

Numerous authors have studied the process $\pi^{(k,T)}$ in its various incarnations and in the setting of various continuous-time Galton–Watson trees (see Section 1.6 for further discussion). Harris, Johnston and Roberts [18] studied the large T asymptotics of the process $\pi^{(k,T)}$ in the setting where the offspring distribution is critical (i.e., $\sum_{i \geq 0} i p_i = 1$) with finite variance (i.e., $\sum_{i \geq 0} i(i-1) p_i < \infty$). Under these conditions they established the convergence in distribution of the renormalised process $(\pi_{sT}^{(k,T)})_{s \in [0,1]}$ to a universal stochastic process $\nu^{(k,2)} := (\nu_s^{(k,2)})_{s \in [0,1]}$ taking values in the set of partitions of $\{1, \dots, k\}$. This limiting process $\nu^{(k,2)}$ is universal in the sense that it does not depend on the precise form of the offspring distribution, only that the distribution is (near) critical and has finite variance. Harris et al. [18] show that $\nu^{(k,2)}$ only exhibits binary splits (i.e., every discontinuity amounts to one block breaking into exactly two subblocks) where if there are currently i blocks of sizes $a_1, a_2, \dots, a_i$, the probability the next block to split is block j is $(a_j - 1)/(k - i)$; for any block of size a that splits, the size of its first subblock is uniformly distributed on $\{1, 2, \dots, a - 1\}$, and independently of the block topology, the joint distribution of the $k - 1$ splitting times $0 < \tau_1 < \cdots < \tau_{k-1} < 1$ is given by

$$f_k(t_1, \dots, t_{k-1}) = k! \int_0^\infty \left(\prod_{i=1}^{k-1} \frac{\varphi}{(1 + \varphi(1 - t_i))^2} \right) \frac{1}{(1 + \varphi)^2} \, \mathrm{d}\varphi \, \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1}. \tag{2}$$

Further, when viewed backward in time, this partition process $\nu^{(k,2)}$ has the same topology as Kingman's coalescent [24], that is, any two blocks are equally likely to be the next to merge.

1.2. *Main results.* In the present article, we will consider Galton–Watson trees whose offspring distribution is critical but with heavy tails, ultimately discovering a new collection of *universal stochastic processes* $\{\nu^{(k,\alpha)} := (\nu_s^{(k,\alpha)})_{s\in[0,1]} : \alpha \in (1,2]\}$ that describe their limiting genealogical structures. We define the probability generating function (PGF) of the offspring distribution $\overline{p} = (p_i)_{i\in\mathbb{N}_0}$ by

$$f(s) := \mathbb{E}[s^L] = \sum_{j\geq 0} p_j s^j \quad \text{for } s \in [0,1].$$

Throughout, we will assume $p_0 > 0$. For some $\alpha \in (1,2]$, suppose that the PGF of $\overline{p}$ can be written as

$$\textbf{(H1)} \qquad f(s) = s + (1-s)^\alpha \ell\left(\frac{1}{1-s}\right),$$

where ℓ is a slowly varying function at infinity, that is, for any $\lambda > 0$,

$$\lim_{x\to\infty} \frac{\ell(\lambda x)}{\ell(x)} = 1.$$

Note, such an α will be unique, and it can be verified that, for f of the form in (**H1**), we have $f'(1) = \sum_{j\geq 0} jp_j = 1$, that is, f is the PGF of a critical offspring distribution. The higher moments are slightly more delicate. If $\alpha \in (1,2)$, we will see later that if L is an integer-valued random variable whose moment generating function takes the form in (**H1**), then $\mathbb{E}[L^\kappa] < \infty$ whenever $\kappa < \alpha$, and $\mathbb{E}[L^\kappa] = \infty$ whenever $\kappa > \alpha$. The moment $\mathbb{E}[L^\alpha]$ itself may be either finite or infinite. In the setting where $\alpha = 2$, however, it turns out that moment generating functions of the form in (**H1**) encompass those of all unit-mean probability distributions on the nonnegative integers with finite variance. On the other hand, as we will observe at (3), it is possible to find infinite variance random variables whose moment generating functions take the form (**H1**) with $\alpha = 2$.

We are now ready to state our first main result on universality classes for the coalescent structure of Galton–Watson trees with heavy-tailed offspring distributions.

THEOREM 1.1. *Consider a continuous-time Galton–Watson tree with branching rate r and a critical offspring distribution $\overline{p}$ whose moment generating function satisfies* (**H1**) *for some $\alpha \in (1,2]$. Conditional on $\{Z_T \geq k\}$, let $(\pi_t^{(k,T)})_{t\in[0,T]}$ denote the ancestral process associated with k particles sampled uniformly at random without replacement from the population alive at time T. Then the ancestral process converges in distribution for large times with*

$$\left(\pi_{sT}^{(k,T)}\right)_{s\in[0,1]} \quad \Longrightarrow \quad \left(\nu_s^{(k,\alpha)}\right)_{s\in[0,1]} \quad \text{as } T\to\infty,$$

where, for each $\alpha \in (1,2]$, $(\nu_s^{(k,\alpha)})_{s\in[0,1]}$ is a stochastic process taking values in the set of partitions of $\{1,\dots,k\}$, *which is* universal *in the sense that its law depends only on the moment index α but not on the specific offspring distribution. In other words, for each $\alpha \in (1,2]$, the set of Galton–Watson trees with offspring generating function of the form* (**H1**) *form a* universality class *in terms of their asymptotic sample coalescent structure.*

We remark that, due to the time scaling in T, the branching rate parameter r plays no role in Theorem 1.1. We will describe the complete structure of these universal partition processes $(\nu_s^{(k,\alpha)})_{s\in[0,1]}$ (which can also be thought of as coalescent trees) in Theorem 1.2.

Before discussing distributional properties of the stochastic process $\nu^{(k,\alpha)} := (\nu_s^{(k,\alpha)})_{s\in[0,1]}$ for general α, we take a moment to elucidate further on the case $\alpha = 2$. We will see shortly (c.f. the case $\alpha = 2$ of Theorem 1.2) that the stochastic process $\nu^{(k,2)}$ coincides with the one mentioned above from the work by Harris et al. [18], in particular, that $\nu^{(k,2)}$ has binary splits and the joint distribution of the $k-1$ splits is given by (2). As such, the case $\alpha=2$ of Theorem 1.1 implies the critical case in Harris et al. [18] (Theorem 3), namely, that $\nu^{(k,2)}$ is a limiting object for the coalescent structure of critical trees with finite variance.

In fact, as touched on above, the class of offspring distributions whose moment generating functions of the form (**H1**) with $\alpha=2$ is broader than critical distributions with finite variance. Indeed, there are cases when $\alpha=2$, but the variance is infinite; an explicit example is the offspring distribution with moment generation function

$$f(s)=s+(1-s)^2\left(\frac{1}{2}+\frac{1}{4}\log\left(\frac{1}{1-s}\right)\right); \tag{3}$$

see Slack [33] for further details. As such, our special case $\alpha=2$ of Theorems 1.1 & 1.2 represents a slight generalisation of the critical finite variance case found in [18]. Essentially, we are only able to extend to infinite variance offspring cases in the present paper by introducing discounting of the total population size into various spine changes of measure whilst, perhaps somewhat surprisingly, simultaneously being able to preserve key properties as well as understanding their more complex structures. This novel approach was not featured in [18], where instead some kth moment assumptions were needed combined with truncation approximation argument. An analogous method of spine changes of measure with discounting has been concurrently developed by Harris, Palau, and Pardo [19] for the (significantly different) setting of a discrete time critical Galton–Watson in varying environment with finite variances.

We now turn to describing the processes $\nu^{(k,\alpha)}$ for $\alpha\in(1,2)$, which are more complicated than their $\alpha=2$ counterpart. In the $\alpha\in(1,2)$ setting, the blocks of the stochastic process $(\nu_s^{(k,\alpha)})_{s\in[0,1]}$ may break into three or more subblocks at any splitting event, and consequently, we need to take care to describe the topology of the process.

Let $\nu:=(\nu_s)_{s\in[0,1]}$ be a stochastic process taking values in the set of partitions of $\{1,\dots,k\}$ with the property that ν_0 is one block, ν_1 is singletons, and each discontinuity of ν is a block breaking into several subblocks. Write $0=:\tau_0<\tau_1<\cdots<\tau_m<1$ for the splitting times (i.e., times of discontinuity) of ν. The *topology* $\mathcal{T}(\nu)$ of ν is the sequence of partitions

$$\mathcal{T}(\nu):=(\mathcal{T}_0,\dots,\mathcal{T}_m)\quad\text{with } \mathcal{T}_i:=\nu_{\tau_i}.$$

The resulting sequence $(\mathcal{T}_0,\dots,\mathcal{T}_m)$ is what we call a *splitting sequence* of $\{1,\dots,k\}$. A splitting sequence is a collection of partitions $(\beta_0,\dots,\beta_m)$ of $\{1,\dots,k\}$ such that β_0 is one block, β_m is the singletons, and each β_{i+1} is obtained from β_i by breaking a single block of β_i into two or more subblocks.

Given a splitting sequence $(\beta_0,\dots,\beta_m)$, we define the ith split size, g_i, to be the number of new blocks created at the ith split time, that is,

$$g_i:=\#\beta_i-\#\beta_{i-1}+1\quad i=1,\dots,m,$$

where $\#\beta_i$ is the number of blocks in the partition β_i. Since β_0 contains *one* block, and β_m contains k blocks, $k-1$ blocks are created across the entire sequence, and as such we have

$$\sum_{i=1}^{m}(g_i-1)=k-1.$$

With this notation at hand, we are now ready to state our second main result, characterising the law of the limit processes $(\nu_s^{(k,\alpha)})_{s\in[0,1]}$ occuring in Theorem 1.1.

THEOREM 1.2.

(a) *For* $\alpha \in (1,2)$, *the probability law* $\mathbb{P}^{(k,\alpha)}$ *of* $\nu^{(k,\alpha)}$ *is given by the formula*

$$\begin{aligned}&\mathbb{P}^{(k,\alpha)}\big(\mathcal{T}\big(\nu^{(k,\alpha)}\big)=\overline{\beta}, \tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m\big)\\ &\quad = \frac{1}{(\alpha-1)(k-1)!}\\ &\qquad \cdot \prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)} \cdot \int_0^1 (1-w)^{k-1} w^{m+\frac{2-\alpha}{\alpha-1}} \prod_{i=1}^m (1-wt_i)^{-g_i}\,\mathrm{d}t_i\,\mathrm{d}w,\end{aligned} \tag{4}$$

where $m \geq 1$, $\overline{\beta} = (\beta_0, \dots, \beta_m)$ *is any splitting sequence for* $\{1,\dots,k\}$ *with split sizes* $g_1,\dots,g_m$, *and* $0 < t_1 < \cdots < t_m < 1$ *are splitting times.*

(b) *For* $\alpha = 2$, *here we have* $\mathbb{P}^{(k,2)}(m = k-1, g_i = 2\ \forall i) = 1$, *and we have*

$$\begin{aligned}&\mathbb{P}^{(k,2)}\big(\mathcal{T}\big(\nu^{(k,2)}\big)=\bar{P}, \tau_1 \in \mathrm{d}t_1, \dots, \tau_{k-1} \in \mathrm{d}t_{k-1}\big)\\ &\quad = \frac{2^{k-1}}{(k-1)!}\int_0^1 (1-w)^{k-1} w^{k-1} \prod_{i=1}^{k-1}(1-wt_i)^{-2}\,\mathrm{d}t_i\,\mathrm{d}w,\end{aligned} \tag{5}$$

where $\bar{P}$ *is any splitting sequence of* $\{1,\dots,k\}$ *consisting only of binary splits.*

Let us make a few remarks. First, we note that in the setting of Theorem 1.2, after fixing the split sizes $g_1,\dots,g_m$, the topology $\mathcal{T}(\nu^{(k,\alpha)})$ has no effect on the formula in (4). As such, conditional on the event that $\nu^{(k,\alpha)}$ has m splits at times $t_1,\dots,t_m$ of sizes $g_1,\dots,g_m$, the topology of the process is uniformly distributed on the set of possible splitting sequences with m splits of sizes $g_1,\dots,g_m$. This property may be regarded as a generalisation of the Kingman topology found in the case $\alpha = 2$: going backward in time, whenever a merger of g ancestral lines is about to occur, it is equally likely to consist of any subcollection of g existing ancestral lines.

Next, let us note that the $\alpha = 2$ formula (5) is consistent with the formula for $\alpha \in (1,2)$ in a limit when $\alpha \uparrow 2$. The fact that the formula is asymptotically supported on $\{g_1 = 2,\dots,g_m = 2\}$ when $\alpha = 2$ corresponds to the fact that $\alpha\Gamma(g-\alpha)/\Gamma(2-\alpha) \to 2\mathbf{1}_{\{g=2\}}$ as $\alpha \uparrow 2$.

Additionally, we observe that after taking the change of variable $\varphi = w/(1-w)$ and accounting for the $k!(k-1)!/2^{k-1}$ different binary trees with k labelled leaves and $k-1$ ranked internal nodes, we can verify that (5) is equivalent to the formula from (2).

Recall that a splitting sequence $(\beta_0,\dots,\beta_m)$ is a sequence of partitions of $\{1,\dots,k\}$ such that β_0 is one block, β_m is the singletons, and for each $i \geq 1$, β_i is obtained from β_{i-1} by breaking a single block of β_{i-1} into $g_i \geq 2$ subblocks. Note that β_i has $g_i - 1$ more blocks than β_{i-1} so that

$$\#\beta_j = 1 + \sum_{i=1}^j (g_i - 1) =: k_j.$$

In particular, $k_0 = 1$, and $k_m = k$. We now observe that

$$\#\{\text{Splitting sequences of } \{1,\dots,k\} \text{ with ordered split sizes } (g_1,\dots,g_m)\} = k! \prod_{i=1}^m \frac{k_i}{g_i!}. \tag{6}$$

To see (6), it is best to work backward. Each splitting sequence arises in precisely the following way: starting with $\{1\},\dots,\{k\}$, we choose g_m of k_m blocks to merge first, then g_{m-1} of k_{m-1} blocks to merge next, and continue until on the mth step we merge $g_1 = k_1$ blocks to one block. Thus, the quantity in (6) is equal to $\binom{k_m}{g_m}\binom{k_{m-1}}{g_{m-1}}\cdots\binom{k_1}{g_1}$. Using $k_j - g_j = k_{j-1} - 1$, a series of cancellations occur, and we obtain (6).

Finally, we remark that equation (6) now supplies a way of describing the marginal distribution of the split times and split sizes without any other regard for the topology of the process. Namely, by multiplying (4) by the number $k!\prod_{j=1}^m \frac{k_j}{g_j!}$ of splitting sequences of partitions with order split sizes $(g_1,\dots,g_m)$, we see that the probability density of splits of ordered sizes $(g_1,\dots,g_m)$ in times $\mathrm{d}t_1,\dots,\mathrm{d}t_m$ is given by

$$\begin{aligned}&\mathbb{P}^{(k,\alpha)}(\text{Splits of sizes } (g_1,\dots,g_m) \text{ at times } \mathrm{d}t_1,\dots,\mathrm{d}t_m)\\ &\quad= \frac{k}{(\alpha-1)}\cdot\prod_{i=1}^m \frac{\alpha k_i \Gamma(g_i-\alpha)}{g_i!\Gamma(2-\alpha)}\cdot\int_0^1 (1-w)^{k-1}w^{m+\frac{2-\alpha}{\alpha-1}}\prod_{i=1}^m (1-wt_i)^{-g_i}\,\mathrm{d}t_i\,\mathrm{d}w,\end{aligned}$$

for all $0<t_1<\cdots<t_m<1$, where, as above, for $j=1,\dots,m$, $k_j := 1+\sum_{i=1}^j (g_i-1)$ is the number of blocks after the jth split.

1.3. *The Lauricella representation.* It turns out that it is possible to describe the distribution of the partition process $\nu^{(k,\alpha)}$ in terms of a certain special function known as the *Lauricella* hypergeometric function [12]. The Lauricella function is given by

$$F_D^{(n)}(c,a;b;t) = \int_{\mathbb{T}_n} \big(1-\langle t,x\rangle\big)^{-c} D_n(a;b;x)\,\mathrm{d}x \quad a\in\mathbb{R}^n_{>0}, b>0, c>0, \tag{7}$$

where $\langle t,x\rangle := \sum_{i=1}^n t_i x_i$ and

$$D_n(a;b;x) = \frac{\Gamma(b+\sum_{i=1}^n a_i)}{\Gamma(b)\Gamma(\sum_{i=1}^n a_i)}.\left(1-\sum_{i=1}^n x_i\right)^{b-1}\prod_{i=1}^n x_i^{a_i-1}$$

is the density of the Dirichlet distribution with parameters $(a,b)=(a_1,\dots,a_n,b)$. In Section 5.5 we show that equation (4), describing the law of $\nu^{(k,\alpha)}$, may alternatively be written

$$\begin{aligned}&\mathbb{P}^{(k,\alpha)}(\mathcal{T}(\nu)=\overline{\beta}, \tau_1\in\mathrm{d}t_1,\dots,\tau_m\in\mathrm{d}t_m)\\ &\quad= \frac{1}{(\alpha-1)}\prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}\frac{\Gamma(m+\frac{1}{\alpha-1})}{\Gamma(k+m+\frac{1}{\alpha-1})} &(8)\\ &\quad\quad\times F_D^{(m)}\left[m+\frac{1}{\alpha-1},g_1,\dots,g_m;k+m+\frac{1}{\alpha-1};t_1,\dots,t_m\right]. &(9)\end{aligned}$$

1.4. *Population size and giant birth events.* In addition to describing the joint ancestral structure of k uniformly sampled particles chosen at a large time T from the population of a branching process in the (**H1**) universality class, in the setting $\alpha\in(1,2)$ it transpires that there are also giant birth events occurring in conjunction with the split times.

To explain this connection, let us begin by noting that in Section 4, using tools from Pakes [32], we undertake a careful analysis of the generating functions of branching trees whose offspring generating functions lies in the universality class (**H1**). Ultimately, we show that the survival probability takes the form

$$\overline{F}(T) := \mathbb{P}(Z_T>0) \sim T^{-1/(\alpha-1)}\tilde{\ell}(T),$$

where $\tilde{\ell}(\cdot)$ is another function slowly varying at infinity that may be described explicitly in terms of the slowly varying function $\ell(\cdot)$ occuring in (**H1**); see (59) below.

Consider now that the criticality $\mathbb{E}[L]=1$ of the offspring distribution entails that the expected number of particles alive at time T satisfies $\mathbb{E}[Z_T]=1$ for all T. This implies that $\mathbb{E}[Z_T|Z_T > 0] = \overline{F}(T)^{-1}$, and as such, we may expect that conditional on the event $\{Z_T > 0\}$, the number of particles alive has the order $\overline{F}(T)^{-1}$. Indeed, the following limit due to Pakes [32] states that conditionally on $\{Z_T > 0\}$, $\overline{F}(T)Z_T$ converges to a limit as T gets large, with

$$\lim_{T\to\infty} \mathbb{E}[e^{-\theta\overline{F}(T)Z_T}|Z_T > 0] = 1 - (1 + \theta^{1-\alpha})^{-1/(\alpha-1)}. \tag{10}$$

Recall the process $(\pi_{sT}^{(k,T)})_{s\in[0,1]}$ characterising the joint ancestral structure of a sample of k particles from the population at time T and that we denote by $\tau_1 < \cdots < \tau_m$ the discontinuities, or splitting times, of this process. According to Theorem 1.3, which we state shortly, when T is large, the splitting times $\tau_1 < \cdots < \tau_m$ coincide with seismic birth events of size of the same order of magnitude as the entire population. To formulate this observation precisely, let

$$L_i := \text{Number of individuals born at time } \tau_i, \quad i = 1, \dots, m.$$

Let $(\pi_{sT}^{k,T})_{s\in[0,1]}$ be the rescaled partition process characterising the joint ancestry of k particles sampled from the population at time T. Our next result is an extension of Theorem 1.2.

THEOREM 1.3. *Let $\alpha \in (1, 2)$. Conditioned on the event $\{Z_T \geq k\}$, as $T \to \infty$, we have the convergence in distribution*

$$\big((\pi_{sT}^{k,T})_{s\in[0,1]}, \overline{F}(T)L_1, \dots, \overline{F}(T)L_m\big) \to \big((\nu_s^{(k,\alpha)})_{s\in[0,1]}, X_1, \dots, X_m\big)$$

of the time-rescaled partition process $(\pi_{sT}^{k,T})_{s\in[0,1]}$ together with the $\overline{F}(T)$-rescaled offspring sizes at the split times, where the limiting object $((\nu_s^{(k,\alpha)})_{s\in[0,1]}, X_1, \dots, X_m)$ is a partition process $(\nu_s^{(k,\alpha)})_{s\in[0,1]}$ together with a random vector $(X_1, \dots, X_m)$ of nonnegative random variables, where m is the (random) number of splitting events of $\nu^{k,\alpha}$.

Further, the joint law of $((\nu_s^{(k,\alpha)})_{s\in[0,1]}, X_1, \dots, X_m)$ is given by

$$\begin{aligned} &\mathbb{P}^{(k,\alpha)}\big(\mathcal{T}(\nu) = \overline{\beta}, \tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m, X_1 \in \mathrm{d}x_1, \dots, X_m \in \mathrm{d}x_m\big) \\ &\quad = \frac{\prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}}{(\alpha-1)(k-1)!} \int_0^1 (1-w)^{k-1} w^{m+\frac{2-\alpha}{\alpha-1}} \prod_{i=1}^m (1-wt_i)^{-g_i} \mathrm{d}t_i \, \Delta_{g_i,t_i}^w(x_i) \mathrm{d}x_i \, \mathrm{d}w, \end{aligned} \tag{11}$$

where $\overline{\beta} = (\beta_0, \dots, \beta_m)$ is any splitting sequence for $\{1, \dots, k\}$ with $m \geq 1$ splits having split sizes $g_1, \dots, g_m$, $0 < t_1 < \cdots < t_m < 1$ are split times, $x_1, \dots, x_m \geq 0$ are split offspring sizes, and where

$$\Delta_{g,t}^w(x) := \frac{x^{g-\alpha-1}}{\Gamma(g-\alpha)(1/w-t)^{\frac{g-\alpha}{\alpha-1}}} \exp\left\{-\frac{x}{(1/w-t)^{\frac{1}{\alpha-1}}}\right\} \tag{12}$$

is the probability density function of a standard Gamma random variable with shape parameter $g - \alpha$ and rate parameter $(w/(1-wt))^{\frac{1}{\alpha-1}}$.

Let us take a moment to unpack Theorem 1.3. First, we note that (11) is an immediate generalisation of (4); indeed, integrating through the variables $x_1, \dots, x_m$ of the former, we immediately obtain the latter. Let us comment further that the variable w is like an (unobservable) mixture random variable that acts as a proxy for the entire population size, and after fixing the value of w, the topology and split times have a joint law proportional to $\prod_{i=1}^m \frac{\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}(1-wt_i)^{-g_i}$. After choosing w, the topology and the split sizes $g_1, \dots, g_m$, the offspring events are conditionally gamma distributed with shape parameter $g-\alpha$ and scale factor $(1/w-t)^{\frac{1}{\alpha-1}}$ (equivalently, rate parameter $w^{\frac{1}{\alpha-1}}/(1-wt)^{\frac{1}{\alpha-1}}$).

Let us touch on the interpretation of Theorem 1.3 as $\alpha \uparrow 2$. We recall from our discussion following the statement of Theorem 1.2 that $\nu^{(k,2)}$ only has binary splitting events, that is, when $\alpha = 2$, each $g_i = 2$. In the case $g = 2$, we can interpret the probability measure $\Delta^w_{g,t}(x)\mathrm{d}x$ as approximating the Dirac mass at zero as $\alpha \uparrow 2$. This captures the vanishing of giant birth events as $\alpha \uparrow 2$.

