Probability Surveys
Vol. 14 (2017) 53–154
ISSN: 1549-5787
DOI: 10.1214/16-PS272

Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees

Cecilia Holmgren* **and Svante Janson†**

Department of Mathematics, Uppsala University, PO Box 480
SE-751 06 Uppsala, Sweden
e-mail: cecilia.holmgren@math.uu.se; svante.janson@math.uu.se

Abstract: This survey studies asymptotics of random fringe trees and extended fringe trees in random trees that can be constructed as family trees of a Crump–Mode–Jagers branching process, stopped at a suitable time. This includes random recursive trees, preferential attachment trees, fragmentation trees, binary search trees and (more generally) m-ary search trees, as well as some other classes of random trees.

We begin with general results, mainly due to Aldous (1991) and Jagers and Nerman (1984). The general results are applied to fringe trees and extended fringe trees for several particular types of random trees, where the theory is developed in detail. In particular, we consider fringe trees of m-ary search trees in detail; this seems to be new.

Various applications are given, including degree distribution, protected nodes and maximal clades for various types of random trees. Again, we emphasise results for m-ary search trees, and give for example new results on protected nodes in m-ary search trees.

A separate section surveys results on the height of the random trees due to Devroye (1986), Biggins (1995, 1997) and others.

This survey contains well-known basic results together with some additional general results as well as many new examples and applications for various classes of random trees.

MSC 2010 subject classifications: Primary 60C05; secondary 05C05, 05C80, 60J80, 60J85, 68P05, 68P10.
Keywords and phrases: Random trees, fringe Trees, extended fringe trees, m-ary search trees, random recursive trees, preferential attachment trees, fragmentation trees, protected nodes, clades, branching processes.

Contents

*CH partly supported by the Swedish Research Council.
†SJ partly supported by the Knut and Alice Wallenberg Foundation.

1. Introduction

Aldous [1] introduced the concept of a random fringe subtree of a random tree. (See Section 4 below for definitions.) This is a useful concept since many properties of a tree can be formulated in terms of fringe trees, and thus results on the asymptotic distribution of fringe trees can imply various other asymptotic results; a simple example is the degree distribution (considered already in [1]) and some other examples are given in Section 10 (protected nodes and rank). (See also Devroye and Janson [37] and Holmgren and Janson [59] for some recent related applications of fringe trees.) Moreover, Aldous [1] also introduced the extended fringe tree that allows for consideration of e.g. parents and siblings of a chosen node; see Section 11 for some applications (e.g. maximal clades).

It is thus of interest to describe the asymptotic distribution of random fringe trees and extended fringe trees for various classes of (random) trees. Aldous [1] gave several examples of asymptotic fringe trees, including the case of random binary search trees; he also, more briefly, gave examples of asymptotic extended fringe trees. One of the purposes of the present paper is to extend these examples. In particular, we describe asymptotic fringe trees and extended fringe trees for m-ary search trees (see Section 3 for a definition and Section 7 for results). We give some applications of these results for m-ary search trees in Sections 10 and 11.

Our characterization uses some of the ideas in Aldous [1], in particular the reduction to results for continuous-time Crump–Mode–Jagers branching processes by Jagers and Nerman [65], [95]. (The m-ary search trees have earlier been studied by similar methods by Pittel [102]; however there the focus was on the height of the trees and not on the fringe trees.) In a sense, the results are implicit in [1], and partly in [65, 95], but the details are not completely trivial so we give a detailed explicit treatment.

We therefore begin with a survey of fringe trees and extended fringe trees for family trees of Crump–Mode–Jagers branching processes, including many other examples besides the m-ary search trees. The general theory is described in Sections 4 and 5. In Section 6, several examples are studied in detail, in particular various versions of preferential attachment trees, which earlier have been studied by these methods by Oliveira and Spencer [98], Rudas, Tóth and Valkó [105] and Rudas and Tóth [104]. We then specialise on m-ary search trees; explicit results for them are given in Section 7. In Section 8 we consider the random median-of-$(2\ell+1)$ binary search tree as yet another example.

Furthermore, as another novel example, we consider in Section 9 the class of fragmentation trees; these too can be constructed using family trees of Crump–Mode–Jagers branching processes, but in a slightly different way from the preceding examples. We extend the results for the asymptotic distribution of random (extended) fringe trees to this case too.

In Sections 10 and 11, as mentioned above, we give some applications of the results on asymptotic fringe trees and extended fringe trees to protected nodes and maximal clades (and related properties). This serves partly to illustrate the general theory and its uses and some results are old, but we also give a number of new results for m-ary search trees. In particular, we give a recursion that yields the asymptotic probability that a random node in an m-ary search tree is k-protected, for general m and k, and a closed formula for the case $k = 2$, together with asymptotics as $m \to \infty$ of this probability for $k = 2$.

In the main part of the paper, we consider the fringe tree or other properties of a uniformly random node in the tree. In Section 12 we consider variations for a random node with a non-uniform distribution. We study first restricted sampling, where we sample only nodes with some given property, for example a random leaf. For m-ary search trees, we study also the node containing a random key.

In Sections 4–12, we study (more or less) local properties of the tree, that are related to (extended) fringe trees. Branching process methods have also for a long time, beginning with Devroye [29], been used to study global properties of random trees, such as the height and other properties related to the distance to the root from the nodes. As a complement to the previous sections, we give in Section 13 a survey of such results for the height. This uses the same set-up as the preceding sections with random trees constructed as family trees of Crump–Mode–Jagers branching processes, but the methods are different and based on results on branching random walks by Biggins [11, 12, 13, 14]. This section is thus essentially independent of the previous sections, except for definitions and some basic results. The main results are well-known,

but we believe that the results on moment convergence in Section 13.2 are new.

In this paper, we concentrate on results obtained by general branching process methods, in particular results on the asymptotic distribution of (extended) fringe trees and applications of such results. Typical results can be expressed as convergence in probability or almost surely (a.s.) of the fraction of fringe trees that are isomorphic to some given tree, see for example (5.22); see also (4.7)–(4.8) and Remark 4.1. Such results can be seen as a law of large numbers for fringe trees, and typical applications yield first-order results for the proportion or number of nodes that have a certain property (see Sections 6–12). In some special cases, for example for some properties of the binary search tree, much more precise results have been derived by other methods. We give some references to such results, but we do not attempt completeness.

A natural next step would be to show a general central limit theorem, i.e., asymptotic normality of the number of fringe trees of a given type, under suitable conditions. This will not be attempted in the present paper, but we give some comments and references in Section 14; in particular we note that such results have been proved, by other methods, for some special cases (the binary search tree and random recursive tree), but that they do not hold in other cases (m-ary search tree with $m \geqslant 27$).

The appendices contain some results that are used in the main part of the paper.