Finally, let us note that using the Laplace transform of the gamma distribution, we have

$$\int_0^\infty e^{-\gamma x}\Delta^w_{g,t}(x)\,\mathrm{d}x = \bigl(1-(1/w-t)^{\frac{1}{\alpha-1}}\gamma\bigr)^{-(g-\alpha)}, \quad \gamma \geq 0.$$

Consequently, for nonnegative parameters $\gamma_1, \dots, \gamma_m$, we may alternatively write

$$\begin{aligned} &\mathbb{E}^{(k,\alpha)}\bigl[e^{-\sum_{j=1}^m \gamma_m X_m}; \mathcal{T}(\nu) = \overline{\beta}, \tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m\bigr] \\ &\quad = \frac{\prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}}{(\alpha-1)(k-1)!}\int_0^1 (1-w)^{k-1} w^{m+\frac{2-\alpha}{\alpha-1}} \prod_{i=1}^m \frac{(1-wt_i)^{-g_i}}{(1-(1/w-t_i)^{\frac{1}{\alpha-1}}\gamma_i)^{g_i-\alpha}}\,\mathrm{d}w. \end{aligned} \tag{13}$$

1.5. *Outline of proofs.* Let us outline briefly our approach to proving Theorems 1.1, 1.2, and 1.3. We begin by introducing a collection of *spines*, which are distinguished lines of descent that flow through a continuous-time Galton–Watson tree forward in time. We adapt and generalise the techniques in Harris et al. [18] by constructing a change of measure $\mathbb{Q}^{(k)}_{\theta,T}$ that encourages the spines to flow through the tree in such a way that at time T they represent a uniform sample of k distinct particles of the tree in such a way that the overall population is size biased by the function $n \mapsto n(n-1)\cdots(n-k+1)e^{-\theta n}$. (Harris et al. [18] have $\theta = 0$ in their set-up.) The parameter θ, which is a discounting parameter controlling the size of the tree, furnishes a simple interpretation of sampling from a k times size biased tree in the absence of a second (let alone kth) moment. Let us mention here that this exponential discounting technique was used by first and third authors in concurrent work with Sandra Palau [19].

A further innovation in our approach here is in accounting for the sizes of the offspring events at spine split times. This tool ultimately leads to the more descriptive limit in Theorem 1.3 of not just the ancestral process of the spines but also the magnitudes of the birth events.

After establishing basic properties of change of measure $\mathbb{Q}^{(k)}_{\theta,T}$, in Section 4 we undertake a careful analysis of the generating functions associated with processes in the universality class (**H1**), which allows us in Section 5 to ultimately tie our work together to prove our main results.

The behaviour of the spines under $\mathbb{Q}^{(k)}_{\theta,T}$ may be understood as a proxy for the ancestral behaviour of uniformly chosen particles under the original measure governing the Galton–Watson tree, though under $\mathbb{Q}^{(k)}_{\theta,T}$ the spines have a tractable distribution which is fairly easily described.

Moreover, under the measure $\mathbb{Q}^{(k)}_{\theta,T}$, we are able to describe in full detail the joint distribution of the entire population size in conjunction with the ancestral behaviour of the spines.

Finally, we mention that various authors, such as Yakymiv [34] and Lagerås and Sagitov [25], have studied the so-called *reduced process* associated with heavy-tailed continuous-time Galton–Watson trees; the reduced process up until a time t is the associated random process of a Galton–Watson process consisting of particles living at times $s \in [0,t]$ who have a descendent alive at time t. In particular, Yakymiv [34], Theorem 3, showed that the reduced process associated with our universality class (**H1**) may be described in terms of a certain deterministic time change of a supercritical Galton–Watson process with a particular offspring distribution. (This offspring distribution also appears in Proposition 19 in Berestycki et al. [7] in the context of Beta coalescents.) In principle, this paves a potential alternative avenue to our intermediate result in Theorem 1.2: in order to obtain the coalescent structure of Galton–Watson trees in our university class (**H1**), one may argue that the coalescent structure of a Galton–Watson tree is identical to that of its reduced tree, and thereafter study the coalescent structure of the particular supercritical tree, and ultimately undo the time change. We nonetheless favour our more direct approach, since it provides a rich probabilistic description of the coalescent structure under the size biased change of measure and, significantly, together with the understanding of the uniform sampling at large times. The latter supplies a more intricate understanding of the relationship between the coalescent structure of the uniform sample and the population size and, in particular, gives us the full description of this relationship afforded by Theorem 1.3.

1.6. *Further discussion of related work.* As mentioned above, countless authors have studied various special cases of the partition process $(\pi_t^{(k,T)})_{t\in[0,T]}$ associated with the ancestry of k uniformly sampled particles from a continuous-time Galton–Watson tree. The case $k=2$, which amounts to sampling two particles from the population and studying their time to most recent common ancestor, is particularly well trodden [3–5, 11, 14, 26, 31]. Work for $k\geq 3$ has appeared only more recently. In the setting of general continuous-time Galton–Watson trees, the second author [21] found an explicit integral formula for the finite dimensional distributions of $\pi^{(k,T)}$ in terms of the generation function $F_t(s) := \mathbb{E}[s^{Z_t}]$ of the process and thereafter considered the large-T asymptotics in the supercritical and subcritical cases; see also Grosjean and Huillet [16] and Le [30]. Zubkov [35] studied a certain aspect of the $k=\infty$ case for critical Galton–Watson trees, showing that asymptotically the most recent common ancestor of the entire population at time T is uniformly distributed on $[0,T]$.

Several related models have also been considered. The first and last author have worked recently on the problem of sampling k particles from a critical branching process in a varying environment (see [19]). A similar result was obtained independently by Boenkost et al. [10]. The second author has worked recently with David Cheek [13] on the problem of sampling a single particle uniformly from the population of a continuous-time Galton–Watson tree and studying the point process of reproduction times along the particles ancestral lineage, and with Amaury Lambert [22] on the ancestry of continuous-state branching processes. The three authors of this manuscript have also considered the problem of sampling k particles from a critical branching process with infinite mean (see [17]). Let us also mention work by Amaury Lambert and coauthors [27, 28] on coalescent point processes associated with trees as well Aldous and Popovic [2] and Gernhard [15].

1.7. *Overview.* We now give the structure of the remainder of the article. In Section 2 we introduce multiple spines and our changes of measure that will underpin our approach. In Section 3 we prove properties of the multiple spines under the change of measure $\mathbb{Q}_{\theta,T}^{(k)}$. Section 4 is dedicated to studying basic properties of trees in the universality class (**H1**) and the large-T asymptotic behaviour of spines under $\mathbb{Q}_{\theta,T}^{(k)}$ for trees in this universality class. In the final section, Section 5, we study the joint law of the population size and the spines under $\mathbb{Q}_{\theta,T}^{(k)}$, thereby ultimately inverting the change of measure.

2. Spines and changes of measures. In this section we introduce various probability measures that will serve as essential tools for our understanding of the genealogies of samples of k individuals drawn from the population without replacement at time T. We start by giving a more precise description of the continuous-time Galton–Watson (GW) population model. Then we will extend the original GW model by identifying k distinguished lines of descent, known as *spines*. Via a change of measure, we will then define our key size-biased and discounted GW process with k-spines under a probability measure $\mathbb{Q}^{(k)}_{\theta,T}$, describing its properties and how it modifies the behaviour of both the population process and the spines.

2.1. *Notation.* Throughout, we use the convention that $\mathbb{Z}^+ := \{0, 1, 2, 3, \dots\}$ and $\mathbb{N} := \{1, 2, 3, \dots\}$. We will make use of the standard Ulam–Harris labelling system to keep track of genealogical information of particles: when an individual labelled u dies and is replaced by L offspring, these are labelled by concatenating the parent label with the number of each child, yielding offspring labels $u1, u2, \dots, uL$ and so on. If there is only one initial particle, for convenience, the single root is usually labelled $\varnothing$. Then, for example, 2.3.1 would represent the first child of the third child of the second child of the initial ancestor. With this notation it is easy to refer to subtrees within, or join subtrees onto, existing trees.

Denote by $\mathcal{Z}_t$ the set of labels of all particles alive at time $t > 0$, and let $Z_t = |\mathcal{Z}_t|$ be the number of particles alive at time $t > 0$. If u is an ancestor of v or equal to v, then we write $u \preceq v$, and if u is a strict ancestor of v, then we write $u \prec v$. An initial labelling $\mathcal{Z}_0$ is said to be *permissible* as long as no initial individual is an ancestor of, or the same as, any other one, that is, $u \npreceq v$ for any distinct pair $u, v \in \mathcal{Z}_0$.

Throughout, we will use a common sample space and σ-algebra $(\Omega, \mathcal{F})$, which is sufficiently enriched to describe the randomness associated with all the various processes encountered below. In the following subsections, we will introduce various probability measures, $\mathbb{P}$, $\mathbb{P}_{\theta,T}$, $\mathbb{P}^{(k)}$, $\mathbb{Q}^{(k)}_{\theta,T}$, and $\mathbb{P}^{(k)}_{\mathrm{unif},T}$, which will all be defined on this same common space $(\Omega, \mathcal{F})$. Considering the same process under different probability laws will be foundational to our approach.

We define $(\mathcal{F}_t, t \geq 0)$ to be the natural filtration of the population process $\mathcal{Z} := (\mathcal{Z}_s)_{s\geq 0}$, that is, $\mathcal{F}_t := \sigma(\mathcal{Z}_s : s \leq t)$. (We will introduce other filtrations in the sequel; see Section 2.4.2.)

2.2. *The continuous-time Galton–Watson process under* $\mathbb{P}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose $\overline{p} = (p_i)_{i\in\mathbb{Z}^+}$ is a probability distribution on $\mathbb{Z}^+$ with finite mean $m := \sum_{i\in\mathbb{Z}^+} i p_i < \infty$.

DEFINITION 2.1 (Galton–Watson process under $\mathbb{P}$). Under the probability measure $\mathbb{P}$, we say that $\mathcal{Z} = (\mathcal{Z}_s)_{s\geq 0}$ is a continuous-time Galton–Watson process with branching rate r and offspring distribution $\overline{p}$, started with one individual, if:

1. The process initially starts with one particle alive (i.e., $Z_0 = 1$ with label $\mathcal{Z}_0 = \{\varnothing\}$).
2. The branching property: Any particles currently alive evolve forward in time independently of one another and of the history of the process.
3. Any particle currently alive at time t undergoes branching at rate r (i.e., the time until the particle branches is exponentially distributed at rate r).
4. Given that particle v branches at time t, it immediately dies and is simultaneously replaced by L_v offspring, where L_v is an independent realisation of the offspring random variable L with $\mathbb{P}(L = i) = p_i$, for $i \in \mathbb{Z}^+$.

Similarly, a Galton–Watson process initially started with $j \geq 2$ particles alive is defined analogously simply by modifying (1), typically with the labelling convention that $\mathcal{Z}_0 = \{1, 2, \dots, j\}$, although any permissible labelling $\mathcal{Z}_0$ can be specified.

2.3. *The Galton–Watson discounted by population size under* $\mathbb{P}_{\theta,T}$. Consider any fixed $T > 0$ and $\theta \geq 0$, and define a new probability measure $\mathbb{P}_{\theta,T}$ on $\mathcal{F}_T$ via the Radon-Nikodym derivative

$$\frac{\mathrm{d}\mathbb{P}_{\theta,T}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_T} := \frac{e^{-\theta Z_T}}{\mathbb{E}[e^{-\theta Z_T}]}. \tag{14}$$

In other words, the law of the process $\mathcal{Z}$ under $\mathbb{P}_{\theta,T}$ is like that of the original law $\mathbb{P}$, except with exponential discounting at rate θ according to the size of population at time T. In fact, the process under $\mathbb{P}_{\theta,T}$ remains a branching process, although the behaviour becomes *time-dependent*, as below (Lemma 2.2).

Introducing discounting by the final population size turns out to be quite natural and indeed will be crucial in allowing our methods to work without any additional moment assumptions on the offspring distribution L, in particular, to encompass the heavy-tailed laws of interest.

The following lemma describes the behaviour of the particles under $\mathbb{P}_{\theta,T}$.

LEMMA 2.2 (Galton–Watson discounted by population size under $\mathbb{P}_{\theta,T}$). *Under the probability law* $\mathbb{P}_{\theta,T}$, *the process* $\mathcal{Z} = (\mathcal{Z}_s)_{s\geq 0}$ *evolves over time period* $[0, T]$ *as a time inhomogeneous Markov branching process*, *where*:

1. *The process initially starts with a single particle alive*.
2. *Any particles currently alive evolve forward in time independently of one another and of the history of the process* (branching property).
3. *Any particle currently alive at time* t *undergoes branching at rate*

$$r\mathbb{E}\big[\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big].$$

4. *Given that particle* v *branches at time* t, *it immediately dies and is simultaneously replaced by* L_v *offspring*, *where* L_v *is an independent realisation of the offspring random variable at time* t, $L(t)$, *where*

$$\mathbb{P}_{\theta,T}\big(L(t) = \ell\big) = p_\ell \frac{(\mathbb{E}[e^{-\theta Z_{T-t}}])^\ell}{\mathbb{E}[(\mathbb{E}[e^{-\theta Z_{T-t}}])^L]}.$$

Lemma 2.2 follows from the case $k = 0$ of Lemma 3.9, which is proven in Section 3.

2.4. *The Galton–Watson process with* k*-spines under* $\mathbb{P}^{(k)}$. For any fixed $k \in \mathbb{N}$, we now proceed to define a measure $\mathbb{P}^{(k)}$ under which the population process $\mathcal{Z}$ has k distinguished lines of descent, known as *spines*. The measure $\mathbb{P}^{(k)}$ will serve as a natural and convenient *reference measure* when looking at the behaviour of other population processes with k distinguished particles (e.g., later we will select a sample of k individuals uniformly at random at time T).

We denote the k spines by $\xi = (\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(k)})$, where $\xi^{(i)}$ corresponds to the distinguished line of decent of the ith *spine*. Each spine is represented by a sequence of Ulam–Harris labels $v_0v_1v_2\ldots$, which start at the initial ancestor and where the next label in the sequence is always an offspring of the previous (i.e., $v_0 = \varnothing$, and for each $i \in \mathbb{N}$, $v_{i+1} = v_i\ell$ for some $\ell \in \{1, \ldots, L_{v_i}\}$). A spine may be an infinite line of descent or a finite path which terminates at a leaf in the underlying genealogical tree of the population. If a particle u has j distinct spines passing though it (i.e., $\#\{i \in \{1, \ldots, k\} : u \in \xi^{(i)}\} = j$), then we say particle u is *carrying* j *spines*.

The process $\mathcal{Z}$ with k spines ξ under measure $\mathbb{P}^{(k)}$ is constructed as a simple extension of $\mathcal{Z}$ under $\mathbb{P}$, in that all particles behave exactly as in the original branching process but some particles are additionally identified as carrying the spines, as follows. First, the population

$\mathcal{Z}$ is constructed according to $\mathbb{P}$. Then, given a realisation of the population $\mathcal{Z}$, the k spine lineages are chosen independently, each spine starts by following the initial ancestor, and then at each subsequent branching event follows one of the offspring chosen uniformly at random (or dies if there are no offspring). More precisely, we have the following construction.

DEFINITION 2.3 (Galton–Watson process with k spines under $\mathbb{P}^{(k)}$). Under probability measure $\mathbb{P}^{(k)}$, the process $\mathcal{Z} = (\mathcal{Z}_s)_{s\geq 0}$ with k spines $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(k)})$ is defined as follows:

1. The process initially starts with one particle alive (i.e., $\mathcal{Z}_0 = \{\varnothing\}$, with $Z_0 = 1$ particle labelled as the root $\varnothing$) which is carrying the k spines.
2. The branching property: Any particle alive—either carrying or not carrying spines—evolves forward in time independently of the other particles in the system and of the history of the process.
3. Any particle currently alive at time t undergoes branching at rate r.
4. Given that a particle v branches at time t, it immediately dies and is simultaneously replaced by L_v offspring, where L_v is an independent realisation of the offspring random variable L with $\mathbb{P}(L = i) = p_i$ for $i \in \mathbb{Z}^+$.
5. Conditional on particle v carrying j spines at the time it branches and on having $L_v = \ell$ offspring:

   (a) If $\ell \geq 1$, each of the j spines chooses independently and uniformly at random which of the L_v offspring to continue to follow.

   (b) $\ell = 0$, there are no offspring for the spines to continue to follow, and those j spine paths terminate and pass into the cemetery state $\overline{\delta}$.

REMARK 1. We note that a Galton–Watson processes with k-spines under $\mathbb{P}^{(k)}$ can also be thought of as a multitype Galton–Watson process, where the type of each particle in $\{0, 1, \dots, k\}$ corresponds to the number of spines passing along it. Here a type j individual can only have offspring types of the same or lower value and where the sum of offspring types must match the parent's type (the number of spines are preserved at branching events). We may enrich this construction further by keeping track of the individual spine trajectories so that the type of each particle could also include of the labels of any spines passing along it (e.g., taking the type space as all possible subsets of $\{1, \dots, k\}$), although again only the number of spines along each particle would affect its rate of branching.

We write $\xi_t = (\xi_t^{(1)}, \dots, \xi_t^{(k)})$ to identify the k spines at time t, where $\xi_t^{(i)}$ is the label of the particle carrying spine i at time t. Since each spine chooses its path independently and uniformly from amongst the available offspring, we can immediately observe that, for any $u_1, \dots, u_k \in \mathcal{Z}_t$ and $\underline{u} = (u_1, \dots, u_k)$,

$$\mathbb{P}^{(k)}(\xi_t = \underline{u} | \mathcal{F}_t) = \prod_{i=1}^{k} \prod_{v \prec u_i} \frac{1}{L_v}. \tag{15}$$

For more details, see the proof of Lemma 6 in Harris et al. [18]. At this stage we also observe that a spine's path may end in a leaf under $\mathbb{P}^{(k)}$.

We emphasise that the law of the underlying Galton–Watson tree is the same under $\mathbb{P}^{(k)}$ as under $\mathbb{P}$, that is

$$\mathbb{P}^{(k)} = \mathbb{P} \quad \text{on } \mathcal{F}_t.$$

We will use this fact without comment in the sequel.

2.4.1. *Some further terminology and notation.* When we are interested in which particular spines pass though a given particle, it is sometimes convenient to think of a particle carrying marks. The k spines are marked (i.e., labelled) as $1, \dots, k$, so we can identify which spines a particle is carrying by their marks. Since all the spines start at the initial ancestor, particle $\varnothing$ carries all k marks, $1, 2, 3, \dots, k$. A particle v through which j spines pass will carry j marks for some $b_1 < b_2 < \cdots < b_j$, where each $b_i \in \{1, \dots, k\}$ uniquely identifes a spine.

The set of distinct spine particles at any time t, and the marks that are following those spine particles, induces a partition $\pi_t^{(k)}$ of $\{1, \dots, k\}$ as follows: we declare i and j to be in the same block of $\pi_t^{(k)}$ if $\xi_t^{(i)} = \xi_t^{(j)}$. The partition-valued stochastic process $(\pi_t^{(k)})_{t\in[0,T]}$ has at most $k-1$ times of discontinuity; we call these times $\tau_1, \dots, \tau_m$ the *split times* of the spines. For convenience we also set $\tau_0 = 0$.

If we then let $\mathcal{T}_i = \pi_{\tau_i}^{(k)}$ for $i = 0, \dots, m$, where m is the number of split times of the partition process, then we have created a splitting sequence of partitions $\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_m$, which describe the topological information about the spines without the information about the spine split times. On occasion in the sequel, it will be useful to consider the σ-algebra $\mathcal{H} = \sigma(\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_m)$.

Further, let n_t be the number of distinct spines (i.e., the number of distinct particles in $\mathcal{Z}_t$ carrying marks) at time t. We also let $\rho_t^{(i)}$ be the total number of spines accompanying spine i at time t, including i itself ($\rho_t^{(i)} = 0$ if $\xi^{(i)} = \overline{\delta}$ corresponding to spine i already in the cemetery state).

For any particle $u \in \mathcal{Z}_t$, there exists a last time at which u was a spine (which may be t). If this time equals τ_i for some i, then we say that u is a *residue* particle; if it does not equal τ_i for any i and u is not a spine, then we say that u is *ordinary*. Each particle is exactly one of residue, ordinary, or a spine (i.e., carrying one or more marks).

2.4.2. *Filtrations.* Whilst all our processes and probability measures are assumed to be carried on a sufficiently rich common space $(\Omega, \mathcal{F})$, it will be very convenient for us to make use of a number of different sub-σ-algebras and filtrations according to how much information we want to know about the particles and spines. To this end, we let:

- $(\mathcal{F}_t^{(k)})_{t\geq 0}$ be the filtration containing all information about the process, including the k spines, up to time t.
- $(\mathcal{F}_t)_{t\geq 0}$ be the filtration containing only the information about the Galton–Watson tree but nothing about the identity of any spines.
- $(\widetilde{\mathcal{G}}_t^{(k)})_{t\geq 0}$ be the filtration containing all the information about the k spines up to time t, including the birth events and numbers of offspring along the k spines, but no information about the rest of the tree.
- $(\mathcal{G}_t^{(k)})_{t\geq 0}$ be the filtration containing information only about the spine splitting events (including which marks follow which spine); $(\mathcal{G}_t^{(k)})_{t\geq 0}$ does not know when births of ordinary particles along the spines occur (i.e., any births coming off the spines when the spines all stay together).

Note, $\mathbb{P}^{(k)}$ can be defined on $\mathcal{F}_\infty^{(k)} \subseteq \mathcal{F}$, and this is the smallest σ-algebra we might use. For further details, see Harris and Roberts [20].

2.5. *The k-spine measure* $\mathbb{Q}_{\theta,T}^{(k)}$. We will now introduce the key probability measure under which the spines will form a uniform choice (without replacement) from the population alive at time T, as required, but where there will also be k-size biasing and discounting by the population size at time T. In particular, the population process with spines under this measure will turn out to have sufficient structural independence properties to greatly facilitate computations and, where the discounting allows us, to develop k-spine methods without requiring any additional moment conditions.

We will often make use of the notation $n^{(k)}$ to represent the number of ways of choosing k distinct objects from n objects, more precisely, for any integers n and k,

$$n^{(k)} := \begin{cases} n(n-1)\cdots(n-k+1) & \text{if } n \geq k \geq 1, \\ 1 & \text{if } n \geq 1, k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let us fix $T > 0$ and $\theta \geq 0$, and introduce a new probability measure $\mathbb{Q}_{\theta,T}^{(k)}$ on $\mathcal{F}_T^{(k)}$ by setting

$$\left.\frac{\mathrm{d}\mathbb{Q}_{\theta,T}^{(k)}}{\mathrm{d}\mathbb{P}^{(k)}}\right|_{\mathcal{F}_T^{(k)}} := \frac{\mathbf{1}_{A_{k,T}}(\prod_{i=1}^{k}\prod_{v \prec \xi_T^{(i)}} L_v)e^{-\theta Z_T}}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]}, \tag{16}$$

where

$$A_{k,T} := \{\text{The } k\text{-spines are alive and following distinct particles at time } T\},$$

and we are noting that $\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}] = \mathbb{E}^{(k)}[Z_T^{(k)}e^{-\theta Z_T}]$, since the underlying Galton–Watson process has the same law under $\mathbb{P}$ as under $\mathbb{P}^{(k)}$.

We now comment on some key properties of $\mathbb{Q}_{\theta,T}^{(k)}$. First, we note that $\mathbb{Q}_{\theta,T}^{(0)} = \mathbb{P}_{\theta,T}$.

Let $\mathcal{Z}_T^{(k)}$ denote the collection of distinct k-tuples of individuals alive at time T so that $\#\mathcal{Z}_T^{(k)} = Z_T^{(k)}$. Using the selection of the spines under $\mathbb{P}^{(k)}$, given the underlying population process (Definition 2.3) and then using (15), we find that

$$\begin{aligned} &\mathbb{E}^{(k)}\left[\mathbf{1}_{A_{k,T}}e^{-\theta Z_T}\prod_{i=1}^{k}\prod_{v \prec \xi_T^{(i)}} L_v \middle| \mathcal{F}_T\right] \\ &\quad = \sum_{u_1,\ldots,u_k \in \mathcal{Z}_T^{(k)}} \mathbb{P}^{(k)}(\xi_T = \underline{u}|\mathcal{F}_T)e^{-\theta Z_T}\prod_{i=1}^{k}\prod_{v \prec u_i} L_v = Z_T^{(k)}e^{-\theta Z_T}. \end{aligned} \tag{17}$$

This confirms that $\mathbb{Q}_{\theta,T}^{(k)}$ is a probability measure and also that

$$\left.\frac{\mathrm{d}\mathbb{Q}_{\theta,T}^{(k)}}{\mathrm{d}\mathbb{P}^{(k)}}\right|_{\mathcal{F}_T} = \frac{Z_T^{(k)}e^{-\theta Z_T}}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]}. \tag{18}$$

In particular, the distribution of the random variable Z_T under $\mathbb{Q}_{\theta,T}^{(k)}$ is that of $\mathbb{P}^{(k)}$ but k-size biased and θ-discounted by the function $n \mapsto n^{(k)}e^{-\theta n}$.

Importantly, for any $u_1, \ldots, u_k \in Z_T$, we have

$$\mathbb{Q}_{\theta,T}^{(k)}(\xi_T = \underline{u}|\mathcal{F}_T) \propto \mathbb{P}^{(k)}(\xi_T = \underline{u}|\mathcal{F}_T)\mathbf{1}_{A_{k,T}}\left(\prod_{i=1}^{k}\prod_{v \prec u_i} L_v\right)e^{-\theta Z_T},$$

then from (15) we deduce the crucial equation

$$\mathbb{Q}_{\theta,T}^{(k)}(\xi_T = \underline{u}|\mathcal{F}_T) = \frac{\mathbf{1}_{A_{k,T}}}{Z_T^{(k)}}, \tag{19}$$

which states the property that, under the measure $\mathbb{Q}_{\theta,T}^{(k)}$, the k-spines are a uniform choice without replacement from those particles alive a time T.

From the previous observation and its definition in (16), we can think of $\mathbb{Q}_{\theta,T}^{(k)}$ by first: (i) k-size biasing and discounting by the population size Z_T, given $\mathcal{F}_T$, and then (ii) choosing

k-spines uniformly without replacement under $\mathbb{Q}^{(k)}_{\theta,T}$, given $\mathcal{F}_T$, that is,

$$\frac{\mathrm{d}\mathbb{Q}^{(k)}_{\theta,T}}{\mathrm{d}\mathbb{P}^{(k)}}\bigg|_{\mathcal{F}^{(k)}_T} = \frac{Z^{(k)}_T e^{-\theta Z_T}}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \cdot \frac{\mathbf{1}_{A_{k,T}}}{Z^{(k)}_T} \cdot \prod_{i=1}^{k} \prod_{v \prec \xi^{(i)}_T} L_v. \tag{20}$$

Also, observe that $\mathbb{Q}^{(k)}_{\theta,T}(Z_T \geq k) = 1$.

Our aim is to provide a complete description of the evolution of the Galton–Watson process with k-spines under $\mathbb{Q}^{(k)}_{\theta,T}$; its desirable properties (including uniformly sampled spines) and tractability, due to the independence within its structure (due to the size-biasing), will prove absolutely crucial to our later analysis of uniform sampling.

We already remarked that a Galton–Watson process with k-spines under $\mathbb{P}^{(k)}_{\theta,T}$ can be thought of as a time inhomogeneous multitype Galton–Watson process, where the type of a particle is the number of spines it carries. Since the branching structure is preserved by the "product" structure of the Radon–Nikodyn derivative in (16), it turns out that the Galton–Watson process with k-spines under $\mathbb{Q}^{(k)}_{\theta,T}$ can be seen as another time inhomogeneous multitype Galton–Watson process (cf. Abraham and Debs [1] for a discrete GW with k-spines described this way). We have the following property.

LEMMA 2.4 (Branching Markov property/Symmetry Lemma). *Suppose that $u \in \mathcal{Z}_t$ is carrying j marks at time t, that is, that u has j spines passing through it. Then under $\mathbb{Q}^{(k)}_{\theta,T}$, the subtree generated by u after time t is independent of the rest of the process and behaves as if under $\mathbb{Q}^{(j)}_{\theta,T-t}$.*

The proof of this Lemma follows from the same arguments of Lemma 8 in Harris et al. [18], where the measure $\mathbb{Q}^{(k)}_{\theta,T}$ is considered without the compensation term θ. Thus, it just remains to identify the branching rates for k-spines to describe $\mathbb{Q}^{(k)}_{\theta,T}$.

The following result gives a full account of the behaviour of the spine and nonspine particles under the change of measure $\mathbb{Q}^{(k)}_{\theta,T}$.

LEMMA 2.5 (Size-biased and discounted Galton–Watson process with k spines under $\mathbb{Q}^{(k)}_{\theta,T}$). *The process $\mathcal{Z} = (\mathcal{Z}_s)_{s\in[0,T]}$ with k spines $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(k)})$ under measure $\mathbb{Q}^{(k)}_{\theta,T}$ evolves as follows*:

1. *The process starts at time* 0 *with one particle carrying all k spines* (*i.e.*, $\mathcal{Z}_0 = \{\varnothing\}$, *and* $\xi_0 = (1, 2, \dots, k)$).
2. *A particle carrying j spines evolves a subtree forward in time independently of the rest of the process* (*branching Markov property*).
3. *A particle carrying $j \geq 1$ spines at time t branches into ℓ offspring, and the j spines split into $g \in \{1, \dots, j\}$ groups of sizes $k_1, \dots, k_g \geq 1$ with $\sum_{i=1}^{g} k_i = j$ at rate*

$$\frac{j!}{\prod_{m=1}^{j} h_m! \prod_{i=1}^{g} k_i!} \cdot \frac{\prod_{i=1}^{g} \mathbb{E}[Z^{(k_i)}_{T-t} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(j)}_{T-t} e^{-\theta Z_{T-t}}]} r\mathbb{E}\big[L^{(g)}\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-g}\big]$$
$$\times \frac{\ell^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-g} p_\ell}{\mathbb{E}[L^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L-g}]},$$

   where $h_m := |\{i : k_i = m\}|$ is the number of spine groups of size m so that $\sum_{m=1}^{j} m h_m = j$ and $\sum_{m=1}^{j} h_m = g$.
4. *Given a particle carrying $j \geq 1$ spines branches into ℓ offspring where the j spines split into g groups of sizes $k_1, \dots, k_g \geq 1$, the spines are assigned between the offspring as follows*:

(a) *Choose g of the ℓ offspring to carry the spine groups uniformly amongst the $\ell!/(g!(\ell-g)!$ distinct ways.*

(b) *Assign the group sizes $k_1, \dots, k_g$ amongst the g offspring chosen to carry the spine groups uniformly amongst the $g!/\prod_{m=1}^{j} h_m!$ distinct allocations.*

(c) *Partition the j (labelled) spines between the g chosen offspring with their given group sizes from $k_1, \dots, k_g$ uniformly amongst the $j!/\prod_{m=1}^{g} k_m!$ distinct ways.*

5. *Finally, any particle v, which is alive at time t and carries no spines, behaves independently of the remainder of the process and undergoes branching into ℓ offspring that carry no spines at rate*

$$r\mathbb{E}\big[\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big] \cdot \frac{(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell} p_\ell}{\mathbb{E}[(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L}]}.$$

For the proof of Lemma 2.5, Part (1) follows from the definition, Part (2) follows from the Markov branching property of Lemma 2.4, Part (3) and Part (4) follow from Lemma 3.6, and finally, Part (5) will be given by Lemma 3.9.

The number of spines following each offspring is of particular interest as whenever spines split apart going forward in time this corresponds to coalescence of family lines when viewing backward in time starting from the individuals sampled at the end. There are various other ways of describing the process with spines, each of which can be extracted from the above description that involves the number of spines particles carry. If instead we are interested in the partitions formed by individual spines and how they break up or want to formulate the process as a multitype Galton–Watson process, the spine splitting rates above can be readily adjusted by the required combinatorial factors.

Observe that the rate a particle carrying j spines at time t branches and all the spines stay together (so $g=1$) is given by

$$r\mathbb{E}\big[L\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big],$$

and the offspring distribution at a branching event at time t when all spine stay together is size-biased and discounted by time to go with the probability of getting ℓ offspring being

$$\frac{\ell(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-1} p_\ell}{\mathbb{E}[L(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L-1}]}.$$

Such branching events, where all the spines follow the same particle ($g=1$), are sometimes referred to as *births off the spine*, as opposed to *spine splitting* branching events where the spine break apart into two or more groups ($g \geq 2$). Importantly here, *the births off any spine branch occur independently of the number of spines following that branch* (i.e., independent of j whenever $g=1$, as above). This very convenient feature means that the subpopulations that have come off any particular spine branch only depend on time period over which that branch has run but not on the number of spines along that branch or whether it changes along it. As the total population is made up of the spine plus the subpopulations coming off each spine branch, this observation will make understanding the total population size relatively straightforward under $\mathbb{Q}^{(k)}_{\theta,T}$.

2.6. *Uniform sampling from a Galton–Watson process.* As was noted in Harris et al. [18], the measure $\mathbb{Q}^{(k)}_{\theta,T}$ provides a tractable tool for studying the forward in time the past genealogies of particles sampled uniformly from the population at time T.

We now introduce a probability measure $\mathbb{P}^{(k)}_{\mathrm{unif},T}$ associated with uniform sampling from our Galton–Watson tree. Let f be a measurable functional on the genealogies of k-tuples of particles. That is, let f be a functional of the Ulam–Harris labelling of the ancestors of the k particles, their birth times, death times, and the number of offspring they have upon death. Let $\xi_T = (\xi_T^{(1)}, \dots, \xi_T^{(k)})$ be a uniform sample without replacement at time T taken from a Galton–Watson process conditioned on the event that $\{Z_T \geq k\}$.

Define the probability measure $\mathbb{P}^{(k)}_{\mathrm{unif},T}$ on $\{Z_T \geq k\}$ as follows:

$$\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T \geq k\big] = \mathbb{E}\bigg[\frac{1}{Z_T^{(k)}} \sum_{\underline{u} \in \mathcal{Z}_T^{(k)}} f(\underline{u})\big|Z_T \geq k\bigg], \tag{21}$$

where the first term of the right-hand side of (21) is the probability for any given choice of distinct $\underline{u} \in \mathcal{Z}_T^{(k)}$. Equivalently, we can make the definition

$$\frac{\mathrm{d}\mathbb{P}^{(k)}_{\mathrm{unif},T}(\cdot|Z_T \geq k)}{\mathrm{d}\mathbb{P}^{(k)}}\bigg|_{\mathcal{F}_T^{(k)}} := \frac{\mathbf{1}_{A_{k,T}}}{Z_T^{(k)}} \cdot \prod_{i=1}^{k} \prod_{v \prec \xi_T^{(i)}} L_v. \tag{22}$$

In particular, recalling (20), we directly see the size-biased relationship between $\mathbb{Q}^{(k)}_{\theta,T}$ and $\mathbb{P}^{(k)}_{\mathrm{unif},T}$ as

$$\frac{\mathrm{d}\mathbb{Q}^{(k)}_{\theta,T}}{\mathrm{d}\mathbb{P}^{(k)}_{\mathrm{unif},T}(\cdot|Z_T \geq k)}\bigg|_{\mathcal{F}_T^{(k)}} = \frac{Z_T^{(k)} e^{-\theta Z_T}}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]}.$$

For the particular case of the splitting times of distinct k-tuples of particles alive at time T under the event of $\{Z_T \geq k\}$, we have the following relationship between $\mathbb{P}^{(k)}_{\mathrm{unif},T}$ and $\mathbb{Q}^{(k)}_{\theta,T}$, which will be very useful for our purposes.

LEMMA 2.6. *Let f be a measurable functional on the genealogies of k-tuples of particles. Then*

$$\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T \geq k\big] = \mathbb{Q}^{(k)}_{\theta,T}\bigg[\frac{f(\xi_T)}{Z_T^{(k)} e^{-\theta Z_T}} 1_{\{Z_T \geq k\}}\bigg] \mathbb{E}\big[Z_T^{(k)} e^{-\theta Z_T}|Z_T \geq k\big]. \tag{23}$$

In particular, if $\tau_1, \dots, \tau_m$ are the split times of the k uniformly chosen individuals $\xi_T \in \mathcal{Z}_T^{(k)}$, then we have

$$\begin{aligned}
&\mathbb{P}^{(k)}_{\mathrm{unif},T}(\tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m|Z_T \geq k) \\
&\quad = \mathbb{Q}^{(k)}_{\theta,T}\bigg(\frac{1}{Z_T^{(k)} e^{-\theta Z_T}} \mathbf{1}_{\{\tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m\}}\bigg) \mathbb{E}\big[Z_T^{(k)} e^{-\theta Z_T}|Z_T \geq k\big].
\end{aligned}$$

PROOF. We first observe from (19) that for a functional along the spines $\xi_T = (\xi_T^{(1)}, \dots, \xi_T^{(k)})$, we have

$$\begin{aligned}
\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)|\mathcal{F}_T\big] &= \mathbb{Q}^{(k)}_{\theta,T}\bigg[\sum_{u \in \mathcal{Z}_T^{(k)}} \mathbf{1}_{\{\xi_T = \underline{u}\}} f(\underline{u})|\mathcal{F}_T\bigg] \\
&= \sum_{\underline{u} \in \mathcal{Z}_T^{(k)}} f(\underline{u}) \mathbb{Q}^{(k)}_{\theta,T}(\xi_T = \underline{u}|\mathcal{F}_T) \\
&= \frac{1}{Z_T^{(k)}} \sum_{\underline{u} \in \mathcal{Z}_T^{(k)}} f(\underline{u}).
\end{aligned}$$

Hence, from (21) and the above identity, it follows that

$$\begin{aligned}\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\ge k\big] &= \mathbb{E}\bigg[\frac{1}{Z_T^{(k)}}\sum_{\underline{u}\in\mathcal{Z}_T^{(k)}} f(\underline{u})|Z_T\ge k\bigg]\\ &= \mathbb{E}\big[\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)|\mathcal{F}_T\big]|Z_T\ge k\big]\\ &= \frac{1}{\mathbb{P}(Z_T\ge k)}\mathbb{E}\big[\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)|\mathcal{F}_T\big]1_{\{Z_T\ge k\}}\big].\end{aligned}$$

Finally, using (18), we now deduce

$$\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\ge k\big] = \mathbb{Q}^{(k)}_{\theta,T}\bigg[\frac{f(\xi_T)}{Z_T^{(k)}e^{-\theta Z_T}}\bigg]\mathbb{E}\big[Z_T^{(k)}e^{-\theta Z_T}|Z_T\ge k\big].$$

This completes the proof. □

As we will show in the next section, one can describe the spine behaviour completely under $\mathbb{Q}^{(k)}_{\theta,T}$ and hence implicitly, thanks to the above result, under $\mathbb{P}^{(k)}_{\mathrm{unif},T}$. We also observe that if we let θ depend on T, according to an appropriate scaling limit, the term on the far right-hand side of (23) can be computed using a relevant Yaglom theorem. (For instance, in the finite variance case, we have Z_T/T converges in distribution to an exponential distribution as $T\to\infty$.) Further, this will similarly be the case for other individual terms appearing in the description of $\mathbb{Q}^{(k)}_{\theta,T}$. This scaling procedure will lead to a limiting k-spine construction and ultimately determine the asymptotic genealogies from uniform samples at large times, as desired.

3. Behaviour of Galton–Watson with k-spines under $\mathbb{Q}^{(k)}_{\theta,T}$.

3.1. *Behaviour of the spines and births off the spine.* In this section we are interested in computing explicitly some functionals of Galton–Watson trees with k-spines under $\mathbb{Q}^{(k)}_{\theta,T}$ that will prove essential in the forthcoming sections. Before delving into the proofs, let us give a brief overview of some key results in this section:

- According to Lemma 3.5, if a particle is carrying $j\ge 1$ spines, then “births-off-the-spine” (i.e., births along this particle after which all of the j spines follow the same offspring particle) occur at rate

$$r\mathbb{E}\big[L\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big]$$

at time t, and when that they do occur, the number of offspring are distributed according to the size-biased probability distribution

$$\mathbb{P}(\text{A birth-off-the-spine at time } t \text{ has size}\ell) = \frac{\ell\mathbb{E}[e^{-\theta Z_{T-t}}]^{\ell-1}]p_\ell}{\mathbb{E}[L\mathbb{E}[e^{-\theta Z_t}]^{L-1}]}.$$

It is important to note that both these quantities (i.e., the rate and offspring distribution) for the births off the spine are independent of j, the number of spines being carried.
- Another main result of this section is Lemma 3.8, which reveals that

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}\big(\tau_1\in \mathrm{d}t_1,\dots,\tau_m\in \mathrm{d}t_m,\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),L_{\tau_1}=\ell_1,\dots,L_{\tau_m}=\ell_m\big)\\ &\quad= \frac{F_T'(e^{-\theta})}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]}\prod_{i=1}^m \ell_i^{(g_i)}\big(\mathbb{E}\big[e^{-\theta Z_{T-t_i}}\big]\big)^{\ell-g}p_{\ell_i}F'_{T-t_i}(e^{-\theta})^{g_i-1},\end{aligned}\tag{24}$$

where $F_t(s)=\mathbb{E}[s^{Z_t}]$, for $|s|<1$ and $t\ge 0$, $0<t_1<\cdots<t_m<T$ are times, $(\beta_0,\dots,\beta_m)$ is a splitting process of $\{1,\dots,k\}$ with split sizes $g_1,\dots,g_m$, and $\ell_1,\dots,\ell_m$ are integers with $\ell_i\ge g_i$.

As mentioned, the spine change of measure results under $\mathbb{Q}^{(k)}_{\theta,T}$ of this section generalise some earlier spine approaches for the continuous time GW found in [21] (Section 4) and [18] (Section 4) with the significant addition of exponentially θ-discounting by the final population size in addition to k-size biasing. Whilst this introduces extra complexity, all key properties are preserved, and the discounting is crucial to permit the spine approach to be applied when heavy-tailed offspring distributions are present.

First, we compute the event that there are no births along the spine by time t. The proof of the following result follows similar arguments as those of Lemma 9 in [18], we provide its proof for the sake of completeness.

LEMMA 3.1. *Let χ_1 be the time of the first birth event in the entire population. Then*

$$\mathbb{Q}^{(k)}_{\theta,T}(\chi_1 > t) = e^{-rt}\frac{\mathbb{E}[Z^{(k)}_{T-t}e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}.$$

PROOF. Recall that $A_{k,T}$ denotes the event that spines are separated and alive by time T. Then

$$\begin{aligned}\mathbb{Q}^{(k)}_{\theta,T}(\chi_1 > t) &= \frac{\mathbb{E}^{(k)}[(\prod_{i=1}^k \prod_{v\prec\xi^{(i)}_T} L_v)e^{-\theta Z_T}\mathbf{1}_{\{\chi_1>t, A_{k,T}\}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \\ &= \mathbb{P}(\chi_1 > t)\frac{\mathbb{E}^{(k)}[(\prod_{i=1}^k \prod_{v\prec\xi^{(i)}_T} L_v)e^{-\theta Z_T}\mathbf{1}_{A_{k,T}}|Z_t = 1]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \\ &= e^{-rt}\frac{\mathbb{E}[Z^{(k)}_{T-t}e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]},\end{aligned}$$

where the third equality follows from the Markov property and identity (17). □

Recall that we call birth events that occur along the spines but which do not occur at spine splitting events, *births off the spine*. The following lemmas tells us the distribution of the number of children at the first birth off spines at time t.

LEMMA 3.2. *Let B_{χ_1} be the event that spines stay together at time χ_1 and L_{χ_1} be the number of offspring at the first birth event*; *then*

$$\mathbb{Q}^{(k)}_{\theta,T}(\chi_1 \in \mathrm{d}t, L_{\chi_1} = \ell; B_{\chi_1}) = \ell\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-1}p_\ell r e^{-rt}\frac{\mathbb{E}[Z^{(k)}_{T-t}e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}\,\mathrm{d}t.$$

PROOF. We first observe that the probability the first particle dies in time dt and has ℓ offspring particles is $re^{-rt}\mathrm{d}tp_\ell$. Moreover, if the first particle has ℓ offspring, then the probability all k spines follow the same one of these offspring is $1/\ell^{k-1}$. Thus, from the Markov property and Lemma 3.1, we have

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}(\chi_1 \in \mathrm{d}t, L_{\chi_1} = \ell; B_{\chi_1}) \\ &\quad= \frac{\mathbb{E}^{(k)}[(\prod_{i=1}^k \prod_{v\prec\xi^{(i)}_T} L_v)e^{-\theta Z_T}\mathbf{1}_{\{A_{k,T};\chi_1\in\mathrm{d}t, L_{\chi_1}=\ell; B_{\chi_1}\}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \qquad (25)\\ &\quad= re^{-rt}\mathrm{d}tp_\ell\left(\frac{1}{\ell}\right)^{k-1}\frac{\mathbb{E}^{(k)}[(\prod_{i=1}^k \prod_{v\prec\xi^{(i)}_T} L_v)e^{-\theta Z_T}\mathbf{1}_{A_{k,T}}|\chi_1\in\mathrm{d}t, L_{\chi_1}=\ell; B_{\chi_1}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}.\end{aligned}$$

Consider that on the event $\{\chi_1 \in \mathrm{d}t, L_{\chi_1} = \ell; B_{\chi_1}\}$, at time t we have $\ell - 1$ nonspine particles alive and a single particle carrying all ℓ spines. Thus, appealing to the identity (17), we have

$$\mathbb{E}^{(k)}\Bigg[\Bigg(\prod_{i=1}^{k} \prod_{v \prec \xi_T^{(i)}} L_v\Bigg) e^{-\theta Z_T} \mathbf{1}_{A_{k,T}} \Bigg| \chi_1 \in \mathrm{d}t, L_{\chi_1} = \ell; B_{\chi_1}\Bigg] \tag{26}$$
$$= \ell^k \big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-1} \mathbb{E}\big[Z_{T-t}^{(k)} e^{-\theta Z_{T-t}}\big].$$

Plugging (26) into (25), we obtain the result. □

From the previous lemma, by summing over all possible values of ℓ, we deduce that

$$\mathbb{Q}_{\theta,T}^{(k)}(\chi_1 \in \mathrm{d}t, B_{\chi_1}) = \mathbb{E}\big[L\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big] r e^{-rt} \frac{\mathbb{E}[Z_{T-t}^{(k)} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]} \,\mathrm{d}t.$$

The following result tells us the probability that the k spines are still following the same particle at time t.