Remark 1.1. In the present paper we consider random trees that are generated by stopping a supercritical branching process at a suitable (random) time, for example when its size (the number of individuals) is a given number.

Note that the results are quite different from the results for fringe trees of conditioned Galton–Watson trees, where we also start with a branching process but instead of stopping it, we let it run until extinction and condition on its total size being a given finite number, see [1, 7, 69, 70].

2. Some notation

The trees considered here are *rooted* and *finite*, unless otherwise indicated. (The infinite sin-trees, that arise as limits in Section 4, are important exceptions.) Furthermore, the trees are *ordered*, again unless otherwise indicated; unordered trees may be considered by giving them an arbitrary (e.g. random) ordering of the children of each node.

Moreover, there may be further information on the children of each node. In a *binary tree*, each child is labelled as left or right (with at most one child of each type at any node); the tree is ordered, with a left child before a right child, but also a single child is labelled left or right. More generally, in an *m-ary* tree, see Section 3, a node has m slots for children and the children are labelled with distinct numbers in $\{1, \dots, m\}$; these numbers determine the order of the children, but not conversely, since a node may have less than m children and thus only use a subset of these labels. (In an extended m-ary tree, each node

has either m children or 0, so these labels are determined by the order and are therefore redundant.)

We write $T_1 \approx T_2$ when T_1 and T_2 are isomorphic rooted trees. We often identify trees that are isomorphic. We may regard all finite rooted trees as subtrees of the infinite Ulam–Harris tree with node set $V_\infty := \bigcup_{n=0}^\infty \mathbb{N}^n$ consisting of all finite strings of natural numbers, where $\emptyset$ is the root and the mother of $i_1 \cdots i_k$ is $i_1 \cdots i_{k-1}$, see e.g. [55, § VI.2] and [96].

Let $|T|$ be the number of nodes in a tree T.

We regard the edges in a tree as directed from the root. Thus the *outdegree* $d^+(v) = d_T^+(v)$ of a node v in a tree T is its number of children.

The *depth* $h(v)$ of a node v is its distance from the root.

Given a tree T and a node $v \in T$, let T^v denote the subtree rooted at v, i.e., the subtree consisting of v and its descendants.

If T and S are trees, let $n_S(T)$ be the number of nodes v in T such that $T^v \approx S$. Similarly, given a property $\mathcal{P}$ of nodes in a tree, let $n_{\mathcal{P}}(T)$ be the number of nodes v in T that have the property $\mathcal{P}$. (Thus $n_S(T) = n_{\mathcal{P}_S}(T)$ if $\mathcal{P}_S$ is the property of v that $T^v \approx S$.)

For a random rooted tree $\mathcal{T}$ and a fixed tree S, let $p_S(\mathcal{T}) = \mathbb{P}(\mathcal{T} \approx S)$. Furthermore, if $\mathcal{P}$ is a property of nodes, let $p_{\mathcal{P}}(\mathcal{T})$ be the probability that the root of $\mathcal{T}$ has the property $\mathcal{P}$. (Note that $p_S(\mathcal{T}) = p_{\mathcal{P}_S}(\mathcal{T})$ with $\mathcal{P}_S$ as in the preceding paragraph, so the notation is consistent.)

Note that when we talk about a property $\mathcal{P}$ of nodes, it is implicit that the property depends also on the tree containing the node, so it is really a property of pairs (v, T) with $v \in T$. We will frequently consider properties of a node v that depend only on v and its descendants, i.e., on the subtree T^v. In this case (but not in general), we may also regard the property $\mathcal{P}$ as a property of rooted trees: we say that a tree T has $\mathcal{P}$ if the root of T has $\mathcal{P}$. In this case we also use $\mathcal{P}$ for the set of rooted trees that have the property $\mathcal{P}$; thus a node v in a tree T has $\mathcal{P} \iff T^v \in \mathcal{P}$.

If $\mathcal{P}$ is a property of nodes, we sometimes write $v \in \mathcal{P}$ for the event that v has $\mathcal{P}$.

By $\mathrm{Exp}(\lambda)$, we mean the exponential distribution with *rate* $\lambda > 0$, and thus mean $1/\lambda$: if $X \sim \mathrm{Exp}(\lambda)$ then $\mathbb{P}(X > t) = e^{-\lambda t}$. $\mathrm{Ge}_0(p)$ denotes the geometric distribution with probability function $p(1-p)^k$, $k \geqslant 0$, and $\mathrm{Ge}_1(p)$ denotes the shifted geometric distribution with probability function $p(1-p)^{k-1}$, $k \geqslant 1$.

We let $\langle x \rangle_k$ and $(x)_k$ denote the rising and falling factorials:

$$\langle x \rangle_k := x(x+1)\cdots(x+k-1) = \Gamma(x+k)/\Gamma(x), \tag{2.1}$$

$$(x)_k := x(x-1)\cdots(x-(k-1)) = \Gamma(x+1)/\Gamma(x-k+1). \tag{2.2}$$

We let $\mathbb{Z}_{\geqslant 0} := \{0, 1, 2, \dots\}$ and $\mathbb{Z}_{\leqslant 0} := \{0, -1, -2, \dots\}$.

We say that a function $f(x)$ is *decreasing* if $x < y$ implies $f(x) \geqslant f(y)$; note that we allow equality. (This is sometimes called *weakly decreasing*.) If $x < y$ implies $f(x) < f(y)$, we may say *strictly decreasing*. *Increasing* and *strictly increasing* are defined similarly.

We consider asymptotics of various random trees when some parameter n (for example the number of nodes, or number of keys in an m-ary search tree) tends to infinity. Similarly, for the continuous-time branching processes, we consider limits as the time t tends to infinity. As usual, *w.h.p.* (*with high probability*) means with probability tending to 1.

3. m-ary search trees

An *m-ary search tree*, where $m \geqslant 2$ is a fixed number, is an m-ary tree constructed recursively from a sequence of distinct *keys* (real numbers) as follows, see e.g. [85] or [40]. (In the case $m = 2$, we say *binary search tree*.) The m-ary search trees were first introduced in [93].

Each node may store up to $m-1$ keys. We start with a tree containing just an empty root. The first $m-1$ keys are stored in the root. When the $(m-1)$:th key is placed in the root, so the root becomes full, we add m new nodes, initially empty, as children of the root. Furthermore, the $m-1$ keys in the root divide the set of real numbers into m intervals $J_1, \dots, J_m$. Each further key is passed to one of the children of the root depending on which interval it belongs to; a key in J_i is passed to the i:th child.

This construction yields the *extended m-ary search tree*. Nodes containing at least one key are called *internal* and empty nodes are called *external*. Usually one eliminates all external nodes and consider the tree consisting of the internal nodes only; this is the *m-ary search tree*.