LEMMA 3.3. *Let τ_1 be the first time that the spines split apart. Then*

$$\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t) = \exp\bigg\{-\int_0^t r\big(1 - \mathbb{E}\big[L\big(\mathbb{E}\big[e^{-\theta Z_{T-s}}\big]\big)^{L-1}\big]\big)\,\mathrm{d}s\bigg\} \frac{\mathbb{E}[Z_{T-t}^{(k)} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]}. \tag{27}$$

PROOF. Using the definition of $\mathbb{Q}_{\theta,T}^{(k)}$, we have

$$\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t) = \frac{\mathbb{E}^{(k)}[(\prod_{i=1}^{k} \prod_{v \prec \xi_T^{(i)}} L_v) e^{-\theta Z_T} \mathbf{1}_{\{\tau_1 > t; A_{k,T}\}}]}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]}.$$

Expanding the product, this further reduces to

$$\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t) = \frac{1}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]} \mathbb{E}^{(k)}\Bigg[\mathbf{1}_{\{\tau_1 > t; A_{k,T}\}} \prod_{u \preceq \xi_t^{(1)}} \bigg(\frac{1}{L_u}\bigg)^{k-1}$$
$$\times \Bigg(\prod_{u \preceq \xi_t^{(1)}} L_u^k\Bigg)\Bigg(\prod_{i=1}^{k} \prod_{\xi_t^{(i)} \prec w \prec \xi_T^{(i)}} L_w\Bigg) e^{-\theta Z_{T-t}^{(1)}} \prod_{u \preceq \xi_t^{(1)}} e^{-\theta \sum_{j=1}^{L_u - 1} Z_{T-\sigma_u}^{u,j}}\Bigg],$$

where σ_u denotes the time of the birth off spine of particle u and $Z_{T-\sigma_u}^{u,j}$ represents the contribution of the j-child of the particle u to the population alive at time T. Moreover, $Z_{T-t}^{(1)}$ denotes the contribution of $\xi_t^{(1)}$ to the population alive at time T, implying that Z_T can be rewritten as follows:

$$Z_T = \sum_{u \preceq \xi_t^{(1)}} \sum_{j=1}^{L_u - 1} Z_{T-\sigma_u}^{u,j} + Z_{T-t}^{(1)}.$$

It is important to note that for each node u, the random variables $(Z_{T-\sigma_u}^{u,j}, j \geq 1)$ are i.i.d., conditionally on σ_u, and that for $u \prec v \preceq \xi_t^{(1)}$ the families $(Z_{T-\sigma_u}^{u,j}, j \geq 1)$ and $(Z_{T-\sigma_v}^{v,j}, j \geq 1)$ are independent, conditionally on (σ_u, σ_v).

Using the Markov branching property at time t, we obtain

$$\begin{aligned}
&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1 > t) \\
&= \frac{\mathbb{E}^{(k)}[\mathbf{1}_{\{\tau_1>t;A_{k,T}\}}\prod_{u\preceq\xi_t^{(1)}} L_u e^{-\theta\sum_{i=1}^{L_u-1} Z^{u,i}_{T-\sigma_u}}(\prod_{i=1}^k\prod_{\xi_t^{(i)}\prec v\prec\xi_T^{(i)}} L_v)e^{-\theta Z^{(1)}_{T-t}}]}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]} \qquad (28)\\
&= \mathbb{E}^{(k)}\Big[\prod_{u\preceq\xi_t^{(1)}} L_u e^{-\theta\sum_{i=1}^{L_u-1} Z^{u,i}_{T-\sigma_u}}\Big]\frac{\mathbb{E}[Z^{(k)}_{T-t}e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]}.
\end{aligned}$$

To complete the proof, we use a Campbell formula technique to compute the expectation of the former term above. Namely, let η_t denote the cardinality of $\{u\preceq\xi_t^{(1)}\}$, that is, the number of births off the first spine up until time t. Then since the particle carrying the $\xi_t^{(1)}$ branches at rate r, η_t is Poisson distributed with mean rt. Thus,

$$\mathbb{E}^{(k)}\Big[\prod_{u\preceq\xi_t^{(1)}} L_u e^{-\theta\sum_{i=1}^{L_u-1} Z^{u,i}_{T-\sigma_u}}\Big] = \sum_{j\geq 0}(rtQ)^j e^{-rt}/j! = e^{-rt(1-Q)}, \qquad (29)$$

where Q is the expected contribution to the product of a birth uniformly distributed on $[0,t]$, that is,

$$Q := \frac{1}{t}\int_0^t \mathbb{E}\big[Le^{-\theta\sum_{i=1}^{L-1} Z^i_{T-s}}\big]\,\mathrm{d}s = \frac{1}{t}\int_0^t \mathbb{E}\big[L\mathbb{E}\big[e^{-\theta Z_{T-s}}\big]^{L-1}\big]\,\mathrm{d}s. \qquad (30)$$

Plugging (30) into (29), we obtain the Campbell formula

$$\mathbb{E}^{(k)}\Big[\prod_{u\preceq\xi_t^{(1)}} L_u e^{-\theta\sum_{i=1}^{L_u-1} Z^{u,i}_{T-\sigma_u}}\Big] = \exp\Big\{-r\int_0^t 1-\mathbb{E}\big[L\mathbb{E}\big[e^{-\theta Z_{T-s}}\big]^{L-1}\big]\,\mathrm{d}s\Big\}. \qquad (31)$$

Using (31) in (28), we obtain the result. $\square$

Whilst the above proof provides some clear probabilistic insight, note that an alternative approach to compute the event that all spines are together at time t was given in [18]. More precisely, in the current context, the probability $\mathbb{Q}^{(k)}_{\theta,T}(\tau_1>t)$ can be computed without using Campbell's formula and equivalently written in terms of the derivatives of $F_t(s)=\mathbb{E}(s^{Z_t})$, as follows.

LEMMA 3.4. *Let τ_1 be the first time that spines split apart. Then*

$$\mathbb{Q}^{(k)}_{\theta,T}(\tau_1>t) = \frac{F^{(k)}_{T-t}(e^{-\theta})}{F^{(k)}_T(e^{-\theta})}\frac{F'_T(e^{-\theta})}{F'_{T-t}(e^{-\theta})}, \qquad (32)$$

where $F_t^{(j)}(s) := \frac{\partial^j}{\partial s^j}F_t(s)$.

PROOF. Using the same notation as in the previous lemma, we observe that on the event $\{\tau_1>t\}$ that the spines are following the same particle at time t, we may decompose the population Z_T as

$$Z_T = Z'_T + Z''_T,$$

where Z'_T are the descendents of the unique particle the k spines are following at time t and Z''_T counts the rest of the particles at time T. Observe that $Z''_T = Z_T - Z'_T$ and that, conditionally on $\{\tau_1 > t\}$, Z'_T and Z''_T are independent. Hence,

$$\begin{aligned}
&\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z_T}\mathbf{1}_{\{\tau_1>t;A_{k,T}\}}\right]\\
&\quad=\mathbb{E}^{(k)}\left[\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z_T}\mathbf{1}_{\{\tau_1>t;A_{k,T}\}}\middle|\mathcal{F}_t^{(k)}\right]\right]\\
&\quad=\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_t^{(i)}} L_v\right)\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{\xi_t^{(i)}\preceq v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z'_T}\mathbf{1}_{\{A_{k,T}\}}\middle|\mathcal{F}_t^{(k)}\right]\mathbb{E}[e^{-\theta Z''_T}|\mathcal{F}_t]\mathbf{1}_{\{\tau_1>t\}}\right].
\end{aligned}$$

On the one hand, we have $\mathbb{E}[e^{-\theta Z''_T}|\mathcal{F}_t] = F_{T-t}(e^{-\theta})^{Z_t-1}$. On the other, we have

$$\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{\xi_t^{(i)}\preceq v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z'_T}\mathbf{1}_{\{A_{k,T}\}}\middle|\mathcal{F}_t^{(k)}\right]=\mathbb{E}[Z_{T-t}^{(k)}e^{-\theta Z_{T-t}}],$$

independently of $\mathcal{F}_t^{(k)}$ given $\{\tau_1 > t\}$. Putting all pieces together, we obtain

$$\begin{aligned}
&\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z_T}\mathbf{1}_{\{\tau_1>t;A_{k,T}\}}\right]\\
&\quad=\mathbb{E}[Z_{T-t}^{(k)}e^{-\theta Z_{T-t}}]\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_t^{(i)}} L_v\right)F_{T-t}(e^{-\theta})^{Z_t-1}\mathbf{1}_{\{\tau_1>t\}}\right].
\end{aligned}\tag{33}$$

Finally, by summing over the possible elements of the time t population, we deduce

$$\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_t^{(i)}} L_v\right)\mathbf{1}_{\{\tau_1>t\}}\middle|\mathcal{F}_t\right]=Z_t.\tag{34}$$

Taking $\mathcal{F}_t$-conditional expectations in (33) and using (34), we deduce

$$\begin{aligned}
\mathbb{E}^{(k)}\left[\left(\prod_{i=1}^{k}\prod_{v\prec\xi_T^{(i)}} L_v\right)e^{-\theta Z_T}\mathbf{1}_{\{\tau_1>t;A_{k,T}\}}\right]&=\mathbb{E}[Z_{T-t}^{(k)}e^{-\theta Z_{T-t}}]\mathbb{E}[Z_tF_{T-t}(e^{-\theta})^{Z_t-1}]\\
&=F_{T-t}^{(k)}(e^{-\theta})F'_t(F_{T-t}(e^{-\theta})).
\end{aligned}\tag{35}$$

Let us note that from the semigroup identity $F_T(s) = F_t(F_{T-t}(s))$, we have

$$F'_t(F_{T-t}(s)) = F'_T(s)/F'_{T-t}(s).\tag{36}$$

Substituting (35) into (33) and using (36) allows us to deduce (32). □

Noting that $\mathbb{E}[Z_t^{(k)}e^{-\theta(Z_t-k)}] = F_t^{(k)}(e^{-\theta})$, the agreement of the two representations of $\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t)$ in Lemmas 3.3 and 3.4 amounts to the identity

$$\exp\left\{-\int_0^t r(1-\mathbb{E}[L(\mathbb{E}[e^{-\theta Z_{T-s}}])^{L-1}])\,\mathrm{d}s\right\}=F'_t(F_{T-t}(e^{-\theta})),$$

the proof of which is a standard calculation in the theory of continuous-time Galton–Watson processes; see, for example, Lemma 2.5 of [13].

We continue our exposition by computing the rate of births off the spine.

LEMMA 3.5. *Let σ_1 represent the first time there is a birth off the spine event, that is, a birth event where all the spines stay together. The rate of births occurring off the k spines is given by*

$$\mathbb{Q}_{\theta,T}^{(k)}(\sigma_1 \in \mathrm{d}t; L_{\sigma_1} = \ell|\tau_1 > t) = \ell\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-1} p_\ell r e^{-rt} \frac{F'_{T-t}(e^{-\theta})}{F'_T(e^{-\theta})}. \tag{37}$$

PROOF. In order to deduce this identity, we will use Lemmas 3.2 and 3.4. Observe that under the event $\{\tau_1 > t\}$, the spines stay together by time t and that, by time t, the first birth event and the first birth off spines are the same. More precisely, we have

$$\begin{aligned}
\mathbb{Q}_{\theta,T}^{(k)}(\sigma_1 \in \mathrm{d}t; L_{\sigma_1} = \ell|\tau_1 > t) &= \frac{\mathbb{Q}_{\theta,T}^{(k)}(\sigma_1 \in \mathrm{d}t; L_{\sigma_1} = \ell, B_{\sigma_1}, \tau_1 > t)}{\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t)} \\
&= \frac{\mathbb{Q}_{\theta,T}^{(k)}(\chi_1 \in \mathrm{d}t; L_{\chi_1} = \ell, B_{\chi_1})}{\mathbb{Q}_{\theta,T}^{(k)}(\tau_1 > t)} \\
&= \frac{\ell(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-1} p_\ell r e^{-rt} \frac{\mathbb{E}[Z_{T-t}^{(k)} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]} \mathrm{d}t}{\frac{F_{T-t}^{(k)}(e^{-\theta})}{F_T^{(k)}(e^{-\theta})} \frac{F'_T(e^{-\theta})}{F'_{T-t}(e^{-\theta})}} \\
&= \ell\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-1} p_\ell r e^{-rt} \frac{F'_{T-t}(e^{-\theta})}{F'_T(e^{-\theta})},
\end{aligned}$$

as required. □

Let us make a brief remark interpreting (37). Note that we may alternatively write

$$\begin{aligned}
&\mathbb{Q}_{\theta,T}^{(k)}(\sigma_1 \in \mathrm{d}t; L_{\sigma_1} = \ell|\tau_1 > t) \\
&\quad = \frac{\ell(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-1} p_\ell}{\mathbb{E}[L(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L-1}]} \times \mathbb{E}\big[L\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{L-1}\big] r e^{-rt} \frac{F'_{T-t}(e^{-\theta})}{F'_T(e^{-\theta})} \mathrm{d}t.
\end{aligned} \tag{38}$$

As mentioned in the opening of this section, the first term

$$\frac{\ell(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-1} p_\ell}{\mathbb{E}[L(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L-1}]}$$

in the above identity is a probability mass function in the variable ℓ. The latter term represents the total rate of births off the spine. In other words, the rate of births off a spine branch and number of births off a spine branch are size biased and discounted by time to go and are independent of the number of spines following that branch k.

Now, we compute the probability of the event that spines split at time t with ℓ offspring. Suppose k spines are split into $g \in \{2, \ldots, k\}$ groups of sizes $k_1, k_2, \ldots k_g \geq 1$. Let h_i be the number of groups of size i. We note that

$$\sum_{i=1}^{g} k_i = k \quad \text{and} \quad \sum_{j=1}^{k} j h_j = k.$$

LEMMA 3.6. *Let $C_{g;k_1,\dots,k_g}$ be the event that at the first splitting event, the spines split into g groups of sizes $k_1,\dots,k_g$. Then*

$$\mathbb{Q}^{(k)}_{\theta,T}(\tau_1\in dt; C_{g;k_1,\dots,k_g}; L_{\tau_1}=\ell) = \ell^{(g)}\bigl(\mathbb{E}[e^{-\theta Z_{T-t}}]\bigr)^{\ell-g}p_\ell \frac{k!}{\prod_{j=1}^k h_j!\prod_{i=1}^g k_i!}\frac{\prod_{i=1}^g\mathbb{E}[Z^{(k_i)}_{T-t}e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}\frac{F'_T(e^{-\theta})}{F'_{T-t}(e^{-\theta})}r\,dt.$$

Observe that the term $\ell^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-g}p_\ell$ represents the g-size biased discounted offspring.

PROOF. Recall that $A_{k_j,T}$ represents the event that k_j spines are separated and alive by time T. Then

$$\mathbb{Q}^{(k)}_{\theta,T}(\tau_1\in dt; C_{g;k_1,\dots,k_g}; L_{\tau_1}=\ell) = \frac{\mathbb{E}^{(k)}[(\prod_{i=1}^k\prod_{v\prec\xi_T^{(i)}}L_v)e^{-\theta Z_T};\tau_1\in dt; C_{g;k_1,\dots,k_g}; L_{\tau_1}=\ell; A_{k,T}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}.$$

Decomposing the product and Z_T, we can now write

$$\begin{aligned}
&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1\in dt; C_{g;k_1,\dots,k_g}; L_{\tau_1}=\ell)\\
&\quad=\frac{1}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}\mathbb{E}^{(k)}\Biggl[\prod_{u\prec\xi_t^{(1)}}\Bigl(\frac{1}{L_u}\Bigr)^{k-1}\binom{\ell}{g}\frac{g!}{\prod_{i=1}^k h_i!}\frac{k!}{\prod_{j=1}^g k_j!}\Bigl(\frac{1}{\ell}\Bigr)^k\mathbf{1}_{\{\tau_1\in dt; L_{\tau_1}=\ell\}}\\
&\qquad\times\prod_{i=g+1}^{\ell}e^{-\theta Z^{(i)}_{T-t}}\prod_{u\preceq\xi_t^{(1)}}e^{-\theta\sum_{j=1}^{L_u-1}Z^{u,j}_{T-\sigma_u}}\Biggl(\prod_{j=1}^g\mathbf{1}_{A_{k_j,T}}\prod_{i=1}^{k_j}\prod_{\xi_t^{(j)}\prec v\prec\xi_T^{(j,i)}}L_v\Biggr)\\
&\qquad\times\ell^k\prod_{u\prec\xi_t^{(1)}}L_u^k\Biggl(\prod_{i=1}^g e^{-\theta Z^{(i)}_{T-t}}\Biggr)\Biggr],
\end{aligned}$$

where $\xi_T^{(j,i)}$ denotes the ith spine at time T whose ancestor is $\xi_t^{(j)}$ at time t and, similarly in the proof of Lemma 3.3, σ_u denotes the time of the birth off spine of particle u and $Z^{u,j}_{T-\sigma_u}$ represents the contribution of the j-child of the particle u to the population alive at time T. Moreover, for $1\le i\le\ell$, $Z^{(i)}_{T-t}$ denotes the contribution of the ith offspring at first spine splitting to the population alive at time T. Without loss of generality, we assume that $(Z^{(i)}_{T-t}: 1\le i\le g)$ are the populations associated with the spine particles born at time t and that $(Z^{(i)}_{T-t}: g+1\le i\le\ell)$ are the populations associated with the $\ell-g$ nonspine particles born at time t. In other words, Z_T can be rewritten as follows:

$$Z_T=\sum_{u\preceq\xi_t^{(1)}}\sum_{j=1}^{L_u-1}Z^{u,j}_{T-\sigma_u}+\sum_{i=1}^{\ell}Z^{(i)}_{T-t}.$$

It is important to note that, for each node u, the random variables $(Z^{u,j}_{T-\sigma_u}, j\ge 1)$ are i.i.d., conditionally on σ_u, and that, for $u\prec v\preceq\xi_t^{(1)}$, the families $(Z^{u,j}_{T-\sigma_u}, j\ge 1)$ and $(Z^{v,j}_{T-\sigma_v}, j\ge 1)$ are independent, conditionally on (σ_u,σ_v).

Before we continue with the simplification of the terms inside the expectation, let us provide some explanations about each term. The term $(1/L_u)^{k-1}$ represents the probability that all k spines follow the same particle following a birth of size L_u.

The next few terms are combinatorial: the term $\binom{\ell}{g}$ represents the way we choose offspring which get spine groups, the next term is the way we choose which group gets which group size, and the third term is the number of ways dividing k spines into groups with sizes $k_1, k_2, \dots, k_g$. All three terms together represents the number of ways to get g groups of sizes $k_1, k_2, \dots, k_g$ with ℓ offsprings from k spines. Finally, the fourth term $(1/\ell)^k$ is due to the individual probabilities of each of k spines following one of ℓ possible particles.

The first term in the second row (just after the product sign) represents the contribution at time T of the nonspine individuals. The next term are the contributions to Z_T from births off spines before time t. The three terms inside the brackets provides the contributions from g groups with $k_1, \dots, k_g$ spines after time t. The last two terms are nothing but k spines and follow the same path before time t and the contribution to Z_T of the spines.

Therefore, from the Markov branching property, we obtain

$$\begin{aligned}
&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1 \in \mathrm{d}t; C_{g;k_1,\dots,k_g}; L_{\tau_1} = \ell) \\
&= r\,\mathrm{d}t\,\mathbb{E}\Big[\prod_{u \prec \xi^{(1)}_t} L_u e^{-\theta \sum_{j=1}^{L_u-1} Z^{u,j}_{T-\sigma_u}}\Big] \\
&\quad \times \frac{\ell!}{(\ell-g)!}\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-g} p_\ell \frac{k!}{\prod_{i=1}^k h_i! \prod_{j=1}^g k_j!} \frac{\prod_{i=1}^g \mathbb{E}[Z^{(k_i)}_{T-t} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]}.
\end{aligned}$$

Using the Campbell formula (31), we complete the proof. $\square$

We now derive the third point of Lemma 2.5 from Lemma 3.6. Combining Lemmas 3.3 and 3.6, we obtain

$$\begin{aligned}
&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1 \in \mathrm{d}t; C_{g;k_1,\dots,k_g}; L_{\tau_1} = \ell \mid \tau_1 > t) \\
&\quad = \frac{\ell^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-g} p_\ell}{\mathbb{E}[L^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{L-g}]} \\
&\qquad \times \frac{k!}{\prod_{j=1}^k h_j! \prod_{i=1}^g k_i!} \frac{\prod_{i=1}^g \mathbb{E}[Z^{(k_i)}_{T-t} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_{T-t} e^{-\theta Z_{T-t}}]} r\,\mathrm{d}t.
\end{aligned} \tag{39}$$

Importantly, there is another way of reinterpreting the above result in terms of partitions. Recall that we write $(\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_m)$ for the topology of the spine process, that is, $\mathcal{T}_i$ is the spine partition after the ith split time. Hence, the joint law of the first spine split time τ_1, the offspring size L_{τ_1} at the split as well as $\mathcal{T}_1$, the partition describing the grouping of the spines after this first split satisfies.

LEMMA 3.7. *Let β_1 be a partition of $\{1, \dots, k\}$ into g blocks of sizes $k_1, \dots, k_g$. Then*

$$\begin{aligned}
&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1 \in \mathrm{d}t; \mathcal{T}_1 = \beta_1; L_{\tau_1} = \ell) \\
&\quad = \ell^{(g)}\big(\mathbb{E}\big[e^{-\theta Z_{T-t}}\big]\big)^{\ell-g} p_\ell \frac{\prod_{i=1}^g \mathbb{E}[Z^{(k_i)}_{T-t} e^{-\theta Z_{T-t}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \frac{F'_T(e^{-\theta})}{F'_{T-t}(e^{-\theta})} r\,\mathrm{d}t.
\end{aligned} \tag{40}$$

PROOF. Let $k_1 \geq \dots \geq k_g$ be integers with $k_1 + \dots + k_g = k$. Let $h_j := \#\{i : k_i = j\}$. Then there are

$$\frac{k!}{k_1! \cdots k_g! h_1! \cdots h_k!}$$

different set partitions of $\{1,\dots,k\}$ such that the block sizes, listed in decreasing order, are given by $k_1,k_2,\dots,k_g$. Since the spines are exchangeable, it follows that, given the event in Lemma 3.6, it is equally likely that any of these partitions occur. The result, therefore, follows by dividing through the formula in Lemma 3.6 by the combinatorial factor $k!/(k_1!\cdots k_g!h_1!\cdots h_k!)$. $\square$

We are now ready to state the main result of this section, providing an explicit description of the joint law of the spine ancestry process $(\pi_t^k)_{t\in[0,T]}$ (through its split times $\tau_1,\dots,\tau_m$ and topology $(\beta_0,\dots,\beta_m)$) and the birth sizes $L_{\tau_1},\dots,L_{\tau_m}$ at these split times.

LEMMA 3.8. *Let $0<t_1<\cdots<t_m<T$. Let $(\beta_0,\dots,\beta_m)$ be a splitting process of $\{1,\dots,k\}$, and let $g_1,\dots,g_m$ be the associated split sizes. Let $\ell_1,\dots,\ell_m$ be integers with $\ell_i\ge g_i$. Then*

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}\big(\tau_1\in dt_1,\dots,\tau_m\in dt_m,\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),L_{\tau_1}=\ell_1,\dots,L_{\tau_m}=\ell_m\big)\\ &\quad=\frac{F_T'(e^{-\theta})}{\mathbb{E}[Z_T^{(k)}e^{-\theta Z_T}]}\prod_{i=1}^m \ell_i^{(g_i)}\big(\mathbb{E}[e^{-\theta Z_{T-t_i}}]\big)^{\ell_i-g_i}p_{\ell_i}F'_{T-t_i}\big(e^{-\theta}\big)^{g_i-1}r\,dt_i.\end{aligned}\tag{41}$$

PROOF. We induct on m. The case $m=1$ follows immediately from the previous lemma.

We now assume the result holds for $m'=1,\dots,m-1$ splits, and prove the result holds for $m'=m$ splits. Let $p_1:=\{\Gamma_1,\dots,\Gamma_{g_1}\}$ be the partition after the first split so that Γ_i contains k_i elements. According to the symmetry lemma, after the first split time $\tau_1=t_1$, the spines in each of the g_1 different groups behave independently from one another and as if under $\mathbb{Q}^{(k_i)}_{\theta,T-t_1}$.

Consider now each of the subsequent splitting events $\tau_2,\dots,\tau_m$. Each of these splitting events corresponds to some subblock of some Γ_j ($1\le j\le g_1$) breaking into smaller blocks. As such, we can reindex the $(\tau_i)_{i=2,\dots,m}$ as $(\tau_{i,j})_{1\le i\le m_j,1\le j\le g_1}$, where $m_j\ge 0$ is the number of subsequent splitting events involving a subblock of Γ_j. Specifically, for $i\ge 1$, $\tau_{i,j}$ is the ith time, after τ_1, some subblock of Γ_j (possibly equal to Γ_j itself) breaks into smaller subblocks. Under the same reindexing, we can reindex $(t_i)_{i=2,\dots,m}$ to $(t_{i,j})_{1\le i\le m_j,1\le j\le g_1}$, $(\ell_i)_{i=2,\dots,m}$ to $(\ell_{i,j})_{1\le i\le m_j,1\le j\le g_1}$, and $(g_i)_{i=2,\dots,m}$ to $(g_{i,j})_{1\le i\le m_j,1\le j\le g_1}$.

Let $\mathcal{T}(\xi):=(\mathcal{T}_0,\dots,\mathcal{T}_m)$ be the splitting sequence of the partition $\{1,\dots,k\}$ associated with the spine splitting times. Given a block Γ_j of $\mathcal{T}_1$, we define $\mathcal{T}^{\Gamma_j}(\xi)=(\mathcal{T}_0^{\Gamma_j},\dots,\mathcal{T}_{m_j}^{\Gamma_j})$ to be the splitting sequence of Γ_j associated with any spine splitting event involving a spine in Γ_j after time t_1. $\mathcal{T}^{\Gamma_j}(\xi)=(\mathcal{T}_0^{\Gamma_j},\dots,\mathcal{T}_{m_j}^{\Gamma_j})$ is a sequence of partitions of Γ_j with the property that $\mathcal{T}_0^{\Gamma_j}=\{\Gamma_j\}$, $\mathcal{T}_{m_j}^{\Gamma_j}$ is the singletons, and each $\mathcal{T}_{\ell+1}^{\Gamma_j}$ is obtained from $\mathcal{T}_\ell^{\Gamma_j}$ by breaking precisely one block of $\mathcal{T}_\ell^{\Gamma_j}$ into two or more subblocks.

Let us work through some aspects of the example in Figure 1. Here after the first split, we break into three blocks, $\Gamma_1=\{1\}$, $\Gamma_2=\{2,5\}$, and $\Gamma_3=\{3,4,6,7,8,9\}$. Each block Γ_i has m_i subsequent splits; here $m_1=0$, $m_2=1$, $m_3=2$. We can reindex the subsequent split times τ_2,τ_3,τ_4, so that, for instance, $\tau_{1,3}=\tau_2$ and $\tau_{2,3}=\tau_4$. Let us consider the subsplitting process associated with the block $\Gamma_3=\{3,4,6,7,8,9\}$. This is given by $\mathcal{T}^{\Gamma_3}=(\mathcal{T}_0^{\Gamma_3},\mathcal{T}_1^{\Gamma_3},\mathcal{T}_2^{\Gamma_3})$, where $\mathcal{T}_0^{\Gamma_3}=\{\Gamma_3\}$ is the partition of Γ_3 into a single block, $\mathcal{T}_1^{\Gamma_3}=\{\{3\},\{4,6,7,9\},\{8\}\}$, and finally $\mathcal{T}_2^{\Gamma_3}=\{\{x\}:x\in\Gamma_3\}$ is the singletons.

Write $\Gamma:=\{\tau_1\in dt_1,\dots,\tau_m\in dt_m,\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),L_{\tau_1}=\ell_1,\dots,L_{\tau_m}=\ell_m\}$. We can then decompose the event Γ as

$$\Gamma=\{\tau_1\in dt_1,\mathcal{T}_1=\beta_1,L_{\tau_1}=\ell_1\}\cap\bigcap_{j=1}^{g_1}\Gamma_j,\tag{42}$$

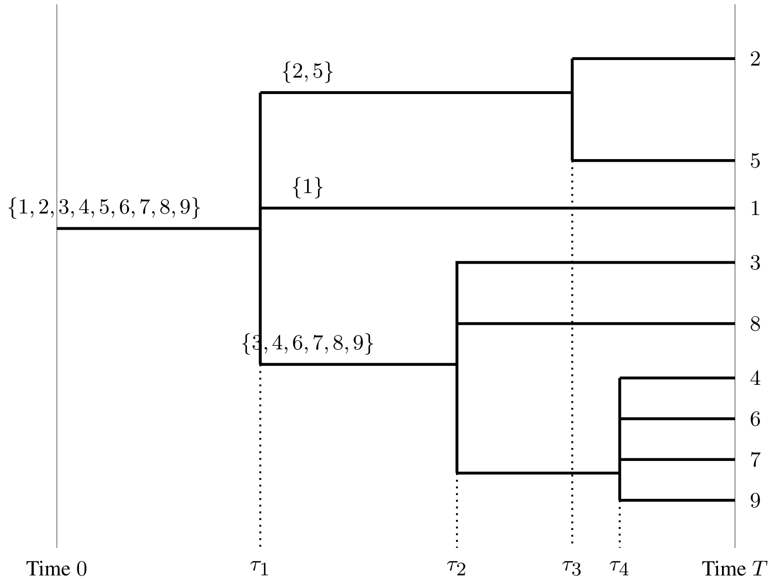

FIG. 1. *The ancestral tree of $k = 9$ spines. After the initial split at time τ_1 into three blocks, the remaining split times and the splitting process can be divided across these three blocks.*

where

$$\Gamma_j = \{\tau_{i,j} \in \mathrm{d}t_{i,j}\ \forall 1 \le i \le m_j, \mathcal{T}^{\Gamma_j} = (\beta_0^{\Gamma_j}, \dots, \beta_{m_j}^{\Gamma_j})\}.$$

Now, according to the symmetry lemma, conditionally on the event $\{\tau_1 \in \mathrm{d}t_1, \mathcal{T}_1 = \beta_1, L_{\tau_1} = \ell_1\}$, the spines in group Γ_j behave as if under $\mathbb{Q}^{(k_1)}_{\theta,T-t_1}$. Using the inductive hypothesis (which we may use since $m_j < m$), we have

$$\begin{aligned} &\mathbb{Q}^{(k)}_{\theta,T}(\Gamma_j | \tau_1 = t_1, \mathcal{T}_1 = \beta_1, L_{\tau_1} = \ell_1) \\ &\quad = \frac{F'_{T-t_1}(e^{-\theta})}{\mathbb{E}[Z^{(k_j)}_{T-t_1} e^{-\theta Z_{T-t_1}}]} \prod_{i=1}^{m_j} \ell_{i,j}^{(g_{i,j})} (\mathbb{E}[e^{-\theta Z_{T-t_{i,j}}}])^{\ell_{i,j}-g_{i,j}} p_{\ell_{i,j}} F'_{T-t_{i,j}}(e^{-\theta})^{g_{i,j}-1}\, \mathrm{d}t_{i,j}, \end{aligned} \tag{43}$$

where we are using the simplification $(T - t_1) - (t_{i,j} - t_1) = T - t_{i,j}$.

Additionally using the fact that the spines in groups $\Gamma_1, \dots, \Gamma_{g_1}$ are independent after time t_1, using (43) in (42), we have

$$\begin{aligned} &\mathbb{Q}^{(k)}_{\theta,T}(\Gamma | \tau_1 = t_1, \mathcal{T}_1 = \beta_1, L_{\tau_1} = \ell_1) \\ &\quad = \prod_{j=1}^{g_1} \frac{F'_{T-t_1}(e^{-\theta})}{\mathbb{E}[Z^{(k_j)}_{T-t_1} e^{-\theta Z_{T-t_1}}]} \\ &\qquad \times \prod_{i=1}^{m_j} \ell_{i,j}^{(g_{i,j})} (\mathbb{E}[e^{-\theta Z_{T-t_{i,j}}}])^{\ell_{i,j}-g_{i,j}} p_{\ell_{i,j}} F'_{T-t_{i,j}}(e^{-\theta})^{g_{i,j}-1}\, \mathrm{d}t_{i,j}. \end{aligned} \tag{44}$$

Reaggregating all of the indices, (44) simplifies to

$$\begin{aligned} &\mathbb{Q}^{(k)}_{\theta,T}(\Gamma | \tau_1 = t_1, \mathcal{T}_1 = \beta_1, L_{\tau_1} = \ell_1) \\ &\quad = \frac{F'_{T-t_1}(e^{-\theta})^{g_1}}{\prod_{j=1}^{g_1} \mathbb{E}[Z^{(k_j)}_{T-t_1} e^{-\theta Z_{T-t_1}}]} \prod_{i=2}^{m} \ell_i^{(g_i)} (\mathbb{E}[e^{-\theta Z_{T-t_{i,j}}}])^{\ell_i-g_i} p_{\ell_i} F'_{T-t_{i,j}}(e^{-\theta})^{g_i-1}\, \mathrm{d}t_i. \end{aligned} \tag{45}$$

Now, according to Lemma 3.7, we have

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}(\tau_1 \in \mathrm{d}t_1; \mathcal{T}_1 = \beta_1; L_{\tau_1} = \ell_1)\\ &\quad = \ell_1^{(g_1)}\big(\mathbb{E}[e^{-\theta Z_{T-t_1}}]\big)^{\ell_1 - g_1} p_\ell \frac{\prod_{i=1}^{g_1} \mathbb{E}[Z^{(k_i)}_{T-t_1} e^{-\theta Z_{T-t_1}}]}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \frac{F'_T(e^{-\theta})}{F'_{T-t_1}(e^{-\theta})} r\,\mathrm{d}t_1.\end{aligned}\tag{46}$$

Multiplying (45) and (46) and using the definition of Γ, we obtain

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}\big(\tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m, \mathcal{T}(\xi) = (\beta_0, \dots, \beta_m), L_{\tau_1} = \ell_1, \dots, L_{\tau_m} = \ell_m\big)\\ &\quad = \frac{F'_T(e^{-\theta})}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \prod_{i=1}^m \ell_i^{(g_i)}\big(\mathbb{E}[e^{-\theta Z_{T-t_i}}]\big)^{\ell_i - g_i} p_{\ell_i} F'_{T-t_i}(e^{-\theta})^{g_i - 1}\,\mathrm{d}t_i,\end{aligned}\tag{47}$$

as required. $\square$

REMARK 2. We remark that one consequence of Lemma 3.8 is that, given that a split of size g_i occurs at time t_i, the size of the birth at this time is distributed according to the probability mass function $\ell \mapsto \ell^{(g)}(\mathbb{E}[e^{-\theta Z_{T-t}}])^{\ell-g} p_\ell/\mathbb{E}[\mathbb{E}[L^{(g)} e^{-\theta Z_{T-t}}]^{L-g}]$. In particular, in the case $g = 1$, we see that births off the spine are (i.e., births after which every spine follows the same particle) are distributed according to the distribution $\ell \mapsto \ell(\mathbb{E}[e^{-\theta Z_{T-t_i}}])^{\ell-1} p_\ell/\mathbb{E}[L\mathbb{E}[e^{-\theta Z_{T-t}}]^{L-1}]$.

REMARK 3. We also note it is possible by summing over ℓ_i to give the following projected version of Lemma 3.8:

$$\begin{aligned}&\mathbb{Q}^{(k)}_{\theta,T}\big(\tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m, \mathcal{T}(\xi) = (\beta_0, \dots, \beta_m)\big)\\ &\quad = \frac{F'_T(e^{-\theta})}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} \prod_{i=1}^m \mathbb{E}\big[L^{(g_i)}\big(\mathbb{E}[e^{-\theta Z_{T-t_i}}]\big)^{L-g_i}\big] F'_{T-t_i}(e^{-\theta})^{g_i-1} r\,\mathrm{d}t_i.\end{aligned}\tag{48}$$

Finally, our results so far have given a detailed description of the spine particles under $\mathbb{Q}^{(k)}_{\theta,T}$. In our final lemma of this section, we describe the behaviour of the *nonspine* particles under $\mathbb{Q}^{(k)}_{\theta,T}$.

LEMMA 3.9. *Under $\mathbb{Q}^{(k)}_{\theta,T}$, nonspine particles behave independently of the other particles and independently of the history of the process. Moreover, at time t any nonspine particles behave like the initial ancestor under a copy of the measure $\mathbb{P}_{\theta,T-t}$ defined in (14).*

In particular, under $\mathbb{Q}^{(k)}_{\theta,T}$, at time t nonspine particles undergo branching at rate

$$r \frac{\mathbb{E}[(\mathbb{E}[e^{-\theta Z_{T-t}}])^L]}{\mathbb{E}[e^{-\theta Z_{T-t}}]},$$

and given that they branch at time t, their offspring distribution is given by

$$\mathbb{P}_{\theta,T}\big(L(t) = \ell\big) = p_\ell \frac{(\mathbb{E}[e^{-\theta Z_{T-t}}])^\ell}{\mathbb{E}[(\mathbb{E}[e^{-\theta Z_{T-t}}])^L]}.$$

PROOF. Suppose v is a nonspine particle alive at time t. Then the Radon–Nikodym derivative of $\mathbb{Q}^{(k)}_{\theta,T}$ against $\mathbb{P}$, defined in (16), may be written

$$\frac{\mathbf{1}_{A_{k,T}}(\prod_{i=1}^k \prod_{u \prec \xi^{(i)}_T} L_u) e^{-\theta Z_T}}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]} = e^{-\theta Z^v_T} \frac{\mathbf{1}_{A_{k,T}}(\prod_{i=1}^k \prod_{u \prec \xi^{(i)}_T} L_u) e^{-\theta(Z_T - Z^v_T)}}{\mathbb{E}[Z^{(k)}_T e^{-\theta Z_T}]},$$

where Z_T^v is the number of particles alive at time t descended from v. Now, under $\mathbb{P}$, since v is not carrying any marks at time t, the random variables $(Z_T - Z_T^v)$ and $\mathbf{1}_{A_{k,T}}(\prod_{i=1}^k \prod_{u \prec \xi_T^{(i)}} L_u)$ are independent of the particle v and of the behaviour of v over the course of $[t, T]$. It follows that we can write

$$\frac{\mathbf{1}_{A_{k,T}}(\prod_{i=1}^k \prod_{u \prec \xi_T^{(i)}} L_u) e^{-\theta Z_T}}{\mathbb{E}[Z_T^{(k)} e^{-\theta Z_T}]} = e^{-\theta Z_T^v} Q_T^{(v)},$$

where $Q_T^{(v)}$ is a random variable independent of v and the descendents of v. By the branching property, under $\mathbb{P}$, Z_T^v is distributed like Z_{T-t}. It follows that, under $\mathbb{Q}_{\theta,T}^{(k)}$, the nonspine particle v behaves independently of the other particles in the system at time t and has its behaviour tilted by the exponential factor $e^{-\theta Z_T^v}$. In short, v behaves like the initial ancestor under a copy of the measure $\mathbb{P}_{\theta,T-t}$.

In order to now describe the branching of v at time t, we, therefore, need only describe the branching of the initial ancestor under a copy of $\mathbb{P}_{\theta,T}$. (Thereafter, we can replace T by $T - t$ as necessary.)

We now note that if $\Gamma(h, \ell)$ is the indicator function of the event that the initial ancestor dies in the time interval $[0, h)$ and has ℓ offspring, then using the definition (14) of the change of measure $\mathbb{P}_{\theta,T}$, we have

$$\mathbb{P}_{\theta,T}\big(\Gamma(h,\ell)\big) := \frac{\mathbb{P}[\Gamma(h,\ell)e^{-\theta Z_T}]}{\mathbb{P}[e^{-\theta Z_T}]}. \tag{49}$$

For small h we then have

$$\begin{aligned}\mathbb{P}\big[\Gamma(h,\ell)e^{-\theta Z_T}\big] &= rp_\ell\big(h + o(h)\big)\mathbb{P}\big[e^{-\theta Z_T}\big|\Gamma(h,\ell)\big] \\ &= rp_\ell\big(h + o(h)\big)\mathbb{P}\big[e^{-\theta Z_{T-h}}\big]^\ell \\ &= rhp_\ell \mathbb{P}\big[e^{-\theta Z_T}\big]^\ell + o(h).\end{aligned}$$

Plugging this into (49), we see that $\mathbb{P}_{\theta,T}(\Gamma(h,\ell)) = rhp_\ell \mathbb{P}[e^{-\theta Z_T}]^\ell / \mathbb{P}[e^{-\theta Z_T}] + o(h)$.

Since under $\mathbb{P}_{\theta,T}$ the particles at time t behave independently and as if under an independent copy of $\mathbb{P}_{\theta,T-t}$, it follows that the time t rate of splitting into ℓ particles is given by $rp_\ell \mathbb{P}[e^{-\theta Z_{T-t}}]^\ell / \mathbb{P}[e^{-\theta Z_{T-t}}]$. Rearranging so that this quantity is proportional to a probability mass function in ℓ, we obtain the stated branching rates and probabilities. $\square$

3.2. *Subpopulation sizes.* It will be important later to understand the sizes of subpopulations coming off the spine under $\mathbb{Q}_{\theta,T}^{(k)}$.

Suppose a nonspine particle v is alive at time t and Z_T^v represents the number of descendants of v alive at time T. Then by Lemma 3.9 and Definition (14), we have

$$\mathbb{Q}_{\theta,T}^{(k)}\big[e^{-\phi Z_T^v}\big|\mathcal{F}_t\big] = \mathbb{P}_{\theta,T-t}\big[e^{-\phi Z_{T-t}}\big] = \frac{\mathbb{E}[e^{-(\phi+\theta)Z_{T-t}}]}{\mathbb{E}[e^{-\theta Z_{T-t}}]}. \tag{50}$$

Similarly, if a particle v alive at time t is carrying j spines and Z_T^v represents the number of descendants of v alive at time T, then recalling Lemma 2.4 and (18), we find

$$\mathbb{Q}_{\theta,T}^{(k)}\big[e^{-\phi Z_T^v}\big|\mathcal{F}_t^{(k)}\big] = \mathbb{Q}_{\theta,T-t}^{(j)}\big[e^{-\phi Z_{T-t}}\big] = \frac{\mathbb{E}[Z_{T-t}^{(j)} e^{-(\phi+\theta)Z_{T-t}}]}{\mathbb{E}[Z_{T-t}^{(j)} e^{-\theta Z_{T-t}}]}.$$

In particular, the Laplace transform for the number of descendents at time T of births coming off a single spine branch ($k = 1$) started at time t (plus the spine itself) is given by

$$\frac{\mathbb{E}[Z_{T-t} e^{-(\phi+\theta)Z_{T-t}}]}{\mathbb{E}[Z_{T-t} e^{-\theta Z_{T-t}}]}. \tag{51}$$

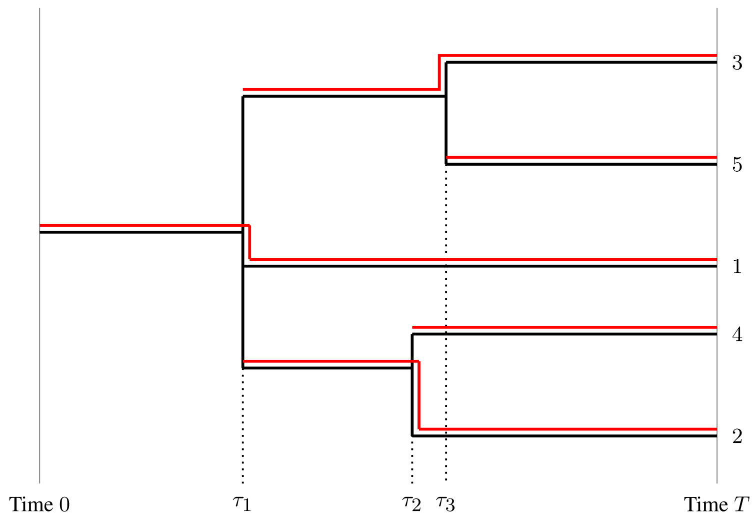

FIG. 2. *A decomposition of the spine tree into lineages. Using* (52), *we see that here the separation times are given by* $\varsigma_1 = 0$, $\varsigma_2 = \tau_1$, $\varsigma_3 = \tau_1$, $\varsigma_4 = \tau_2$, $\varsigma_5 = \tau_3$).

Moreover, since we observed that the rate of births off any spine particle and the corresponding offspring distribution are *independent* of the actual number of spines following it, the above expression remains unchanged for the number of descendants coming off any single spine branch over the time period $[T - t, T]$.

3.3. *Lineage and population decompositions for the ancestral tree.* We now define the subpopulations off spines. We begin by decomposing the spine tree into lineages. Recall that the split times $\tau_1 < \cdots < \tau_m$ associated with the k spines are the times at which a particle carrying some spines dies, and these spines do not all follow the same child. With these in mind, we now define the *separation times* of the spines; see Figure 2. Define $\varsigma_1 := 0$, and for $1 \le j \le k - 1$, define the $(j + 1)$th separation time ς_{j+1} to be the first time that the $(j + 1)$th spine is separate from spines $1, \dots, j$, that is,

$$\varsigma_{j+1} := \inf\{t \ge 0 : \xi_t^{(j+1)} \ne \xi_t^{(i)} \text{ for all } 1 \le i \le j\}. \tag{52}$$

We note that, for all $j \ge 1$, ς_{j+1} is equal to some τ_i. In fact, if g_i is the size of the split at time τ_i, then there are exactly $g_i - 1$ elements $j \in \{1, \dots, k\}$ for which $\varsigma_j = \tau_i$.

With the separation times at hand, we can decompose the entire ancestral tree of the spine into j lineages, where the length of lineage j is given by $T - \varsigma_j$. Of course, then the total length of the tree is given by $kT - \sum_{j=1}^{k} \varsigma_j$. This lineage decomposition affords us a decomposition of the entire population at time T. Giving first an informal definition, let us define

$$\mathcal{Z}_{T,j} := \{u \in \mathcal{Z}_T : u \text{ is a descendant of a birth off the spine of the } j\text{th lineage}\}.$$

Now, let us a give a more precise definition. We say a particle carrying spines is *nonsplitting* if its death is a birth off the spine, that is, at the death of this particle, all of the spines it was carrying follow the same child of this particle. Now, define, for $1 \le j \le k - 1$, $\mathcal{Z}_{T,j+1}$ to be the set of particles at time T (excluding $\xi_T^{(j+1)}$) who are descended from any nonsplitting ancestor of $\xi_T^{(j+1)}$ who is not an ancestor of any $\xi_T^{(1)}, \dots, \xi_T^{(j)}$, that is,

$$\begin{aligned} &\mathcal{Z}_{T,j+1} \\ &\quad := \{u \in \mathcal{Z}_T - \{\xi_T^{(j+1)}\} : \\ &\qquad \exists t, \exists v \in \mathcal{Z}_t \text{ nonsplitting} : v \prec u, v \prec \xi_T^{(j+1)}, v \not\prec \xi_T^{(i)} \ \forall 1 \le i \le j\}. \end{aligned}$$

We now define $\tilde{\mathcal{Z}}_{T,1}, \dots, \tilde{\mathcal{Z}}_{T,m}$ by letting

$$\tilde{\mathcal{Z}}_{T,j} := \{u \in \mathcal{Z}_T : u \text{ is a descendent of a nonspine particle born at splitting event } \tau_j\}.$$

We know that every particle alive at time T is either a spine, a descendent of a birth off the spine, or a descendent of a nonspine particle born at a splitting event. This creates a decomposition of the population at time T via

$$\mathcal{Z}_T := \{\xi_T^{(1)}, \dots, \xi_T^{(k)}\} \cup \bigcup_{j=1}^{k} \mathcal{Z}_{T,j} \cup \bigcup_{j=1}^{m} \tilde{\mathcal{Z}}_{T,j}.$$

Accordingly, if $Z_{T,j} := \#\mathcal{Z}_{T,j}$ and $\tilde{Z}_{T,j} := \#\tilde{\mathcal{Z}}_{T,j}$ denote cardinalities, then we have

$$Z_T = k + \sum_{j=1}^{k} Z_{T,j} + \sum_{j=1}^{m} \tilde{Z}_{T,j}.$$

Further, conditional on the separation times, the Laplace transforms for each subpopulation appearing in this decomposition can readily be written down using (50) and (51).