For both versions, we often wish to keep track of the number of keys in each node, so we regard the trees as labelled trees where each node has a label in $\{0, \dots, m-1\}$ indicating the number of keys. (Thus external nodes have label 0 while internal nodes have labels in $\{1, \dots, m-1\}$.)

We assume that the keys are i.i.d. random variables with a continuous distribution, for example $U[0,1]$. With a given number n of keys, this gives a *random m-ary search tree* T_n. (As is customary, we usually omit the word "random" for convenience. Also, we regard m as fixed, and omit it from the notation.) Note that only the order of the keys matter; hence we obtain the same random m-ary search tree T_n also if we instead let the keys be a uniformly random permutation of $\{1, \dots, n\}$.

Note that in T_n we have fixed the number of keys; not the number of nodes. A node may contain $1, \dots, m-1$ keys, and the total number of nodes will be random when $m \geqslant 3$. (The binary case $m = 2$ is an exception; each internal node contains exactly one key, so the number of (internal) nodes equals the number n of keys, and the number of external nodes is $n+1$.)

In an extended m-ary search tree, say that a node with $i \leqslant m-2$ keys has $i+1$ *gaps*, while a full node has no gaps. It is easily seen that an extended m-ary search tree with n keys has $n+1$ gaps; the gaps correspond to the intervals of real numbers between the keys (and $\pm\infty$), and a new key has the same probability $1/(n+1)$ of belonging to any of the gaps. Thus the evolution of the extended m-ary search tree may be described by choosing a gap uniformly at random at

each step. Equivalently, the probability that the next key is added to a node is proportional to the number of gaps at that node. For the m-ary search tree (with only internal nodes) the same holds with minor modifications; a full node now has one gap for each external node in the extended version, i.e., $m-d$ gaps if there are d children, and a key added to one of its gaps now starts a new node.

4. Fringe trees and extended fringe trees

Given a (finite, rooted) tree T, the *random fringe tree* of T is the random tree obtained by taking the subtree T^v with v chosen uniformly at random from the nodes of T; we denote the random fringe tree of T by T^*.

Consider a sequence T_n of (possibly random) trees such that the random fringe tree T_n^* converges in distribution to some random tree $\mathcal{F}$:

$$T_n^* \xrightarrow{\mathrm{d}} \mathcal{F}, \tag{4.1}$$

which simply means (since the set of finite trees is countable)

$$\mathbb{P}(T_n^* \approx S) \to \mathbb{P}(\mathcal{F} \approx S) \tag{4.2}$$

for every finite rooted tree S. We then say, following Aldous [1], that $\mathcal{F}$ (or rather its distribution) is the asymptotic fringe distribution of T_n.

If the trees T_n are deterministic, then (4.2) can be written

$$\frac{n_S(T_n)}{|T_n|} \to p_S(\mathcal{F}), \tag{4.3}$$

for every tree S; this is equivalent to the seemingly more general

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \to p_{\mathcal{P}}(\mathcal{F}), \tag{4.4}$$

for every property $\mathcal{P}$ of a node v that depends only on the subtree T^v, i.e., on v and its descendants.

In the more general case when T_n are random (which is the case we are interested in), (4.2) instead can be written

$$\mathbb{E}\,\frac{n_S(T_n)}{|T_n|} \to p_S(\mathcal{F}) \tag{4.5}$$

or, more generally but equivalently,

$$\mathbb{E}\,\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \to p_{\mathcal{P}}(\mathcal{F}) \tag{4.6}$$

for properties $\mathcal{P}$ as above. In interesting cases, we may typically strengthen (4.5)–(4.6) to convergence in probability:

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\mathrm{p}} p_{\mathcal{P}}(\mathcal{F}); \tag{4.7}$$

Aldous [1, Proposition 7] gives a general criterion for this (the distribution of $\mathcal{F}$ is extremal in the set of fringe distributions), but we will instead prove (4.7) directly in the cases considered here; moreover, we will in our cases prove convergence almost surely:

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\mathcal{F}). \tag{4.8}$$

Remark 4.1. Note that

$$\frac{n_S(T_n)}{|T_n|} = \mathbb{P}\bigl(T_n^* \approx S \mid T_n\bigr) \tag{4.9}$$

and, more generally, for a property $\mathcal{P}$ as above,

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} = \mathbb{P}\bigl(T_n^* \in \mathcal{P} \mid T_n\bigr). \tag{4.10}$$

It follows from (4.9) that (4.7) and (4.8) (for all properties $\mathcal{P}$ considered there) are equivalent to conditional versions of (4.1):

$$\mathcal{L}\bigl(T_n^* \mid T_n\bigr) \xrightarrow{\text{p}} \mathcal{L}\bigl(\mathcal{F}\bigr) \tag{4.11}$$

and

$$\mathcal{L}\bigl(T_n^* \mid T_n\bigr) \xrightarrow{\text{a.s.}} \mathcal{L}\bigl(\mathcal{F}\bigr), \tag{4.12}$$

respectively, with convergence in probability or a.s. of the conditional distribution, in the space of probability distributions on trees. (Note that any such property $\mathcal{P}$ corresponds to a set of finite rooted trees T, and conversely.)

Results such as (4.11) and (4.12), where we fix a realization T_n of a random tree and then study the distribution of its fringe tree (or something else), as a random variable depending on T_n, are usually called *quenched*, while results such as (4.1), where we consider the random fringe tree of a random tree as a combined random event, are called *annealed*. See further e.g. [37] and [70].

4.1. Extended fringe trees

The fringe tree T^* considers only the descendants of a random node. Aldous [1] introduced also the *extended fringe trees* that include the nearest ancestors and other close relatives. If $k \geqslant 0$ and $v \in T$ with $h(v) \geqslant k$, let $v^{(k)}$ be the ancestor of v that is k generations earlier (i.e., with $h(v^{(k)}) = h(v) - k$), and let $T^{v,-k}$ be the subtree rooted at $v^{(k)}$, with the node v marked. (Or, equivalently, with the path from the root $v^{(k)}$ to v marked.) Thus $T^{v,-k}$ is a rooted tree with a distinguished node of depth k. (Note that $T^{v,-0} = T^v$.)

We define the random extended fringe tree $T^{*,-k}$ as $T^{v,-k}$ for a uniformly random node $v \in T$; this is really not defined when $h(v) < k$, but we may

define $T^{v,-k}$ in this case too by some supplementary definition, for example as a path of length $k-h(v)$ with a copy of T attached, with v marked. We are only interested in asymptotics of the random extended fringe trees for sequences of trees T such that

$$h(v) \xrightarrow{\mathrm{p}} \infty \tag{4.13}$$

for a random node v, i.e., $\mathbb{P}(h(v) < k) \to 0$ for every fixed k, and thus each $T^{*,-k}$ is well-defined w.h.p., and then the supplementary definition does not matter.

Aldous [1] showed that if T_n is a sequence of (possibly random) trees such that (4.13) holds and an asymptotic fringe distribution exists, i.e., (4.1) holds, then, more generally, each $T_n^{*,-k}$ converges in distribution to some random tree $\mathcal{F}^{-k}$ with a distinguished node o of depth k. Note that the trees $T_n^{v,-k}$ are consistent in an obvious way, with $T_n^{v,-(k-1)}$ a subtree of $T_n^{v,-k}$, and thus the same holds for the limits $\mathcal{F}^{-k}$ (after a suitable coupling). Hence it is possible to regard the trees $\mathcal{F}^{-k}$ as subtrees of a (random) infinite tree $\widetilde{\mathcal{F}}$ with a distinguished node o and an infinite line $o, o^{(1)}, o^{(2)}, \dots$ of ancestors of o, such that $\mathcal{F}^{-k} = \widetilde{\mathcal{F}}^{o,-k} = \widetilde{\mathcal{F}}^{o^{(k)}}$. Furthermore, every node in $\widetilde{\mathcal{F}}$ has a finite number of descendants; thus there are no other infinite paths from o. (Aldous [1] calls such a tree a *sin-tree*, for single infinite path.) We may then say that the extended fringe trees converge to the random sin-tree $\widetilde{\mathcal{F}}$, in the sense that $T_n^{*,-k} \xrightarrow{\mathrm{d}} \widetilde{\mathcal{F}}^{o^{(k)}}$ for each k, or, equivalently, using the product topology on the set of sequences of (finite) trees,

$$\bigl(T_n^{*,-k}\bigr)_{k=0}^{\infty} \xrightarrow{\mathrm{d}} \bigl(\widetilde{\mathcal{F}}^{o^{(k)}}\bigr)_{k=0}^{\infty}. \tag{4.14}$$

For a random sin-tree $\widetilde{\mathcal{F}}$ and a property $\mathcal{P}$ of nodes, let $p_{\mathcal{P}}(\widetilde{\mathcal{F}})$ be the probability that the distinguished node o has the property $\mathcal{P}$. Then, cf. (4.6) (which is the case $k=0$), (4.14) implies, and is equivalent to,

$$\mathbb{E}\,\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \to p_{\mathcal{P}}(\widetilde{\mathcal{F}}), \tag{4.15}$$

for every property $\mathcal{P}$ that depends only on $T^{v,-k}$ for some k, i.e., on v and its descendants and the descendants of its ancestors at most a fixed number of generations back. Again, we may typically strengthen (4.15) to convergence in probability, and in our cases we shall prove convergence a.s.:

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\mathrm{a.s.}} p_{\mathcal{P}}(\widetilde{\mathcal{F}}). \tag{4.16}$$

By standard truncation arguments, it may be possible to extend (4.15) or (4.16) also to some more general properties $\mathcal{P}$, depending on an unlimited number of ancestors, see Sections 5.1 and 11 for some examples.

Remark 4.2. Similarly to Remark 4.1, (4.16) is equivalent to a conditional version of (4.14):

$$\mathcal{L}\bigl(\bigl(T_n^{*,-k}\bigr)_{k=0}^{\infty} \mid T_n\bigr) \xrightarrow{\mathrm{a.s.}} \mathcal{L}\bigl(\bigl(\widetilde{\mathcal{F}}^{o^{(k)}}\bigr)_{k=0}^{\infty}\bigr). \tag{4.17}$$

5. Family trees of general branching processes

A *Crump–Mode–Jagers* process is a general branching process defined as follows, see e.g. [64] for further details and for basic facts used below.

The branching process starts with a single individual born at time 0. This individual has a random number N of children, born at random times $(\xi_i)_{i=1}^N$; here $0 \leqslant N \leqslant \infty$, and we assume $0 \leqslant \xi_1 \leqslant \xi_2 \leqslant \cdots$. It is convenient to describe the birth times $\{\xi_i\}_1^N$ as a point process Ξ on $[0,\infty)$. Every child that is born evolves in the same way, i.e., every individual x has its own copy Ξ_x of Ξ (where now ξ_i means the age of the mother when child i is born); these copies are assumed to be independent and identically distributed. Denote the time an individual x is born by σ_x.

Recall that formally a point process Ξ is best defined as an integer-valued random measure, where $\Xi(A)$ is the number of points in a set A, see e.g. [75]. In our setting, $\Xi = \sum_{i=1}^N \delta_{\xi_i}$, where δ_t is a point mass (Dirac measure) at t. In particular, we have $N = \Xi([0,\infty))$. Furthermore, for $j \leqslant N$, $\Xi([0,t]) \geqslant j$ if and only if $\xi_j \leqslant t$. We let $\mu := \mathbb{E}\,\Xi$ denote the intensity measure of Ξ, and write $\mu(t) := \mu([0,t]) = \mathbb{E}\,\Xi([0,t])$. In particular, $\mu(\infty) = \mathbb{E}\,N$.

Usually one also assumes that each individual has a random lifetime $\lambda \in [0,\infty]$; for our purposes this plays no role, so we ignore it. (Formally, we may assume that $\lambda = \infty$.) There may also be other random variables associated to the individuals. Formally, we give each possible individual x its own copy $(\Omega_x, \mathcal{F}_x, \mu_x)$ of some probability space $(\Omega, \mathcal{F}, \mu)$ on which there are defined some given functions defining N, ξ_i (and thus Ξ), and possibly other random variables describing the life history such as the marks ν_i or label $\ell(t)$ in Remarks 5.1 and 5.2 below; the branching process then is defined on the product $\prod_x(\Omega_x, \mathcal{F}_x, \mu_x)$ of these probability spaces. (The individuals may be labelled in a natural way by strings in $V_\infty := \bigcup_{n=0}^\infty \mathbb{N}^n$; hence the set of individuals that are realized in the branching process is a random subset of V_∞, and we may extend the product over $x \in V_\infty$.)

Let Z_t be the number of individuals at time $t \geqslant 0$; since we assume no deaths, this equals the number of individuals born in $[0,t]$. (We follow standard custom and let all processes be right-continuous; thus an individual born at t exists at t and is included.) We say that the process is *finite* (or *dies out*) if $Z_\infty < \infty$, i.e., only a finite number of individuals are ever born.

Let $\mathcal{T}_\infty$ be the *family tree* of the branching process. This is a (generally infinite) tree obtained from the branching process by ignoring the time structure; in other words, it has the individuals as nodes, with the initial individual as the root, and the children of a node in the tree are the same as the children in the branching process. Let $\mathcal{T}_t$ be the subtree consisting of all individuals born up to time t. Note that the number of nodes $|\mathcal{T}_t| = Z_t$. (We are mainly interested in cases where $Z_t < \infty$ for every finite t, but $Z_\infty = \infty$.)