4. Uniform sampling for critical Galton–Watson trees in the regularly varying regime. Throughout this section, whenever $A(\lambda)$ and $B(\lambda)$ are functions depending on a real or integer-valued parameter, we use the notation

$$A(\lambda) \sim B(\lambda) \quad \text{as } \lambda \to \infty$$

to denote $\lim_{\lambda\to\infty} B(\lambda)/A(\lambda) = 1$. Let us consider a random variable L taking values in $\mathbb{Z}_+ = \{0, 1, \dots\}$ with the same distribution as $L_\varnothing$ and recall that Z denotes a critical Galton–Watson process whose offspring distribution is given by L. In other words,

$$p_n = \mathbb{P}(L = n) \quad \text{for } n \geq 0, \quad \mathbb{E}[L] = \sum_{j\geq 0} j p_j = 1 \quad \text{and} \quad f(s) = \mathbb{E}[s^L].$$

Recall that we are assuming that (**H1**) holds, that is $p_0 > 0$ and that, for $\alpha \in (1, 2]$,

$$f(s) = s + (1-s)^\alpha \ell\left(\frac{1}{1-s}\right),$$

where ℓ is a slowly varying function at ∞. The asymptotic behaviour of f provides enough information about the behaviour of the offspring probabilities p_k, for k large.

Indeed, first let us observe from using the geometric sum and interchanging the order of summation

$$1 - f(s) = \sum_{j\geq 1} p_j (1 - s^j) = (1-s) \sum_{j\geq 1} p_j \sum_{k=0}^{j-1} s^k = (1-s) \sum_{k\geq 0} \overline{p}_k s^k,$$

where $\overline{p}_j := \sum_{i>j} p_i$.

Using a similar argument to obtain the final equality below, we have

$$\frac{f(s) - s}{1-s} = 1 - \frac{1 - f(s)}{1-s} = 1 - \sum_{k\geq 0} \overline{p}_k s^k = (1-s) \sum_{k\geq 0} \overline{\overline{p}}_k s^k,$$

where $\overline{\overline{p}}_k := \sum_{j>k} \overline{p}_j$.

In other words, we have

$$(1-s)^{\alpha-2} \ell\left(\frac{1}{1-s}\right) = \frac{f(s) - s}{(1-s)^2} = \sum_{k\geq 0} \overline{\overline{p}}_k s^k.$$

Hence, Karamata's Tauberian Theorem for power series (see, for instance, Corollary 1.7.3 in Bingham, Goldie and Teugels [9]) allows us to deduce in the case $\alpha \in (1, 2)$ that

$$\overline{\overline{p}}_k \sim \frac{1}{\Gamma(2-\alpha)} k^{1-\alpha}\ell(k) \quad \text{as } k \to \infty \tag{53}$$

and in the case $\alpha = 2$ that

$$\sum_{j=0}^{k} \overline{\overline{p}}_j \sim \ell(k) \quad \text{as } k \to \infty.$$

Now, since the sequence $\{\overline{\overline{p}}_k\}_{k\geq 0}$ is monotone, from the monotone density theorem (see, for instance, Theorem 1.7.2 in [9]) we obtain in the case $\alpha = 2$ that

$$\overline{\overline{p}}_k \sim k^{-1}\ell(k) \quad \text{as } k \to \infty. \tag{54}$$

Consolidating (53) and (54), we have

$$\overline{\overline{p}}_k \sim \tilde{Q}_\alpha k^{1-\alpha}\ell(k) \quad \text{as } k \to \infty, \tag{55}$$

where $\tilde{Q}_\alpha := \frac{1}{\Gamma(2-\alpha)}$ in the case $\alpha \in (1, 2)$, and $Q_2 := 1$.

Since the sequences $\{\overline{p}_k\}_{k\geq 0}$ is monotone, it follows from the monotone density theorem (see, for instance, Theorem 1.7.2 in [9]) we obtain in the case $\alpha \in (1, 2]$ that

$$\overline{p}_k \sim Q_\alpha k^{-\alpha}\ell(k) \quad \text{as } k \to \infty,$$

where $Q_\alpha = (\alpha - 1)\tilde{Q}_\alpha$ so that $Q_\alpha = \frac{\alpha-1}{\Gamma(2-\alpha)}$ when $\alpha \in (1, 2)$ and $Q_\alpha = 1$ when $\alpha = 2$.

Our assumption (**H1**) also implies that the probability of survival is regularly varying at infinity with a precise index that we deduce below. Recall that the probability generating function of Z_t, for $t > 0$, is denoted by $F_t(s)$. According to the backward Kolmogorov equation, $F_t(s)$ satisfies the partial differential equation $\frac{\partial}{\partial t}F_t(s) = r(f(F_t(s)) - F_t(s))$ with boundary condition $F_0(s) = s$, where $f(s)$ is the probability generating function of the offspring distribution; see Athreya and Ney [6], Chapter 3. Writing this partial differential equation in integrated form, we obtain

$$\int_s^{F_t(s)} \frac{\mathrm{d}y}{f(y) - y} = rt.$$

When $s = 0$, it is clear that

$$F(t) := F_t(0) = \mathbb{P}_1(Z_t = 0) = \mathbb{P}_1(T_0 \leq t),$$

where $T_0 = \inf\{s : Z_s = 0\}$ and, therefore,

$$\int_0^{F(t)} \frac{\mathrm{d}y}{f(y) - y} = rt.$$

We also introduce

$$V(x) := \int_0^{1-1/x} \frac{\mathrm{d}y}{f(y) - y} \quad \text{for } x \geq 1,$$

which is increasing and concave, and denote by R its inverse, which is convex and increasing (see Lemma 2.1 in Pakes [32]).

Note then that we have

$$V\big(1/\overline{F}(t)\big) = rt, \tag{56}$$

where $\overline{F}(t) := \mathbb{P}_1(Z_t > 0)$.

Under our assumptions, the function V can be rewritten as follows:

$$V(x) = \int_1^x \frac{\mathrm{d}u}{u^{2-\alpha}\ell(u)}.$$

Following the same ideas as in the proof of Proposition 1.5.8, in [9] we see that

$$V(x) \sim \frac{x^{\alpha-1}}{(\alpha-1)\ell(x)} = x^{\alpha-1}\ell_1(x) \quad \text{as } x \to \infty,$$

where $\ell_1(x) = \frac{1}{(\alpha-1)\ell(x)}$ is also slowly varying. From Theorem 1.5.12 in [9], the inverse function R of V is regularly varying at infinity with index $1/(\alpha-1)$, that is,

$$R(x) \sim (\alpha-1)^{\frac{1}{\alpha-1}} x^{\frac{1}{\alpha-1}} \ell_2(x) \quad \text{as } x \to \infty, \tag{57}$$

where ℓ_2 is another slowly varying function at ∞ and the constant $(\alpha-1)^{\frac{1}{\alpha-1}}$ in front is chosen to lighten formulas in the sequel. Moreover, a calculation using $R(V(x)) = 1$ tells us that

$$\ell_2(x) \sim \ell\big(x^{\frac{1}{\alpha-1}}\big)^{\frac{1}{\alpha-1}}. \tag{58}$$

We refer the reader to Theorem 1.5.13 and Proposition 1.5.15 in [9] for further information on inverses of regularly varying functions and the associated *De Bruijn conjugation*.

Thus, under assumption (**H1**), by applying R to both sides of (56) and thereafter using (57), we obtain

$$\overline{F}(t) \sim \frac{1}{R((\alpha-1)rt)} = c_{\alpha,r} t^{-1/(\alpha-1)}/\ell_2(t) \quad \text{as } t \to \infty, \tag{59}$$

where $c_{\alpha,r} = ((\alpha-1)r)^{-1/(\alpha-1)}$.

We will also use the fact that

$$\mathbb{E}\big[e^{-\theta\overline{F}(t)Z_t} | Z_t > 0\big] \xrightarrow[t\to\infty]{} 1 - \big(1+\theta^{1-\alpha}\big)^{-1/(\alpha-1)}; \tag{60}$$

see, for example, Theorem 3.1 of Pakes [32].

Let us denote by $W_{\alpha-1}$ for the r.v. whose Laplace transform is such that

$$\mathbb{E}\big[e^{-\theta W_{\alpha-1}}\big] = 1 - \big(1+\theta^{1-\alpha}\big)^{-1/(\alpha-1)}.$$

We note that when $\alpha = 2$, the limiting random variable W_1 is an exponential random variable with parameter equals 1 regardless of whether σ^2 is finite or infinite.

4.1. *Properties of $W_{\alpha-1}$.* When $\alpha = 2$, it is not so difficult to verify that W_1 possesses all positive moments. More precisely, for $k \geq 1$, we have

$$\mathbb{E}\big[W_1^k\big] = k! \quad \text{and} \quad \mathbb{E}\big[W_1^k e^{-\theta W_1}\big] = \frac{k!}{(1+\theta)^{k+1}}. \tag{61}$$

The compensated moments of $W_{\alpha-1}$ in the setting $\alpha \in (1,2)$ are more involved. To compute these, write

$$f_\alpha(\theta) := \mathbb{E}\big[e^{-\theta W_{\alpha-1}}\big] = 1 - \big(1+\theta^{1-\alpha}\big)^{-1/(\alpha-1)},$$

where $g_\alpha(\theta) = 1 - (1+\theta)^{-\frac{1}{\alpha-1}}$ and $h_\alpha(\theta) = \theta^{1-\alpha}$. We can compute the derivatives of $f_\alpha(\theta)$ using Faà di Bruno formula's (equation (2) of [23]), which states that given sufficiently differentiable functions f and g we have

$$\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} g\big(h(\theta)\big) = \sum_{\pi\in\mathcal{P}_k} g^{(\#\pi)}\big(h(\theta)\big) \prod_{\Gamma\in\pi} h^{(\#\Gamma)}(\theta),$$

where the sum is taken over all set partitions π of $\{1,\dots,k\}$, and given a partition π, the product is taken over all blocks Γ of π.

Indeed, note the jth derivatives of g_α and h_α are given by

$$g_\alpha^{(j)}(\theta) = (-1)^{j-1}(1+\theta)^{-\frac{1}{\alpha-1}-j}\prod_{i=1}^{j}\left(\frac{1}{\alpha-1}+i-1\right),$$

and

$$h_\alpha^{(j)}(\theta) = (-1)^{j}\theta^{1-\alpha-j}\prod_{i=1}^{j}(\alpha+i-2).$$

It follows from Faà di Bruno's formula that

$$\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} f_\alpha(\theta) = \sum_{\pi\in\mathcal{P}_k}(-1)^{\#\pi-1}\left(1+\theta^{1-\alpha}\right)^{-\frac{1}{\alpha-1}-\#\pi}\prod_{i=1}^{\#\pi}\left(\frac{1}{\alpha-1}+i-1\right) \times \prod_{\Gamma\in\pi}\left\{(-1)^{\#\Gamma}\theta^{1-\alpha-\#\Gamma}\prod_{i=1}^{\#\Gamma}(\alpha+i-2)\right\},$$

which, using $\sum_{\Gamma\in\pi}\#\Gamma = k$, simplifies to

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} f_\alpha(\theta) &= (-1)^{k-1}\theta^{-k}\sum_{\pi\in\mathcal{P}_k}(-1)^{\#\pi}\left(1+\theta^{1-\alpha}\right)^{-\frac{1}{\alpha-1}-\#\pi}\theta^{(1-\alpha)\#\pi} \\ &\quad \times \prod_{i=1}^{\#\pi}\left(\frac{1}{\alpha-1}+i-1\right)\prod_{\Gamma\in\pi}\prod_{i=1}^{\#\Gamma}(\alpha+i-2).\end{aligned} \tag{62}$$

It follows that the first compensated moment is

$$\mathbb{E}\left[W_{\alpha-1}e^{-\theta W_{\alpha-1}}\right] = -f'_\alpha(\theta) = \frac{1}{(1+\theta^{\alpha-1})^{1+1/(\alpha-1)}}, \tag{63}$$

and, in particular, $\mathbb{E}[W_{\alpha-1}] = 1$. For the second compensated moment, we get

$$\mathbb{E}\left[W_{\alpha-1}^2 e^{-\theta W_{\alpha-1}}\right] = f''_\alpha(\theta) = \frac{\alpha}{(1+\theta^{\alpha-1})^{2+1/(\alpha-1)}\theta^{2-\alpha}}.$$

Finally, the third compensated moment satisfies

$$\begin{aligned}\mathbb{E}\left[W_{\alpha-1}^3 e^{-\theta W_{\alpha-1}}\right] = -f_\alpha^{(3)}(\theta) &= \frac{\alpha}{(1+\theta^{\alpha-1})^{2+1/(\alpha-1)}}\left(\frac{2\alpha-1}{(1+\theta^{\alpha-1})\theta^{2(2-\alpha)}}+\frac{2-\alpha}{\theta^{3-\alpha}}\right) \\ &= \frac{\alpha}{\theta^2(1+\theta^{\alpha-1})^{3+1/(\alpha-1)}}\left((\alpha+1)\theta^{2(\alpha-1)}+(2-\alpha)\theta^{\alpha-1}\right).\end{aligned}$$

We note at this stage that $\mathbb{E}[W_{\alpha-1}^2]$ is infinite whenever $\alpha\in(1,2)$, and hence every higher moment $\mathbb{E}[W_{\alpha-1}^k]$ is also infinite for $k\geq 2$. We can, however, study the asymptotics of $\mathbb{E}[W_{\alpha-1}^k e^{-\theta W_{\alpha-1}}]$ as $\theta\to 0$. Indeed, from (62) we see that these asymptotics concentrate on the partition minimising the power $(1-\alpha)\#\pi$, that is, the partition of $\{1,\dots,k\}$ into k singletons. In particular, one can show that

$$\mathbb{E}\left[W_{\alpha-1}^k e^{-\theta W_{\alpha-1}}\right] \sim \alpha(k-1-\alpha)\cdots(2-\alpha)\theta^{\alpha-k} = \alpha\frac{\Gamma(k-\alpha)}{\Gamma(2-\alpha)}\theta^{\alpha-k} \quad \text{as } \theta\to 0. \tag{64}$$

4.2. *Critical Galton–Watson processes in the regularly varying regime.* In the sequel we will require an understanding of the factorial moments of our Galton–Watson processes at large times. To begin computing these in this section, we start by noting from (59) to obtain the second line below and (60) to obtain the third, for $\rho \in [0, 1)$ we have

$$
\begin{aligned}
&1-\mathbb{E}\big[e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}\big]\\
&\quad=\overline{F}\big(T(1-\rho)\big)\mathbb{E}\big[1-e^{-\theta\frac{\overline{F}(T)}{\overline{F}(T(1-\rho))}\overline{F}(T(1-\rho))Z_{T(1-\rho)}}|Z_{T(1-\rho)}>0\big]\\
&\quad\sim(1-\rho)^{-\frac{1}{\alpha-1}}\overline{F}(T)\mathbb{E}\big[1-e^{-\theta(1-\rho)^{\frac{1}{\alpha-1}}\overline{F}(T(1-\rho))Z_{T(1-\rho)}}|Z_{T(1-\rho)}>0\big]\\
&\quad\sim\frac{\overline{F}(T)}{(1-\rho+\theta^{-(\alpha-1)})^{\frac{1}{\alpha-1}}}
\end{aligned}
\tag{65}
$$

as $T\to\infty$. We remark that in each case above, the convergence in question is a consequence of the monotone convergence theorem.

Recall that $n^{(k)}=n(n-1)\cdots(n-k+1)$ for $n\ge k$. Then for $k\ge 1$ and $\rho\in[0,1)$, as a consequence of (60), we have

$$
\mathbb{E}\big[Z_{T(1-\rho)}^{(k)}e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}\big]\sim\overline{F}\big(T(1-\rho)\big)^{-(k-1)}\mathbb{E}\big[W_{\alpha-1}^{k}e^{-\theta(1-\rho)^{1/(\alpha-1)}W_{\alpha-1}}\big],
\tag{66}
$$

as T goes to infinity.

When $\alpha=2$, the previous asymptotic is much simpler to write. Indeed, for T large enough, we have

$$
\mathbb{E}\big[Z_{T(1-\rho)}^{(k)}e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}\big]\sim r^{k-1}(1-\rho)^{k-1}T^{k-1}\ell_2^{k-1}(T)\frac{k!}{(1+\theta(1-\rho))^{k+1}},
$$

and

$$
\mathbb{E}\big[Z_{(1-p)T}^{(k)}e^{-\theta\overline{F}(T)Z_{(1-p)T}}|Z_{(1-p)T}\ge k\big]\sim r^{k}(1-\rho)^{k}T^{k}\ell_2^{k}(T)\frac{k!}{(1+\theta(1-p))^{k+1}}.
$$

Recalling the definition (16) of $\mathbb{Q}_{\theta,T}^{(k)}$ we study the asymptotics of Z_T under the rescaling $\theta\to\theta\overline{F}(T)$. From the asymptotic in (66) with $\rho=0$, we have

$$
\mathbb{Q}_{\theta\overline{F}(T),T}^{(k)}\big[e^{-\varphi\overline{F}(T)Z_T}\big]=\frac{\mathbb{E}[Z_T^{(k)}e^{-(\theta+\varphi)\overline{F}(T)Z_T}]}{\mathbb{E}[Z_T^{(k)}e^{-\theta\overline{F}(T)Z_T}]}\xrightarrow[T\to\infty]{}\frac{\mathbb{E}[W_{\alpha-1}^{k}e^{-(\theta+\varphi)W_{\alpha-1}}]}{\mathbb{E}[W_{\alpha-1}^{k}e^{-\theta W_{\alpha-1}}]}.
\tag{67}
$$

In particular, the population off a single spine satisfies

$$
\begin{aligned}
\mathbb{Q}_{\theta\overline{F}(T),T}^{(1)}\big[e^{-\varphi\overline{F}(T)Z_T}\big]&\xrightarrow[T\to\infty]{}\frac{\mathbb{E}[W_{\alpha-1}e^{-(\theta+\varphi)W_{\alpha-1}}]}{\mathbb{E}[W_{\alpha-1}e^{-\theta W_{\alpha-1}}]}\\
&=\left(\frac{1+\theta^{\alpha-1}}{1+(\theta+\varphi)^{\alpha-1}}\right)^{1+1/(\alpha-1)}.
\end{aligned}
$$

Moreover, subpopulations along spine branch of length $(1-\rho)T$ with $\rho\in[0,1)$ satisfy

$$
\mathbb{Q}_{\theta F(T),T(1-\rho)}^{(1)}\big[e^{-\varphi\overline{F}(T)Z_{(1-\rho)T}}\big]\xrightarrow[T\to\infty]{}\left(\frac{1+\theta^{\alpha-1}(1-\rho)}{1+(\theta+\varphi)^{\alpha-1}(1-\rho)}\right)^{1+1/(\alpha-1)}.
$$

When $\alpha=2$, we have an explicit expression, for $k\ge1$, that is

$$
\mathbb{Q}_{\theta\overline{F}(T),T(1-\rho)}^{(k)}\big[e^{\mathrm{red}\varphi\overline{F}(T)Z_T}\big]\xrightarrow[T\to\infty]{}\frac{\mathbb{E}[W_1^{k}e^{-(\theta+\varphi)W_1}]}{\mathbb{E}[W_1^{k}e^{-\theta W_1}]}=\left(\frac{1+\theta}{1+\theta+\varphi}\right)^{k+1},
$$

where the right-hand side of the previous asymptotic is nothing but the Laplace exponent of a Gamma random variable with parameters $k+1$ and $1+\theta$, here denoted by $\Gamma_{k+1,1+\theta}$. In other words, when $\alpha=2$, $\overline{F}(T)Z_T$, under $\mathbb{Q}_{\theta\overline{F}(T),T}^{(k)}$, tends to $\Gamma_{k+1,1+\theta}$, under a new probability measure that we denote as $\mathbb{Q}_{\theta,\infty}^{(k)}$. That is to say, the sum of $k+1$ independent $\mathbf{e}_{1+\theta}$, exponential random variables with parameter $1+\theta$. In particular, the mass coming off a single spine branch is distributed as $\Gamma_{2,1+\theta}$.

4.3. *Large offspring.*

LEMMA 4.1. *If L is an offspring variable satisfying* **(H1)**, *then as $\lambda \to 0$, we have*

$$\mathbb{E}[Le^{-\lambda L}] \sim 1. \tag{68}$$

More generally, for either $g = 2$, $\alpha = 2$, or for any $g \geq 2$ and $\alpha \in (1, 2)$, we have

$$\mathbb{E}[L^{(g)} e^{-\lambda L}] \sim \frac{\Gamma(g - \alpha)}{\Gamma(-\alpha)} \lambda^{\alpha - g} \ell\left(\frac{1}{\lambda}\right), \tag{69}$$

as $\lambda \to 0$.

PROOF. The equation (68) is a consequence of the fact that $\mathbb{E}[L] = 1$ and the monotone convergence theorem.

As for (69), the assumption **(H1)** states that $f(s) = \mathbb{E}[s^L]$ takes the form $f(s) = s + (1 - s)^\alpha \ell(\frac{1}{1-s})$ as $s \to 1$, where $\ell(\cdot)$ is slowly varying at ∞. We note that every derivative of f is monotone decreasing, and accordingly, by the monotone density theorem, whenever either $g = 2$, $\alpha = 2$ or $g \geq 2$, $\alpha \in (1, 2)$, we have

$$f^{(g)}(s) := \mathbb{E}[L^{(g)} s^{L-g}] \sim (-1)^g \alpha(\alpha - 1) \cdots (\alpha - g + 1)(1 - s)^{\alpha - g} \ell\left(\frac{1}{1 - s}\right),$$

as $s \to 1$. Now, set $s = e^{-\lambda}$, and use the definition of the Gamma function. □

The previous results imply the following useful Lemma.

LEMMA 4.2. *For any $\alpha \in (1, 2]$, as $T \to \infty$, we have*

$$\mathbb{E}[L(\mathbb{E}[e^{-\theta \overline{F}(T) Z_{T(1-\rho)}}])^{L-1}] \sim 1, \tag{70}$$

and for either $g = 2$, $\alpha = 2$, or for any $g \geq 2$ and $\alpha \in (1, 2)$, we have

$$\mathbb{E}[L^{(g)}(\mathbb{E}[e^{-\theta \overline{F}(T) Z_{T(1-\rho)}}])^{L-g}] \sim \frac{\Gamma(g - \alpha)}{\Gamma(-\alpha)} \left(\frac{\overline{F}(T)}{(1 - \rho + \theta^{-(\alpha-1)})^{\frac{1}{\alpha-1}}}\right)^{\alpha - g} \ell_2(T)^{(\alpha-1)}, \tag{71}$$

as $T \to \infty$.

PROOF. Recall that by (65) $1 - \mathbb{E}[e^{-\theta \overline{F}(T) Z_{T(1-\rho)}}] \sim \lambda_T$, where

$$\lambda_T = \frac{\overline{F}(T)}{(1 - \rho + \theta^{-(\alpha-1)})^{\frac{1}{\alpha-1}}} \to 0$$

as $T \to \infty$. The first part (70) now follows from (68).

As for the latter equation, (71), we note first that, by (69), we have

$$\mathbb{E}[L^{(g)}(\mathbb{E}[e^{-\theta \overline{F}(T) Z_{T(1-\rho)}}])^{L-1}] \sim \frac{\Gamma(g - \alpha)}{\Gamma(-\alpha)} \lambda_T^{\alpha - g} \ell\left(\frac{1}{\lambda_T}\right). \tag{72}$$

Now, note that, using the definition of λ_T to obtain the first equality below, the asymptotics (59) for $\overline{F}(T)$ to obtain the second, and (58) to obtain the third, we have

$$\ell(1/\lambda_T) \sim \ell(1/\overline{F}(T)) \sim \ell(T^{\frac{1}{\alpha-1}}) = \ell_2(T)^{\alpha-1};$$

using this fact in (72), we obtain the result. □

The next result tells us that the asymptotic law of the offspring size at a size g splitting event under $\mathbb{Q}^{(k)}_{\theta \overline{F}(T), T}$.