Remark 5.1. This defines the family tree $\mathcal{T}_t$ as an unordered tree. Sometimes we want an ordered tree, so we have to add an ordering of the children of each individual. This can be done by taking the children in order of birth (which is

the standard custom), but in our examples we rather want a random order. In general, we can obtain ordered family trees by assuming that each individual has a marked point process Ξ^* (augmenting the plain Ξ above), where each point ξ_i has a mark $\nu_i \in \{1, \dots, i\}$ telling at which position the new child is inserted among the existing ones. (This includes both the birth order case, with $\nu_i = i$, and the random order case, with ν_i uniform and independent of everything else.)

For the m-ary search trees in Section 7, we want further information; this is obtained by instead giving each of the m children a distinct mark $\nu_i \in \{1, \dots, m\}$ telling the position of the child among all (existing and future) children. (Equivalently, we may equip each individual with a random permutation of $\{1, \dots, m\}$ giving the order of birth of the children.)

Remark 5.2. We may also have labels on the nodes of $\mathcal{T}_t$; this is important for our application to m-ary search trees, since they have nodes labelled with the number of keys, see Section 3. In general, we may assume that each individual has a label given by some random function $\ell(t)$ of its age. We assume that the set of possible labels is countable (with the discrete topology); we may assume that the labels are integers. We also assume that the function $\ell(t) \in D[0,\infty)$; thus $\ell(t)$ is constant on some intervals $[t_i, t_{i+1})$. (As everything else in the branching process, the label may depend on Ξ and other properties of the same individual, but not on other individuals, and they have the same distribution for all individuals; this is also a consequence of the formalism with probability spaces $(\Omega_x, \mathcal{F}_x, \mu_x)$ above.)

A *characteristic* of an individual, see e.g. [64, 65, 94, 95], is a random function $\phi(t)$ of the age $t \geqslant 0$; we assume that $\phi(t) \geqslant 0$ and that ϕ belongs to the space $D[0,\infty)$ of right-continuous functions with left limits. (Note that we consider only $t \geqslant 0$. We may extend ϕ to $(-\infty, \infty)$ by setting $\phi(t) = 0$ for $t < 0$.) We assume that each individual has its own copy ϕ_x, and we at first for simplicity assume that the pairs (Ξ_x, ϕ_x) for all individuals are independent and identically distributed; this assumption can (and will) be relaxed, see Remark 5.10 below.

Given a characteristic ϕ, let

$$Z_t^\phi := \sum_{x:\sigma_x \leqslant t} \phi_x(t - \sigma_x) \tag{5.1}$$

be the total characteristic at time t of all individuals that have been born so far. (Recall that x is born at time σ_x, and thus has age $t - \sigma_x$ at time t.)

The random tree $\mathcal{T}_t$ has a random size. We are usually interested in random trees with a given number of nodes, or trees where something else is given, for example the number of keys in an m-ary search tree. We can obtain such random trees by stopping the branching process as follows. Fix a characteristic $\psi(t)$, which we shall call *weight*, and let $\tau(n) := \inf\{t : Z_t^\psi \geqslant n\}$, i.e., the first time the total weight is at least n. (As usual, we define $\inf \emptyset = \infty$.) We exclude the trivial case when $\psi(t) = 0$ for all $t \geqslant 0$ a.s. (which would give $\tau(n) = \infty$ a.s.). Define $T_n := \mathcal{T}_{\tau(n)}$, the family tree at the time the total weight reaches n (provided this ever happens).

Random trees T_n defined in this way, for some Crump–Mode–Jagers branching process and some weight $\psi(t)$, are the focus of the present paper. We shall always denote the weight by ψ and the random tree, stopped as above, by T_n omitting ψ from the notation for simplicity. (In all our examples, ψ is integer-valued, so it is natural to let n be an integer. This is not necessary, however, and all our results are valid for arbitrary real $n \to \infty$.)

Example 5.3. If $\psi(t) = 1$, $t \geqslant 0$, then $Z_t^\psi = Z_t$, and T_n is the family tree of the branching process stopped when there are n nodes or more; if the birth times have continuous distributions and there are no twins, then a.s. no two nodes are born simultaneously, and thus we stop when there are exactly n nodes, so $|T_n| = n$. (This weight is used in all examples in Section 6, but not for the m-ary search trees in Section 7.)

We define the Laplace transform of a function f on $[0,\infty)$ by

$$\widehat{f}(\theta) = \theta \int_0^\infty e^{-\theta t} f(t)\,\mathrm{d}t, \qquad \theta > 0, \tag{5.2}$$

and the Laplace transform of a measure m on $[0,\infty)$ by

$$\widehat{m}(\theta) = \int_0^\infty e^{-\theta t} m(\mathrm{d}t), \qquad -\infty < \theta < \infty. \tag{5.3}$$

(Note that there is a factor θ in (5.2) but not in (5.3). A justification of this difference is that a measure m has the same Laplace transform $\widehat{m}$ as the function $m(t) := m([0,t])$, as is easily verified by an integration by parts, or by Fubini's theorem for the integral $\iint_{s\leqslant t} \theta e^{-\theta t} m(\mathrm{d}s)$.)

Some standing assumptions in this paper are:

(A1) $\mu\{0\} = \mathbb{E}\,\Xi\{0\} < 1$. (This rules out a rather trivial case with explosions already at the start. In all our examples, $\mu\{0\} = 0$.)

(A2) μ is not concentrated on any lattice $h\mathbb{Z}$, $h > 0$. (The results extend to the lattice case with suitable modifications, but we ignore it.)

(A3) $\mathbb{E}\,N > 1$. (This is known as the *supercritical* case.) For simplicity, we further assume that $N \geqslant 1$ a.s., but see Remark 5.5. (In this case, every individual has at least one child, so the process never dies out and $Z_\infty = \infty$.)

(A4) There exists a real number α (the Malthusian parameter) such that $\widehat{\mu}(\alpha) = 1$, i.e.,

$$\int_0^\infty e^{-\alpha t} \mu(\mathrm{d}t) = 1. \tag{5.4}$$

(By (A3), $\alpha > 0$.)

(A5) $\widehat{\mu}(\theta) < \infty$ for some $\theta < \alpha$.

(A6) The random variable $\sup_t \bigl(e^{-\theta t}\phi(t)\bigr)$ has finite expectation for some $\theta < \alpha$.