LEMMA 4.3. *Let $\alpha \in (1,2)$ and $g \geq 2$. Condition on the event that some spines are following a particle just before time ρT, at this point in time this particle dies, and after this event the spines following this particle split into g groups. Let J denote the number of offspring of the dying particle at this splitting event. Then the rescaled conditional Laplace transform of J is given by*

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)J}\big|\mathcal{G}^{(k)}_T, \textit{split of size } g \textit{ at time } \rho T\big] \sim \big(1+\varphi\big(1-\rho+\theta^{1-\alpha}\big)^{\frac{1}{\alpha-1}}\big)^{-(g-\alpha)}.$$

PROOF. According to Remark 2, given that a splitting event of size g occurs at time t, under $\mathbb{Q}^{(k)}_{\theta,T}$, the conditional distribution of the size of offspring event is given by $\ell^{(g)}\mathbb{E}[e^{-\theta Z_{T-t}}]^{\ell-g}p_\ell/\mathbb{E}[L^{(g)}\mathbb{E}[e^{-\theta Z_{T-t}}]^{L-g}]$.

In particular, setting $t=\rho T$, we have

$$\begin{aligned}\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}&\big[e^{-\varphi\overline{F}(T)J}\big|\mathcal{G}^{(k)}_T, \text{split of size } g \text{ at time } \rho T\big]\\ &\sim \frac{\mathbb{E}[L^{(g)}\mathbb{E}[e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}]^{L-g}e^{-\varphi\overline{F}(T)L}]}{\mathbb{E}[L^{(g)}\mathbb{E}[e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}]^{L-g}]}.\end{aligned}$$

Now, with λ_T as in the proof of Lemma 4.2, we have $\mathbb{E}[e^{-\theta\overline{F}(T)Z_{T(1-\rho)}}]=e^{-(1+o(1))\lambda_T}$ so that

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)J}\big|\mathcal{G}^{(k)}_T, \text{split of size } g \text{ at time } \rho T\big] \sim \frac{\mathbb{E}[L^{(g)}e^{-(\lambda_T L+\varphi\overline{F}(T))L}]}{\mathbb{E}[L^{(g)}e^{-\lambda_T L}]}.$$

Now, using (69), we have

$$\begin{aligned}\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)J}\big|\mathcal{G}^{(k)}_T, \text{split of size } g \text{ at time } \rho T\big] &\sim \left(\frac{\lambda_T}{\lambda_T+\varphi\overline{F}(T)}\right)^{g-\alpha}\\ &\sim \big(1+\varphi\big(1-\rho+\theta^{1-\alpha}\big)^{\frac{1}{\alpha-1}}\big)^{-(g-\alpha)},\end{aligned}$$

where the final display follows from using the definition of λ_T in the proof of the previous lemma. That completes the proof. □

We note that the previous lemma states that $\overline{F}(T)J$ is asymptotically distributed like a Gamma random variable with shape parameter $g-\alpha$ and scale parameter $(1-\rho+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}$. More explicitly, we have the following immediate corollary.

COROLLARY 4.4. *Let $\alpha \in (1,2)$ and $g \geq 2$. Let J_j denote the number of offspring at a splitting event of size g_j at time $\rho_j T$. Then*

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big(\overline{F}_T J_j \in \mathrm{d}x\big|\mathcal{G}^{(k)}_T, \textit{split of size } g_j \textit{ at time } \rho_j T\big) = \Delta^\theta_{g_j,\rho_j}(x)\,\mathrm{d}x,$$

where

$$\Delta^\theta_{g,\rho}(x) := \Delta^\theta_{g,\rho,\alpha,\theta}(x) := \frac{x^{g-\alpha-1}}{\Gamma(g-\alpha)(1-\rho+\theta^{1-\alpha})^{\frac{g-\alpha}{\alpha-1}}}\exp\left\{-\frac{x}{(1-\rho+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}}\right\} \tag{73}$$

is the density function of $(1-\rho+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}$ times a Gamma random variable with parameter $g-\alpha$.

Our final result tells us about the asymptotic distribution of the number of children descended from nonspine particles born at a splitting event. Recall from Section 3.3 that $\tilde{Z}_{T,j}$ counts the number of particles who are descended from nonspine particles born at the jth splitting event τ_j.

LEMMA 4.5. *Let $\alpha \in (1,2]$. Then under $\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}$, the asymptotic distribution of $\tilde{Z}_{T,j}$, conditional on $\overline{F}(T)J_j = x_j$, is given by*

$$\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}[e^{-\varphi \overline{F}(T)\tilde{Z}_{T,j}} | \mathcal{G}^{(k)}_T, \text{split of size } g \text{ at time } \rho T, \overline{F}(T)J_j = x_j]$$
$$\sim \exp\left\{-\left(\frac{1}{(1-\rho_j+(\varphi+\theta)^{1-\alpha})^{\frac{1}{\alpha-1}}} - \frac{1}{(1-\rho_j+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}}\right)x_j\right\},$$

regardless of the size g of the split.

PROOF. Recall that particles not carrying spines behave at time t as if under the original measure $\mathbb{P}$ but with discounting by $e^{-\theta \overline{F}(T)Z_{T-t}}$ under $\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}$. In other words, if Y_T denotes the number of descendants at time T of a nonspine particle u living at time t, we have

$$\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}[e^{-\varphi \overline{F}(T)Y_T} | \mathcal{F}^{(k)}_t] = F_T(e^{-(\theta+\varphi)\overline{F}(T)})/F_T(e^{-\theta \overline{F}(T)}).$$

Moreover, Y_T is independent of the remainder of the population. In particular, since $\frac{x_j}{\overline{F}(T)} - g$ nonspine particles are born at time τ_j, we have

$$\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}[e^{-\varphi \overline{F}(T)\tilde{Z}_{T,j}} | \mathcal{G}^{(k)}_T, \text{split of size } g \text{ at time } \rho T, \overline{F}(T)J_j = x_j]$$
$$= \left(\frac{F_T(e^{-(\theta+\varphi)\overline{F}(T)})}{F_T(e^{-\theta \overline{F}(T)})}\right)^{\frac{x_j}{\overline{F}(T)} - g}.$$

The result now follows from (65). □

We remark that in Lemma 4.5, the linear dependence of the Laplace transform on x is due to a branching property.

5. Inverting the change of measure.

5.1. *The ancestral tree probability under $\mathbb{Q}^{(k),T}_{\theta \overline{F}(T)}$.*

LEMMA 5.1. *Let $(\beta_0, \dots, \beta_m)$ be a splitting process of $\{1, \dots, k\}$. Let $0 < t_1 < \dots < t_m < 1$. Then*

$$\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}(\tau_1/T \in \mathrm{d}t_1, \dots, \tau_m/T \in \mathrm{d}t_m, \mathcal{T}(\xi) = (\beta_0, \dots, \beta_m))$$
$$= \frac{\prod_{i=1}^m \mathrm{d}t_i}{\mathbb{E}[W^k_{\alpha-1} e^{-\theta W_{\alpha-1}}]} \prod_{i=1}^m \frac{\alpha \Gamma(g_i - \alpha)}{\Gamma(2-\alpha)} (1 - t_i + \theta^{1-\alpha})^{\frac{g_i - \alpha}{\alpha-1}} \tag{74}$$
$$\times \prod_{i=0}^m \theta^{-\alpha(g_i-1)} (\theta^{1-\alpha} + 1 - t_i)^{-\frac{\alpha}{\alpha-1}(g_i-1)}.$$

PROOF. According to (48), we have

$$\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}(\tau_1/T \in \mathrm{d}t_1, \dots, \tau_m/T \in \mathrm{d}Tt_m, \mathcal{T}(\xi) = (\beta_0, \dots, \beta_m))$$
$$= \lim_{T\to\infty} \frac{F'_T(e^{-\theta \overline{F}(T)})}{\mathbb{E}[Z^{(k)}_T e^{-\theta \overline{F}(T) Z_T}]} \tag{75}$$
$$\times \prod_{i=1}^m \mathbb{E}[L^{(g_i)}(\mathbb{E}[e^{-\theta \overline{F}(T) Z_{T(1-t_i)}}])^{L-g_i}] F'_{T(1-t_i)}(e^{-\theta \overline{F}(T)})^{g_i - 1} rT \,\mathrm{d}t_i.$$

Now, according to (66) and then (63), we have $F'_{T(1-\rho)}(e^{-\theta\overline{F}(T)}) \sim (1+\theta^{\alpha-1}(1-\rho))^{-\frac{\alpha}{\alpha-1}}$. Using this fact in conjunction with the asymptotics in (71) and (66), we obtain

$$
\begin{aligned}
&\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big(\tau_1/T \in \mathrm{d}t_1, \ldots, \tau_m/T \in \mathrm{d}t_m, \mathcal{T}(\xi) = (\beta_0, \ldots, \beta_m)\big) \\
&\quad = \lim_{T\to\infty} \frac{\overline{F}(T)^{(k-1)}}{\mathbb{E}[W^k_{\alpha-1} e^{-\theta W_{\alpha-1}}]} \prod_{i=1}^m (rT\,\mathrm{d}t_i) \\
&\qquad \times \prod_{i=1}^m \frac{\Gamma(g_i-\alpha)}{\Gamma(-\alpha)} \bigg(\frac{\overline{F}(T)}{(1-t_i+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}}\bigg)^{\alpha-g_i} \ell_2(T)^{(\alpha-1)} \\
&\qquad \times \prod_{i=0}^m \big(1+\theta^{\alpha-1}(1-t_i)\big)^{-\frac{\alpha}{\alpha-1}(g_i-1)},
\end{aligned}
\tag{76}
$$

where we are using the conventions $g_0 = 2$ and $t_0 = 0$.

We now gather the asymptotic terms in T. Recall from (59) that $\overline{F}(T) \sim c_{\alpha,r} T^{-\frac{1}{\alpha-1}}/\ell_2(T)$ where $c_{\alpha,r} = ((\alpha-1)r)^{-\frac{1}{\alpha-1}}$. Using this fact in conjunction with the simple identity $\sum_{i=1}^m (g_i - 1) = k-1$, we have

$$
T^m \overline{F}(T)^{k-1} \prod_{i=1}^m \overline{F}(T)^{\alpha-g_i} \ell_2(T)^{(\alpha-1)} \sim \frac{1}{((\alpha-1)r)^m} \tag{77}
$$

as $T \to \infty$. Plugging (77) into (76) and using the identity $(\alpha-1)\Gamma(-\alpha) = \Gamma(2-\alpha)/\alpha$, we obtain

$$
\begin{aligned}
&\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big(\tau_1/T \in \mathrm{d}t_1, \ldots, \tau_m/T \in \mathrm{d}t_m, \mathcal{T}(\xi) = (\beta_0, \ldots, \beta_m)\big) \\
&\quad = \frac{\prod_{i=1}^m \mathrm{d}t_i}{\mathbb{E}[W^k_{\alpha-1} e^{-\theta W_{\alpha-1}}]} \prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)} \big(1-t_i+\theta^{1-\alpha}\big)^{\frac{g_i-\alpha}{\alpha-1}} \\
&\qquad \times \prod_{i=0}^m \theta^{-\alpha(g_i-1)} \big(\theta^{1-\alpha}+1-t_i\big)^{-\frac{\alpha}{\alpha-1}(g_i-1)},
\end{aligned}
$$

completing the proof. $\square$

Combining the previous result and Corollary 4.4, and aggregating the product using the identity $\sum_{i=1}^m (g_i-1) = k-1$, we obtain the following result.

COROLLARY 5.2. *We have*

$$
\begin{aligned}
&\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\bigg(\mathcal{T}(\xi) = (\beta_0, \ldots, \beta_m), \frac{\tau_i}{T} \in \mathrm{d}t_i, \overline{F}(T)L_i \in \mathrm{d}x_i \textit{ for all } i = 1, \ldots, m\bigg) \\
&\quad = \frac{\theta^{-\alpha k}(\theta^{1-\alpha}+1)^{-\frac{\alpha}{\alpha-1}}}{\mathbb{E}[W^k_{\alpha-1} e^{-\theta W_{\alpha-1}}]} \prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)} \big(1-t_i+\theta^{1-\alpha}\big)^{-g_i} \mathrm{d}t_i \prod_{i=1}^m \Delta^\theta_{g_i,t_i}(\mathrm{d}x_i).
\end{aligned}
\tag{78}
$$

5.2. *The conditional distribution of Z_T given ancestral tree under $\mathbb{Q}^{(k),T}_{\theta\overline{F}(T)}$.* In the previous section, we computed the $\mathbb{Q}^{(k)}_{\theta,\infty}$ probabilities associated with the limiting ancestral tree. In this section we compute the conditional Laplace transform for the entire population Z_T at time T *given* a certain ancestral tree with certain offspring sizes at splitting events.

Let us begin by recalling from Section 3.3 the decomposition

$$Z_T = k + \sum_{j=1}^{k} Z_{T,j} + \sum_{j=1}^{m} \tilde{Z}_{T,j} \tag{79}$$

of the entire population at time T into conditionally independent constituents. Here $Z_{T,j}$ counts the number of individuals born off the spine of lineage j, and $\tilde{Z}_{T,j}$ counts the number of descendants of nonspine particles born at the time of the jth splitting event. Our next two lemmas characterise the conditional laws of $Z_{T,j}$ and $\tilde{Z}_{T,j}$, given the ancestral tree, split times, and offspring sizes.

Recall the definition of ς_j given in Section 3.3. Each ς_j is the initial time of a lineage and is equal to some τ_i. It transpires that asymptotically under $\mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}$, the spine subpopulations have a highly tractable form.

LEMMA 5.3. *Conditional on $\mathcal{G}_T^{(k)}$ given $\varsigma_j = \rho T$, we have*

$$\lim_{T\to\infty} \mathbb{Q}^{(k),T}_{\theta \overline{F}(T)}\big[e^{-\varphi \overline{F}(T) Z_{T,j}} \big| \mathcal{G}_T^{(k)}, \varsigma_j = \rho T\big] = \left(\frac{1+(1-\rho)\theta^{\alpha-1}}{1+(1-\rho)(\theta+\varphi)^{\alpha-1}}\right)^{\frac{\alpha}{\alpha-1}}, \tag{80}$$

as $T \to \infty$.

PROOF. Note that, according to Lemma 3.5, the rate of births occuring off the spine is independent of the number of spines following a particle. In particular, it follows that the contribution to births off the spine along a lineage $[\rho T, T]$ is identical to what it would be under the 1-spine measure run across $[0, (1-\rho)T]$. It follows that

$$\mathbb{Q}^{(k),T}_{\theta \overline{F}(T)}\big[e^{-\varphi \overline{F}(T) Z_{T,j}} \big| \mathcal{G}_T^{(k)}, \varsigma_j = \rho T\big] = \mathbb{Q}^{(1),T(1-\rho)}_{\theta \overline{F}(T)}\big[e^{-\varphi \overline{F}(T) Z_{T(1-\rho)}}\big].$$

Now, appealing to the case $k=1$ of (18), we have

$$\mathbb{Q}^{(1),T(1-\rho)}_{\theta \overline{F}(T)}\big[e^{-\varphi \overline{F}(T) Z_{T(1-\rho)}}\big] = \frac{\mathbb{E}[Z_{T(1-\rho)} e^{-(\varphi+\theta)\overline{F}(T) Z_{T(1-\rho)}}]}{\mathbb{E}[Z_{T(1-\rho)} e^{-\theta\overline{F}(T) Z_{T(1-\rho)}}]}.$$

Using (59) to replace $\overline{F}(T)$ with $\overline{F}(T(1-\rho))$ and using the monotone convergence theorem, we have

$$\lim_{T\to\infty} \mathbb{Q}^{(1),T(1-\rho)}_{\theta \overline{F}(T)}\big[e^{-\varphi \overline{F}(T) Z_{T(1-\rho)}}\big] = \lim_{T\to\infty} \frac{\mathbb{E}[Z_{T(1-\rho)} e^{-(\varphi+\theta)(1-\rho)^{\frac{1}{\alpha-1}}\overline{F}(T(1-\rho)) Z_{T(1-\rho)}}]}{\mathbb{E}[Z_{T(1-\rho)} e^{-\theta(1-\rho)^{\frac{1}{\alpha-1}}\overline{F}(T(1-\rho)) Z_{T(1-\rho)}}]}.$$

Now, use (63). □

Let us mention here that a simple calculation using (80) establishes the following corollary in the setting $\alpha = 2$:

COROLLARY 5.4. *Let $\alpha = 2$. Then the distribution of $Z_{T,j}$ converges in distribution as $T \to \infty$ to the sum of two independent exponential random variables with parameter $\frac{1}{1-\rho} + \theta$.*

We also recall now from Lemma 4.5 that

$$\begin{aligned}&\lim_{T\to\infty} \mathbb{Q}^{(k)}_{\theta \overline{F}(T),T}\big[e^{-\varphi \overline{F}(T) \tilde{Z}_{T,j}} \big| \mathcal{G}_T^{(k)}, \text{split of size } g \text{ at time } \rho_j T, \overline{F}(T) J_j = x_j\big] \\ &\quad = \exp\{-G(\rho_j, \varphi) x_j\},\end{aligned} \tag{81}$$

where

$$G(\rho_j,\varphi):=\left(\frac{1}{(1-\rho_j+(\varphi+\theta)^{1-\alpha})^{\frac{1}{\alpha-1}}}-\frac{1}{(1-\rho_j+\theta^{1-\alpha})^{\frac{1}{\alpha-1}}}\right).$$

The main result of this section is a simple consequence of what we have seen in this section so far and gives a characterisation of the asymptotic conditional law of Z_T under $\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}$ after conditioning on the spine ancestral tree and the offspring sizes at split times.

LEMMA 5.5. *For short, write*

$$\Gamma_T:=\{\tau_1/T\in \mathrm{d}t_1,\dots,\tau_m/T\in \mathrm{d}t_m,\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),\\ \overline{F}(T)L_{\tau_1}\in \mathrm{d}x_1,\dots,\overline{F}(T)L_{\tau_m}\in \mathrm{d}x_m\}.$$

Then

$$\lim_{T\to\infty}\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)Z_T}|\Gamma_T\big]\\ =e^{-\sum_{j=1}^m x_jG(t_j,\varphi)}\prod_{j=0}^m\left(\frac{1+(1-t_j)\theta^{\alpha-1}}{1+(1-t_j)(\theta+\varphi)^{\alpha-1}}\right)^{(g_j-1)\frac{\alpha}{\alpha-1}},\tag{82}$$

where we are using the convention $\rho_0=0$, $g_0=2$.

PROOF. Using the decomposition (79), we have

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)Z_T}|\Gamma_T\big]\\ =e^{-\theta\overline{F}(T)k}\prod_{j=1}^k\mathbb{Q}^{(k),T}_{\theta\overline{F}(T)}\big[e^{-\varphi\overline{F}(T)Z_{T,j}}|\Gamma_T\big]\prod_{j=1}^m\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)\tilde{Z}_{T,j}}|\Gamma_T\big].$$

Recall that each ς_j is the initial time of a lineage and is equal to some τ_i. Moreover, for each $j=0,\dots,m$, there are exactly g_j-1 different i such that $\varsigma_i=\tau_j$. It follows, using (80), that

$$\lim_{T\to\infty}\prod_{j=1}^k\mathbb{Q}^{(k),T}_{\theta\overline{F}(T)}\big[e^{-\varphi\overline{F}(T)Z_{T,j}}|\Gamma_T\big]=\prod_{j=0}^m\left(\frac{1+(1-t_j)\theta^{\alpha-1}}{1+(1-t_j)(\theta+\varphi)^{\alpha-1}}\right)^{(g_j-1)\frac{\alpha}{\alpha-1}}.$$

As for the other term, by (81) we have immediately

$$\lim_{T\to\infty}\prod_{j=1}^m\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-\varphi\overline{F}(T)\tilde{Z}_{T,j}}|\Gamma_T\big]=\exp\left(-\sum_{j=1}^m x_jG(t_j,\varphi)\right).$$

Combining the two equations completes the proof. □

5.3. *The joint law of the ancestral tree, split offspring sizes, and entire population.* With Corollary 5.2 and Lemma 5.5 at hand, we now prove the following.

LEMMA 5.6. *With Γ_T as in Lemma 5.5, we have*

$$\lim_{T\to\infty}\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-(\varphi-\theta)\overline{F}(T)Z_T};\Gamma_T\big]=\frac{1}{\mathbb{E}[W^k_{\alpha-1}e^{-\theta W_{\alpha-1}}]}\varphi^{-\alpha k}(1+\varphi^{1-\alpha})^{-\frac{\alpha}{\alpha-1}}\\ \times\prod_{i=1}^m\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}(1-t_i+\varphi^{1-\alpha})^{-g_i}\Delta^{\varphi}_{g_i,t_i}(x_i).$$

PROOF. Simply multiplying the main equations of Corollary 5.2 and Lemma 5.5 (and carefully separating the $j=0$ term out in the product in (82)), we have immediately

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-(\varphi-\theta)\overline{F}(T)Z_T};\Gamma_T\big]=\frac{\theta^{-\alpha k}(\theta^{1-\alpha}+1)^{-\frac{\alpha}{\alpha-1}}}{\mathbb{E}[W_{\alpha-1}e^{-\theta W_{\alpha-1}}]}\left(\frac{1+\theta^{\alpha-1}}{1+\varphi^{\alpha-1}}\right)^{\frac{\alpha}{\alpha-1}}\prod_{i=1}^{m}R_i\,\mathrm{d}x_i\,\mathrm{d}t_i, \tag{83}$$

where

$$R_i=\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}(1-t_i+\theta^{1-\alpha})^{-g_i}\left(\frac{1+(1-t_i)\theta^{\alpha-1}}{1+(1-t_i)\varphi^{\alpha-1}}\right)^{(g_i-1)\frac{\alpha}{\alpha-1}}$$
$$\times\,\Delta^{\theta}_{g_i,t_i}(x_i)\exp\big(-x_iG(t_i,\varphi)\big).$$

Using the definition (73) of $\Delta^{\theta}_{g,\rho}(x)$, a calculation verifies that

$$R_i=\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}(\theta/\varphi)^{(g_i-1)\alpha}\big(1-t_i+\varphi^{1-\alpha}\big)^{-g_i}\Delta^{\varphi}_{g_i,t_i}(x_i), \tag{84}$$

where we note the latter expression above involves $\Delta^{\varphi}_{g_i,t_i}$ as opposed to $\Delta^{\theta}_{g_i,t_i}$.

Plugging (83) into (84) and making good use of the identity $\sum_{i=1}^{m}(g_i-1)=k-1$, we obtain the result. □

The reader will note from Lemma 5.6, we have

$$\mathbb{Q}^{(k)}_{\theta\overline{F}(T),T}\big[e^{-(\varphi-\theta)\overline{F}(T)Z_T};\Gamma_T\big]=\mathbb{Q}^{(k)}_{\varphi\overline{F}(T),T}[\Gamma_T]\frac{\mathbb{E}[W_{\alpha-1}e^{-\varphi W_{\alpha-1}}]}{\mathbb{E}[W_{\alpha-1}e^{-\theta W_{\alpha-1}}]}.$$

This identity may alternatively be derived from considering the respective changes of measure.

5.4. *Tree probabilities for uniform choice under $\mathbb{P}^{(k)}_{\mathrm{unif},T}$.* We are almost ready to prove the main results, Theorems 1.1, Theorem 1.2, and Theorem 1.3. We begin with the following formulation. Recall that L_i denotes the number of particles born at time τ_i.