Nerman [94, Theorem 6.3] (see also Jagers [64, Section 6.10] for related results) shows that under the conditions (A1)–(A6), as $t \to \infty$,

$$\frac{Z_t^\phi}{Z_t} \xrightarrow{\text{a.s.}} m_\phi := \mathbb{E}\,\hat{\phi}(\alpha) = \widehat{\mathbb{E}\,\phi}(\alpha) = \alpha \int_0^\infty e^{-\alpha t}\,\mathbb{E}\,\phi(t)\,\mathrm{d}t. \tag{5.5}$$

The right-hand side of (5.5) is finite by (A6). Thus, if we exclude the trivial case when $\phi(t) = 0$ for all $t \geqslant 0$ a.s., $0 < m_\phi < \infty$.

Note that (A1)–(A5) are conditions on the branching process, while (A6) is a condition on the characteristic ϕ (and α), and thus is relevant only when we consider some ϕ. When discussing trees T_n defined by stopping using a weight ψ as above, we sometimes want (A6) to hold for ψ; we denote this version of the condition by (A6ψ). (However, for most of our results, (A6ψ) is not required. In any case, in Example 5.3 and in all our examples in Sections 6 and 7, $\psi(t)$ is bounded, so (A6ψ) holds trivially.)

Remark 5.4. As a consequence of (A4), $\mu(t) < \infty$ for every $t < \infty$. (However, $\mu(\infty) = \mathbb{E}\, N$ may be infinite.) It is a standard result that this implies that Z_t and $\mathbb{E}\, Z_t$ are finite for every $t < \infty$.

Remark 5.5. We do not really need the assumption $N \geqslant 1$ in (A3); it suffices that $\mathbb{E}\, N > 1$. In this case, the extinction probability $q := \mathbb{P}(Z_\infty < \infty) < 1$, so there is a positive probability that the process is infinite, and (5.5) and the results below hold conditioned on the event $Z_\infty = \infty$. (This is the standard setting in [94, 65, 95].)

Remark 5.6. By (5.4), $e^{-\alpha t}\mu(\mathrm{d}t)$ is a probability measure on $[0, \infty)$. See Remark 5.22 for an interpretation of this distribution.

Remark 5.7. By the definitions,

$$\widehat{\Xi}(\theta) := \int_0^\infty e^{-\theta t}\Xi(\mathrm{d}t) = \sum_{i=1}^N e^{-\theta\xi_i}. \tag{5.6}$$

Since $\mu = \mathbb{E}\,\Xi$, we have $\widehat{\mu}(\theta) = \mathbb{E}\,\widehat{\Xi}(\theta)$ and (5.6) yields

$$\widehat{\mu}(\theta) = \mathbb{E}\sum_{i=1}^N e^{-\theta\xi_i}. \tag{5.7}$$

Thus, (5.4) can be written $\mathbb{E}\,\widehat{\Xi}(\alpha) = 1$, or

$$\mathbb{E}\sum_{i=1}^N e^{-\alpha\xi_i} = 1. \tag{5.8}$$

Similarly, (A5) says that the random variable $\widehat{\Xi}(\theta)$ has finite expectation $\mathbb{E}\,\widehat{\Xi}(\theta) < \infty$ for some $\theta < \alpha$.

Remark 5.8. The conditions (A5) and (A6) may be weakened somewhat if we further assume $\mathbb{E}\,\widehat{\Xi}(\alpha)\log^+\widehat{\Xi}(\alpha) < \infty$, see [94, Conditions 5.1 and 5.2], but the versions above are sufficient for our applications.

Remark 5.9. The results can be extended to multi-type branching processes, see Jagers and Nerman [66].

Remark 5.10. We have for simplicity assumed above that the characteristic $\phi_x(t)$ associated to an individual x is independent of the life histories of all other individuals. As shown by [94, Section 7], the results extend to characteristics $\phi_x(t)$ that may depend also on the descendants of x; we may let $\phi_0(t)$ be any non-negative random function that depends on the entire branching process (and belongs to $D[0,\infty)$ and satisfies (A6)), and then define $\phi_x(t)$ as $\phi_0(t)$ evaluated for the branching process consisting of x and its descendants (shifting the origin of time to the birth of x). This will be important below.

Remark 5.11. Nerman [94] showed also that, under the assumptions (A1)–(A5) above, there exists a random variable W such that, as $t\to\infty$,

$$e^{-\alpha t}Z_t \xrightarrow{\text{a.s.}} W \tag{5.9}$$

and, more generally, for every ϕ satisfying (A6),

$$e^{-\alpha t}Z_t^\phi \xrightarrow{\text{a.s.}} m_\phi W. \tag{5.10}$$

If furthermore

$$\mathbb{E}\,\widehat{\Xi}(\alpha)\log^+\widehat{\Xi}(\alpha)<\infty, \tag{5.11}$$

which is the case in our applications, then $W>0$ a.s. (on the event $\{Z_\infty=\infty\}$), see also [39], so Z_t and Z_t^ϕ grow exactly at rate $e^{\alpha t}$; moreover, (5.5) then follows from (5.9)–(5.10). However, if (5.11) fails, then $W=0$ a.s., so $Z_t^\phi=o(e^{\alpha t})$ a.s.; nevertheless, also in this case

$$e^{-\theta t}Z_t^\phi \xrightarrow{\text{a.s.}} \infty \tag{5.12}$$

for every $\theta<\alpha$, as follows easily by truncating the offspring distribution Ξ to at most M children, for some large M, and applying (5.10) to the truncated process. It follows easily from (5.10) and (5.12) that if (A1)–(A6) hold, then

$$\log Z_t^\phi/t \xrightarrow{\text{a.s.}} \alpha. \tag{5.13}$$

(See also [14, Theorem 2.1] for this result under a slightly weaker condition.)

Furthermore, the expectation $\mathbb{E}\,Z_t^\phi$ always grows as $e^{\alpha t}$, even when (5.11) fails so $W=0$ a.s.; more precisely [94, Proposition 2.1], with $\beta>0$ is given by (5.38) below,

$$\mathbb{E}\,Z_t^\phi \sim (\alpha\beta)^{-1}m_\phi e^{\alpha t}. \tag{5.14}$$

If (5.11) holds, then

$$\mathbb{E}\,W=(\alpha\beta)^{-1}, \tag{5.15}$$

so (5.14) says that the expectation converges in (5.9). (However, as just said, (5.14) holds also when (5.9) holds with $W=0$.)

Our main results are now simple consequences of the general results by Nerman [94] above. We consider the random trees T_n defined by stopping the branching process according to some fixed weight ψ as above. We begin by noting that $\tau(n)<\infty$ so that T_n really is well-defined. (See e.g. [34].)

Theorem 5.12. (i) *Under the assumptions (A1)–(A5), and for any weight ψ, $Z_t^\psi \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$; thus a.s. $\tau(n) < \infty$ for every $n \geqslant 0$ and $T_n := \mathcal{T}_{\tau(n)}$ is a well-defined finite random tree. Furthermore, $\tau(n) \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$.*

(ii) *If moreover* (A6ψ) *holds, then*

$$\frac{|T_n|}{n} \xrightarrow{\text{a.s.}} \frac{1}{m_\psi} \in (0, \infty) \tag{5.16}$$

and

$$\frac{\tau(n)}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{\alpha}. \tag{5.17}$$

Proof. (i): If (A6ψ) holds, then (5.12) (with $\theta = 0$) shows that $Z_t^\psi \xrightarrow{\text{a.s.}} \infty$. In general, apply (5.12) to the truncated weight $\psi_1(t) := \psi(t) \wedge 1$ and deduce $Z_t^\psi \geqslant Z_t^{\psi_1} \xrightarrow{\text{a.s.}} \infty$. Hence, $\tau(n) < \infty$ for every n. Finally, for every finite t, $Z_t^\psi < \infty$ a.s., and thus $\tau(n) > t$ for large n; hence $\tau(n) \to \infty$.

(ii): By the definition of $\tau(n)$ (and right-continuity of the process and ψ), $Z^\psi_{\tau(n)} \geqslant n$, while $Z^\psi_{\tau(n)-\delta} < n$ for any $\delta > 0$. Fix $\delta > 0$ and define the characteristic $\psi_\delta(t) := \psi(t-\delta)$, with $\psi_\delta(t) := 0$ for $t < \delta$. Then $Z_t^{\psi_\delta} = Z^\psi_{t-\delta}$. Consequently,

$$Z^{\psi_\delta}_{\tau(n)} = Z^\psi_{\tau(n)-\delta} < n \leqslant Z^\psi_{\tau(n)}. \tag{5.18}$$

Recall also that $|T_n| = |\mathcal{T}_{\tau(n)}| = Z_{\tau(n)}$.

We have assumed (A6) for ψ, and it follows that it holds for ψ_δ too. Consequently, (5.5) applies to both ψ and ψ_δ, which by (5.18) implies

$$\limsup_{n\to\infty} \frac{n}{|T_n|} \leqslant \limsup_{n\to\infty} \frac{Z^\psi_{\tau(n)}}{Z_{\tau(n)}} = m_\psi, \tag{5.19}$$

$$\liminf_{n\to\infty} \frac{n}{|T_n|} \geqslant \liminf_{n\to\infty} \frac{Z^{\psi_\delta}_{\tau(n)}}{Z_{\tau(n)}} = m_{\psi_\delta}. \tag{5.20}$$

Furthermore,

$$\begin{aligned} m_{\psi_\delta} &= \alpha \int_0^\infty e^{-\alpha t}\, \mathbb{E}\, \psi_\delta(t)\, \mathrm{d}t = \alpha \int_\delta^\infty e^{-\alpha t}\, \mathbb{E}\, \psi(t-\delta)\, \mathrm{d}t \\ &= \alpha \int_0^\infty e^{-\alpha(u+\delta)}\, \mathbb{E}\, \psi(u)\, \mathrm{d}u = e^{-\alpha\delta} m_\psi. \end{aligned} \tag{5.21}$$

Since $\delta > 0$ is arbitrary, the result (5.16) follows from (5.19)–(5.20) and (5.21) by letting $\delta \to 0$.

Similarly, (5.13) applies to both ψ and ψ_δ, which by (5.18) yields $\log n/\tau(n) \xrightarrow{\text{a.s.}} \alpha$, which is (5.17). □

Remark 5.13. Note that (5.16) does *not* hold in the lattice case (in this paper excluded by (A2)), since then the population and Z_t^ψ grow in discrete steps with asymptotically a fixed factor > 1 each time.

We next study the fringe tree T_n^*. Note that the following theorem (and its proof) applies both if we consider $\mathcal{T}_t$ as an unordered tree and if we consider it as an ordered (or m-ary) tree as in Remark 5.1; in the latter case T_n and the fringe tree T_n^* are random ordered (or m-ary) trees, and T below should be an ordered (or m-ary) tree. We may also have labels on the nodes, defined by some random function $\ell(t)$ as in Remark 5.2; then T should be a tree with (arbitrary) labels on the nodes.

Recall from Section 2 that a property of a node v that depends only on v and its descendants may also be regarded as a property of rooted trees (and conversely).

Theorem 5.14 (Jagers, Nerman, Aldous)**.** *Under the assumptions (A1)–(A5), the following hold:*

(i) *(Annealed version.) The random fringe tree T_n^* converges in distribution as $n \to \infty$ to the random tree $\overline{\mathcal{T}} := \mathcal{T}_{\overline{\tau}}$, where $\overline{\tau} \sim \mathrm{Exp}(\alpha)$ is a random time, independent of the branching process.*

(ii) *(Quenched version.) For every finite tree T, as $n \to \infty$,*

$$\mathbb{P}\bigl(T_n^* \approx T \mid T_n\bigr) = \frac{n_T(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(\overline{\mathcal{T}} \approx T). \tag{5.22}$$

More generally, for every property $\mathcal{P}$ of a node v that depends only on v and its descendants,

$$\mathbb{P}\bigl(T_n^* \in \mathcal{P} \mid T_n\bigr) = \frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\overline{\mathcal{T}}). \tag{5.23}$$

Furthermore, for a property of this type,

$$p_{\mathcal{P}}(\overline{\mathcal{T}}) = \int_0^\infty \alpha e^{-\alpha t} p_{\mathcal{P}}(\mathcal{T}_t)\,\mathrm{d}t = m_\phi = \mathbb{E}\,\hat{\phi}(\alpha), \tag{5.24}$$

where $\phi(t)$ is the characteristic $\mathbf{1}\{\mathcal{T}_t \in \mathcal{P}\}$.

More precisely, the characteristic ϕ in (5.24) is defined as in Remark 5.10 with $\phi_0(t) := \mathbf{1}\{\mathcal{T}_t \in \mathcal{P}\}$.

Proof. This is a special case of the main results in Jagers and Nerman [65] and [95], and is one of the main examples in Aldous [1], but we give the simple proof for completeness and in our setting.

Note first that (5.22) is a special case of (5.23), with $\mathcal{P} = \mathcal{P}_T$, see Section 2, and thus $\phi(t) = \mathbf{1}\{\mathcal{T}_t \approx T\}$. (Some readers might prefer to consider this case first.)

The function $\phi(t)$ is clearly a $\{0,1\}$-valued random function in $D[0,\infty)$, so ϕ is a characteristic. The assumption (A6) holds trivially since ϕ is bounded. Furthermore, (5.24) is a consequence of the definition $\overline{\mathcal{T}} := \mathcal{T}_{\overline{\tau}}$, the fact that $\tau \sim \mathrm{Exp}(\alpha)$ has the density function $\alpha e^{-\alpha t}$, $\mathbb{E}\,\phi(t) = \mathbb{P}(\mathcal{T}_t \in \mathcal{P}) = p_{\mathcal{P}}(\mathcal{T}_t)$, and (5.5).