THEOREM 5.7. *We have*

$$\begin{aligned}&\lim_{T\to\infty}\mathbb{P}^{(k)}_{\mathrm{unif},T}\big(\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),\tau_i/T\in\mathrm{d}t_i,\overline{F}(T)L_i\in\mathrm{d}x_i,i=1,\dots,m\,|\,Z_T\geq k\big)\\&\quad=\int_0^\infty\frac{\varphi^{-(\alpha-1)k-1}}{(k-1)!}\big(1+\varphi^{1-\alpha}\big)^{-\frac{\alpha}{\alpha-1}}\prod_{i=1}^{m}\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}\big(1-t_i+\varphi^{1-\alpha}\big)^{-g_i}\Delta^{\varphi}_{g_i,t_i}(x_i)\,\mathrm{d}\varphi,\end{aligned} \tag{85}$$

where, *recalling* (73), *we have*

$$\Delta^{\varphi}_{g,\rho}(x):=\frac{x^{g-\alpha-1}}{\Gamma(g-\alpha)(1-\rho+\varphi^{1-\alpha})^{\frac{g-\alpha}{\alpha-1}}}\exp\left\{-\frac{x}{(1-\rho+\varphi^{1-\alpha})^{\frac{1}{\alpha-1}}}\right\}. \tag{86}$$

PROOF. Taking an asymptotic version of (23) (simply replacing $Z^{(k)}_T$ with Z^k_T) with $\theta\overline{F}(T)$ in place of θ, we have

$$\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\geq k\big]\sim\mathbb{E}\big[Z^k_Te^{-\theta\overline{F}(T)Z_T}|Z_T\geq k\big]\mathbb{Q}^{(k)}_{\theta,T}\left[\frac{f(\xi_T)}{Z^k_Te^{-\theta\overline{F}(T)Z_T}}\right]. \tag{87}$$

Now, according to the Gamma integral, we have $\frac{1}{z^k}=\frac{1}{(k-1)!}\int_0^\infty\varphi^{k-1}e^{-\varphi z}\mathrm{d}\varphi$ so that using Fubini's theorem we may instead write

$$\begin{aligned}&\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\geq k\big]\\&\quad\sim\frac{1}{(k-1)!}\mathbb{E}\big[Z^k_Te^{-\theta\overline{F}(T)Z_T}|Z_T\geq k\big]\int_0^\infty\varphi^{k-1}\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)e^{Z_T}e^{-(\varphi-\theta)\overline{F}(T)Z_T}\big]\mathrm{d}\varphi.\end{aligned} \tag{88}$$

Changing variable from φ to $\varphi\overline{F}(T)$, we obtain

$$\begin{aligned}(89)\quad &\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\ge k\big]\\ &\sim \frac{\overline{F}(T)^k}{(k-1)!}\mathbb{E}\big[Z_T^k e^{-\theta\overline{F}(T)Z_T}|Z_T\ge k\big]\int_0^\infty \varphi^{k-1}\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)e^{-(\varphi-\theta)\overline{F}(T)Z_T}\big]\,\mathrm{d}\varphi.\end{aligned}$$

Now, using the case $\rho=0$ in (66), this reduces to

$$\begin{aligned}(90)\quad &\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\ge k\big]\\ &\sim \frac{\mathbb{E}[W^k_{\alpha-1}e^{-\theta W_{\alpha-1}}]}{(k-1)!}\int_0^\infty \varphi^{k-1}\mathbb{Q}^{(k)}_{\theta,T}\big[f(\xi_T)e^{-(\varphi-\theta)\overline{F}(T)Z_T}\big]\,\mathrm{d}\varphi.\end{aligned}$$

Now, we consider the case where $f(\xi_T)$ is the indicator function of the event $\{\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),\tau_i/T\in\mathrm{d}t_i,\overline{F}(T)L_i\in\mathrm{d}x_i, i=1,\dots,m\}$. Using Lemma 5.6 and the bounded convergence theorem, we obtain

$$\begin{aligned}(91)\quad &\lim_{T\to\infty}\mathbb{E}^{(k)}_{\mathrm{unif},T}\big[f(\xi_T)|Z_T\ge k\big]\\ &=\int_0^\infty \frac{\varphi^{k-1}\varphi^{-\alpha k}}{(k-1)!}\big(1+\varphi^{1-\alpha}\big)^{-\frac{\alpha}{\alpha-1}}\prod_{i=1}^m\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}\big(1-t_i+\varphi^{1-\alpha}\big)^{-g_i}\Delta^{\varphi}_{g_i,t_i}(x_i)\,\mathrm{d}\varphi,\end{aligned}$$

thereby completing the proof. □

Consider now taking the change of variable $w=(1+\varphi^{1-\alpha})^{-1}$. The following equations are easily verified:

$$(92)\qquad \varphi^{-(\alpha-1)}=\frac{1-w}{w},$$

$$(93)\qquad \frac{\mathrm{d}\varphi}{\varphi}=\frac{\mathrm{d}w}{(\alpha-1)w(1-w)}.$$

Using (92) and (93) in (85), we obtain

$$\begin{aligned}&\lim_{T\to\infty}\mathbb{P}^{(k)}_{\mathrm{unif},T}\big(\{\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),\tau_i/T\in\mathrm{d}t_i,\overline{F}(T)L_i\in\mathrm{d}x_i,i=1,\dots,m\}|Z_T\ge k\big)\\ (94)\quad &=\frac{1}{(\alpha-1)(k-1)!}\\ &\quad\times\int_0^1\frac{\mathrm{d}w}{w(1-w)}\Big(\frac{1-w}{w}\Big)^k w^{\frac{\alpha}{\alpha-1}}\prod_{i=1}^m\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}\Big(\frac{1}{w}-t_i\Big)^{-g_i}\Delta^{w}_{g_i,t_i}(x_i),\end{aligned}$$

where we are abusing notation slightly and writing

$$(95)\qquad \Delta^w_{g,\rho}(x):=\frac{x^{g-\alpha-1}}{\Gamma(g-\alpha)(1/w-\rho)^{\frac{g-\alpha}{\alpha-1}}}\exp\Big\{-\frac{x}{(1/w-\rho)^{\frac{1}{\alpha-1}}}\Big\}$$

for the probability density function of $(1/w-\rho)^{\frac{1}{\alpha-1}}$ times a Gamma random variable with shape $g-\alpha$. Tidying (94) and using the identity $\sum_{i=1}^m(g_i-1)=k-1$, we ultimately obtain

$$\begin{aligned}&\lim_{T\to\infty}\mathbb{P}^{(k)}_{\mathrm{unif},T}\big(\{\mathcal{T}(\xi)=(\beta_0,\dots,\beta_m),\tau_i/T\in\mathrm{d}t_i,\overline{F}(T)L_i\in\mathrm{d}x_i,i=1,\dots,m\}|Z_T\ge k\big)\\ (96)\quad &=\frac{1}{(\alpha-1)(k-1)!}\\ &\quad\times\int_0^1(1-w)^{k-1}w^{m+\frac{2-\alpha}{\alpha-1}}\prod_{i=1}^m\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}(1-wt_i)^{-g_i}\Delta^w_{g_i,t_i}(x_i)\,\mathrm{d}w.\end{aligned}$$

Let us now recapitulate and prove Theorem 1.1, Theorem 1.2, and Theorem 1.3 explicitly.

PROOF OF THEOREMS 1.1, 1.2, AND 1.3. Consider a continuous-time Galton–Watson tree with offspring distribution in the universality class (**H1**). Under a probability measure $\mathbb{P}^{(k)}_{\mathrm{unif},T}(\cdot|Z_T \geq k)$, condition on the event $\{Z_T \geq k\}$ that there are at least k particles alive at time T, and sample k particles uniformly from the population at time T. Let $(\pi_t^{(k,T)})_{t\in[0,T]}$ denote the joint ancestral process of these particles. Let $\mathcal{T}(\pi^{(k,T)})$ denote the splitting process associated with this ancestry, and let $\tau_1, \dots, \tau_m$ denote the split times. Finally, let $L_1, \dots, L_m$ denote the offspring sizes at these split times.

Asymptotically in T, the equation (96) holds, which, in particular, does not depend on the explicit form of the offspring distribution but only the parameter α governing the universality class (**H1**). As such, that proves Theorem 1.1.

Next, we note that the convergence in distribution for such trees in this universality class proves Theorem 1.3.

Finally, we note that Theorem 1.2 is obtained from Theorem 1.3 by integrating against x_i. $\square$

5.5. *The Lauricella representation.* In this section we derive the Lauricella representation for the joint density of the split times. According to Theorem 1.2, we have

$$
\begin{aligned}
&\mathbb{P}(\mathcal{T}(\nu) = \overline{\beta}, \tau_1 \in \mathrm{d}t_1, \dots, \tau_m \in \mathrm{d}t_m) \\
&\quad = \frac{1}{(\alpha-1)(k-1)!} \prod_{i=1}^m \frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)} \int_0^1 (1-w)^{k-1} w^{m+\frac{2-\alpha}{\alpha-1}} \prod_{i=1}^m (1-wt_i)^{-g_i}\,\mathrm{d}w.
\end{aligned}
\tag{97}
$$

We now derive our alternative representation for the right-hand side. Let us begin by noting that

$$
\begin{aligned}
&\int_0^1 w^{m+\frac{2-\alpha}{\alpha-1}} (1-w)^{k-1} \prod_{i=1}^m (1-wt_i)^{-g_i}\,\mathrm{d}w \\
&\quad = \int_0^1 \mathrm{d}w\, w^{m+\frac{2-\alpha}{\alpha-1}} (1-w)^{k-1} \sum_{j_i,\dots,j_m=0}^m \frac{(g_1)_{j_1}\cdots(g_m)_{j_m}}{j_1!\cdots j_m!} (wt_1)^{j_1}\cdots(wt_m)^{j_m} \\
&\quad = \sum_{j_i,\dots,j_m=0}^m \frac{(g_1)_{j_1}\cdots(g_m)_{j_m}}{j_1!\cdots j_m!} t_1^{j_1}\cdots t_m^{j_m} \int_0^1 w^{m+\frac{1}{\alpha-1}+\sum_{i=1}^m j_i - 1}(1-w)^{k-1}\,\mathrm{d}w \\
&\quad = \frac{\Gamma(m+\frac{1}{\alpha-1})\Gamma(k)}{\Gamma(k+m+\frac{1}{\alpha-1})} \sum_{j_i,\dots,j_m=0}^m \frac{(m+\frac{1}{\alpha-1})_{j_1+\cdots+j_m}(g_1)_{j_1}\cdots(g_m)_{j_m}}{(k+m+\frac{1}{\alpha-1})_{j_1+\cdots+j_m} j_1!\cdots j_m!} t_1^{j_1}\cdots t_m^{j_m} \\
&\quad = \frac{\Gamma(m+\frac{1}{\alpha-1})\Gamma(k)}{\Gamma(k+m+\frac{1}{\alpha-1})} F_D^{(m)}\left[m+\frac{1}{\alpha-1}, g_1, \dots, g_m; k+m+\frac{1}{\alpha-1}; t_1, \dots, t_m\right],
\end{aligned}
\tag{98}
$$

where $F_D^{(m)}$ is the Lauricella hypergeometric function in m variables $t_1, \dots, t_m$, which was introduced by Lauricella [29]. We note now that by plugging (98) into (97) we obtain (8), as stated in the Introduction.

We now appeal to a probabilistic representation of Chamayou and Wesolowski [12]. To set this up, let $(X_1, \dots, X_n)$ be a random vector with Dirichlet distribution with parameters

$a=(a_1,\dots,a_n)$ and $b>0$, here denoted by $\mathtt{Dir}(a;b)$. In other words, its distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ and

$$D_n(a;b;x)=C\left(1-\sum_{i=1}^n x_i\right)^{b-1}\prod_{i=1}^n x_i^{a_i-1}\mathbf{1}_{\mathbb{T}_n}(x),$$

where $x=(x_1,\dots,x_n)\in\mathbb{R}^n$,

$$\mathbb{T}_n=\left\{(x_1,\dots,x_n):x_i>0,i=1,\dots,n,\sum_{i=1}^n x_i<1\right\}\quad\text{and}\quad C=\frac{\Gamma(b+\sum_{i=1}^n a_i)}{\Gamma(b)\Gamma(\sum_{i=1}^n a_i)}.$$

An important property of the Dirichlet distribution is that it can be represented through independent gamma distributions; that is, let $U_1,\dots,U_n$ be independent Gamma r.v.'s with parameters (σ,a_i), for $i=1,\dots,n$, then

$$(X_1,\dots,X_n)\stackrel{d}{=}\frac{(U_1,\dots,U_n)}{\sum_{i=1}^n U_i},$$

where the latter is independent of $\sum_{i=1}^n U_i$.

Moreover, the Laplace exponent of the Dirichlet distribution satisfies

$$\mathbb{E}\big[e^{\langle t,X\rangle}\big]=C\int_{\mathbb{T}_n}e^{\langle t,x\rangle}\left(1-\sum_{i=1}^n x_i\right)^{b-1}\prod_{i=1}^n x_i^{a_i-1}\,\mathrm{d}x_1\cdots\mathrm{d}x_n=\Phi_2^{(n)}(a;b;t),$$

where $t=(t_1,\dots,t_n)$. The function $\Phi_2^{(n)}$ can be viewed as multivariate version of the hypergeometric function ${}_1F_1$, that is,

$$\Phi_2^{(1)}(a,b,t)={}_1F_1(a;a+b;t),$$

where the right-hand side is the Laplace transform of a beta r.v. with parameters (a,b).

Next, let Z be a Gamma r.v. with parameters $(1,c)$. If $X\sim\mathtt{Dir}(a;b)$ and X and Z are independent, then the random vector $Y=ZX$ satisfies

$$\mathbb{E}\big[\exp\{\langle t,Y\rangle\}\big]=F_D^{(n)}(c,a;b;t).$$

Thus, conditioning with respect to Z, we have

$$F_D^{(n)}(c,a;b;t)=\frac{1}{\Gamma(c)}\int_0^\infty z^{c-1}e^{-z}\Phi_2^{(n)}(a;b;tz)\,\mathrm{d}z.$$

Conditioning with respect to X, we have

$$F_D^{(n)}(c,a;b;t)=\int_{\mathbb{T}_n}\big(1-\langle t,x\rangle\big)^{-c}D_n(a;b;x)\,\mathrm{d}x. \tag{99}$$

Using (99) in (98), we obtain

$$\begin{aligned}&\int_0^1 w^{m+\frac{2-\alpha}{\alpha-1}}(1-w)^{k-1}\prod_{i=1}^m(1-wt_i)^{-g_i}\,\mathrm{d}w\\&\quad=\frac{\Gamma(m+\frac{1}{\alpha-1})\Gamma(k)}{\Gamma(k+m+\frac{1}{\alpha-1})}\int_{\mathbb{T}_m}\big(1-\langle t,x\rangle\big)^{-m-\frac{1}{\alpha-1}}D_m\left(g_1,\dots,g_m,k+m+\frac{1}{\alpha-1};\mathrm{d}x\right).\end{aligned}$$

We finally note that if $(E_1,\dots,E_m)$ is distributed according to $D_m(g_1,\dots,g_m,k+m+\frac{1}{\alpha-1};\mathrm{d}x)$, then we have the identity in distribution

$$E_i=\frac{W_i}{W_1+\cdots+W_m+Q},\qquad W_i\sim\Gamma(g_i),\,Q\sim\Gamma\left(k+m+\frac{1}{\alpha-1}\right).$$

As such, we may instead write

$$\int_0^1 w^{m+\frac{2-\alpha}{\alpha-1}}(1-w)^{k-1}\prod_{i=1}^m(1-wt_i)^{-g_i}\,\mathrm{d}w$$
$$=\frac{\Gamma(m+\frac{1}{\alpha-1})\Gamma(k)}{\Gamma(k+m+\frac{1}{\alpha-1})}\mathbb{E}\bigg[\bigg(1-\frac{t_1W_1+\cdots+t_mW_m}{W_1+\cdots+W_m+Q}\bigg)^{-m-\frac{1}{\alpha-1}}\bigg].$$

Using (97), we obtain

$$\begin{aligned}&\mathbb{P}(\mathcal{T}(\nu)=\overline{\beta},\tau_1\in\mathrm{d}t_1,\ldots,\tau_m\in\mathrm{d}t_m)\\ &=\frac{1}{(\alpha-1)(k-1)!}\prod_{i=1}^m\frac{\alpha\Gamma(g_i-\alpha)}{\Gamma(2-\alpha)}\frac{\Gamma(m+\frac{1}{\alpha-1})\Gamma(k)}{\Gamma(k+m+\frac{1}{\alpha-1})}\\ &\quad\times\mathbb{E}\bigg[\bigg(1-\frac{t_1W_1+\cdots+t_mW_m}{W_1+\cdots+W_m+Q}\bigg)^{-m-\frac{1}{\alpha-1}}\bigg],\end{aligned}\tag{100}$$

where, as above, W_1 are Gamma distrbuted with parameter g_i and Q is Gamma distributed with parameter $k+m+\frac{1}{\alpha-1}$.

Funding. The second author was supported in the early stages by the EPSRC funded Project EP/S036202/1 *Random fragmentation-coalescence processes out of equilibrium*.

REFERENCES

[1] ABRAHAM, R. and DEBS, P. (2020). Penalization of Galton–Watson processes. *Stochastic Process. Appl.* **130** 3095–3119. MR4080739 https://doi.org/10.1016/j.spa.2019.09.005

[2] ALDOUS, D. and POPOVIC, L. (2005). A critical branching process model for biodiversity. *Adv. in Appl. Probab.* **37** 1094–1115. MR2193998 https://doi.org/10.1239/aap/1134587755

[3] ATHREYA, K. B. (2012). Coalescence in the recent past in rapidly growing populations. *Stochastic Process. Appl.* **122** 3757–3766. MR2965924 https://doi.org/10.1016/j.spa.2012.06.015

[4] ATHREYA, K. B. (2012). Coalescence in critical and subcritical Galton–Watson branching processes. *J. Appl. Probab.* **49** 627–638. MR3012088 https://doi.org/10.1239/jap/1346955322

[5] ATHREYA, K. B. (2016). Coalescence in branching processes. In *Branching Processes and Their Applications. Lect. Notes Stat.* **219** 3–22. Springer, Cham. MR3587978

[6] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes. Die Grundlehren der Mathematischen Wissenschaften, Band* 196. Springer, New York. MR0373040

[7] BERESTYCKI, J., BERESTYCKI, N. and SCHWEINSBERG, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** 1835–1887. MR2349577 https://doi.org/10.1214/009117906000001114

[8] BERTOIN, J. and LE GALL, J.-F. (2000). The Bolthausen–Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields* **117** 249–266. MR1771663 https://doi.org/10.1007/s004400050006

[9] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR0898871 https://doi.org/10.1017/CBO9780511721434

[10] BOENKOST, F., FOUTEL-RODIER, F. and SCHERTZER, E. (2022). The genealogy of nearly critical branching processes in varying environment. Available at arXiv:2207.11612.

[11] BÜHLER, W. J. (1971). Generations and degree of relationship in supercritical Markov branching processes. *Z. Wahrsch. Verw. Gebiete* **18** 141–152. MR0295444 https://doi.org/10.1007/BF00569184

[12] CHAMAYOU, J.-F. and WESOŁOWSKI, J. (2009). Lauricella and Humbert functions through probabilistic tools. *Integral Transforms Spec. Funct.* **20** 529–538. MR2543763 https://doi.org/10.1080/10652460802645750

[13] CHEEK, D. and JOHNSTON, S. G. G. (2023). Ancestral reproductive bias in branching processes. *J. Math. Biol.* **86** Paper No. 70, 24. MR4572129 https://doi.org/10.1007/s00285-023-01907-7

[14] DURRETT, R. (1978). The genealogy of critical branching processes. *Stochastic Process. Appl.* **8** 101–116. MR0511879 https://doi.org/10.1016/0304-4149(78)90071-6

[15] GERNHARD, T. (2008). The conditioned reconstructed process. *J. Theoret. Biol.* **253** 769–778. MR2964590 https://doi.org/10.1016/j.jtbi.2008.04.005

[16] GROSJEAN, N. and HUILLET, T. (2018). On the genealogy and coalescence times of Bienaymé–Galton–Watson branching processes. *Stoch. Models* **34** 1–24. MR3776534 https://doi.org/10.1080/15326349.2017.1375958

[17] HARRIS, S. C., JOHNSTON, S. G. G. and PARDO, J.-C. (2023). Universality classes for the coalescent structure of critical Galton–Watson trees with infinite mean. In preparation.

[18] HARRIS, S. C., JOHNSTON, S. G. G. and ROBERTS, M. I. (2020). The coalescent structure of continuous-time Galton–Watson trees. *Ann. Appl. Probab.* **30** 1368–1414. MR4133376 https://doi.org/10.1214/19-AAP1532

[19] HARRIS, S. C., PALAU, S. and PARDO, J.-C. (2022). The coalescent structure of Galton–Watson trees in varying environments. Available at arXiv:2207.10923.

[20] HARRIS, S. C. and ROBERTS, M. I. (2017). The many-to-few lemma and multiple spines. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 226–242. MR3606740 https://doi.org/10.1214/15-AIHP714

[21] JOHNSTON, S. G. G. (2019). The genealogy of Galton–Watson trees. *Electron. J. Probab.* **24** 1–35. MR4003147 https://doi.org/10.1214/19-ejp355

[22] JOHNSTON, S. G. G. and LAMBERT, A. (2023). The coalescent structure of uniform and Poisson samples from multitype branching processes. *Ann. Appl. Probab.* To appear.

[23] JOHNSTON, S. G. G. and PROCHNO, J. (2022). Faà di Bruno's formula and inversion of power series. *Adv. Math.* **395** 108080. MR4363579 https://doi.org/10.1016/j.aim.2021.108080

[24] KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

[25] LAGERÅS, A. N. and SAGITOV, S. (2008). Reduced branching processes with very heavy tails. *J. Appl. Probab.* **45** 190–200. MR2409320 https://doi.org/10.1239/jap/1208358961

[26] LAMBERT, A. (2003). Coalescence times for the branching process. *Adv. in Appl. Probab.* **35** 1071–1089. MR2014270 https://doi.org/10.1239/aap/1067436335

[27] LAMBERT, A. and POPOVIC, L. (2013). The coalescent point process of branching trees. *Ann. Appl. Probab.* **23** 99–144. MR3059232 https://doi.org/10.1214/11-AAP820

[28] LAMBERT, A. and STADLER, T. (2013). Birth-death models and coalescent point proceses: The shape and probability of reconstructed phylogenies. *Theor. Popul. Biol.* **90** 113–128.

[29] LAURICELLA, G. (1893). Sulle funzioni ipergeometriche a più variabili. *Rend. Circ. Mat. Palermo* **7** 111–158.

[30] LE, V. (2014). Coalescence times for the Bienaymé–Galton–Watson process. *J. Appl. Probab.* **51** 209–218. MR3189452 https://doi.org/10.1239/jap/1395771424

[31] O'CONNELL, N. (1995). The genealogy of branching processes and the age of our most recent common ancestor. *Adv. in Appl. Probab.* **27** 418–442. MR1334822 https://doi.org/10.2307/1427834

[32] PAKES, A. G. (2010). Critical Markov branching process limit theorems allowing infinite variance. *Adv. in Appl. Probab.* **42** 460–488. MR2675112 https://doi.org/10.1239/aap/1275055238

[33] SLACK, R. S. (1968). A branching process with mean one and possibly infinite variance. *Z. Wahrsch. Verw. Gebiete* **9** 139–145. MR0228077 https://doi.org/10.1007/BF01851004

[34] YAKYMIV, A. L. (1980). Reduced branching processes. *Theory Probab. Appl.* **20** 602–612.

[35] ZUBKOV, A. M. (1975). Limit distributions of the distance to the nearest common ancestor. *Theory Probab. Appl.* **20** 602–612. MR0397915