Moreover, the characteristic $\phi_x(t-\sigma_x)$ of x at time t is the indicator $\mathbf{1}\{\mathcal{T}_t^x \in \mathcal{P}\}$ that the subtree $\mathcal{T}_t^x$ of $\mathcal{T}_t$ rooted at x satisfies $\mathcal{P}$. Thus, the total characteristic Z_t^ϕ is the number of nodes $v \in \mathcal{T}_t$ such that $\mathcal{T}_t^v \in \mathcal{P}$, which by definition holds if and only if v has the property $\mathcal{P}$; hence, $Z_t^\phi = n_{\mathcal{P}}(\mathcal{T}_t)$. Consequently, (5.5) yields

$$\frac{n_{\mathcal{P}}(\mathcal{T}_t)}{|\mathcal{T}_t|} = \frac{Z_t^\phi}{Z_t} \xrightarrow{\text{a.s.}} m_\phi. \tag{5.25}$$

By Theorem 5.12, we also have (a.s.) $\tau(n) < \infty$ for every n and $\tau(n) \to \infty$ as $n \to \infty$; thus (5.25) implies, as $n \to \infty$,

$$\frac{n_{\mathcal{P}}(T_n)}{|T_n|} = \frac{n_{\mathcal{P}}(\mathcal{T}_{\tau(n)})}{|\mathcal{T}_{\tau(n)}|} \xrightarrow{\text{a.s.}} m_\phi. \tag{5.26}$$

The result (5.22) follows from (5.26) and (5.24). As said above, (5.22) is a special case, and the annealed version (i) follows by taking the expectation in (5.22), yielding (by dominated convergence) $\mathcal{P}(T_n^* \approx T) \to \mathbb{P}(\overline{\mathcal{T}} \approx T)$ for every fixed tree T. (Recall that there is only a countable set of finite trees T, so this shows convergence in distribution. Alternatively, one can take the expectation of (5.23).) □

Remark 5.15. As said above, (5.22) is a special case of (5.23). Conversely, again because there is only a countable set of finite trees T, (5.22) is equivalent to the a.s. convergence of the distributions in (4.12), and thus to (5.23), cf. Remark 4.1. (In general, for distributions on a countable sample space, convergence of the individual point probabilities is equivalent to convergence in total variation [54, Theorem 5.6.4].) Hence, (5.22) and (5.23) are equivalent. (We state both versions for convenience in later applications.)

Remark 5.16. We have stated the result (5.22) for the stopped trees T_n, but proved it by proving the corresponding result for the full branching process, see (5.25) and (5.26). In fact, the two types of results are equivalent; by choosing the weight $\psi = 1$ as in Example 5.3, the trees $\mathcal{T}_t$ run through the same (countable) set of trees as $t \to \infty$ as T_n does as $n \to \infty$; hence (5.25) and (5.26) are equivalent. The same holds for (5.23) and for (5.43) and (5.51) in Theorems 5.25 and 5.26 below, where again we state the results for T_n, in view of our applications in later sections, but the results also hold for $\mathcal{T}_t$.

Remark 5.17. Note that the asymptotics in Theorem 5.14 do not depend on the choice of weight ψ; any weight gives the same asymptotic fringe tree distribution. Of course, this is an immediate consequence of the proof using (5.26) and (5.25), see also Remark 5.16. Note that for this proof, it is essential that we consider convergence almost surely (and not, e.g., in probability).

Remark 5.18. In cases when $|T_n|$ is random, it is often of interest to study the number $n_{\mathcal{P}}(T_n)$ rather than the fraction $n_{\mathcal{P}}(T_n)/|T_n|$ in (5.23). Assuming (A6ψ), we can combine (5.23) and (5.16) and obtain

$$\frac{n_{\mathcal{P}}(T_n)}{n} \xrightarrow{\text{a.s.}} \frac{p_{\mathcal{P}}(\overline{\mathcal{T}})}{m_\psi}. \tag{5.27}$$

Remark 5.19. Since the fractions in (5.22) and (5.23) are bounded by 1, the a.s. convergence above immediately yields also convergence of the expectation by the dominated convergence theorem; thus

$$\mathbb{E}\,\frac{n_{\mathcal{P}}(T_n)}{|T_n|} \to p_{\mathcal{P}}(\overline{\mathcal{T}}). \tag{5.28}$$

This is particularly nice in the common case when the weight $\psi(t) = 1$, so $|T_n| = n$ deterministically; then (5.28) can be written

$$\mathbb{E}\, n_{\mathcal{P}}(T_n) = p_{\mathcal{P}}(\overline{\mathcal{T}})n + o(n). \tag{5.29}$$

For other weights ψ, we can (assuming (A6ψ)) use (5.27). If we furthermore have a deterministic bound $|T_n| \leqslant Cn$ for some constant C (which, for example, is the case for the m-ary search trees in Sections 7.1 and 7.2), then dominated convergence applies again and yields

$$\mathbb{E}\, n_{\mathcal{P}}(T_n) = \frac{p_{\mathcal{P}}(\overline{\mathcal{T}})}{m_\psi} n + o(n). \tag{5.30}$$

We give a simple but important corollary to Theorem 5.14, showing that the degree distribution in T_n converges to the distribution of $D := \Xi([0,\overline{\tau}])$ (with Ξ and $\overline{\tau}$ independent).

Corollary 5.20. *Let $n_k(T_n)$ be the number of nodes in T_n with outdegree k. Under the assumptions (A1)–(A5) above,*

$$\frac{n_k(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(D = k), \tag{5.31}$$

where $D = \Xi([0,\overline{\tau}])$ is the degree of the root of $\overline{\mathcal{T}}$. In other words, if D_n denotes the outdegree of a uniformly random node in T_n, then

$$\mathcal{L}(D_n \mid T_n) \xrightarrow{\text{a.s.}} \mathcal{L}(D). \tag{5.32}$$

Proof. Let $\mathcal{P}$ be the property of a node that it has outdegree k. Then $n_k(T) = n_{\mathcal{P}}(T)$. Hence, (5.23) shows that $n_k(T_n)/|T_n|$ a.s. converges to the probability that the root of $\overline{\mathcal{T}}$ has (out)degree k. However, the root of $\mathcal{T}_t$ has degree $\Xi([0,t])$, so the degree D of the root of $\overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$ equals $\Xi([0,\overline{\tau}])$ and (5.31) follows. □

Remark 5.21. Since $D = \Xi([0,\overline{\tau}])$ and $\overline{\tau} \sim \operatorname{Exp}(\alpha)$,

$$\mathbb{P}(D \geqslant i) = \mathbb{P}(\overline{\tau} \geqslant \xi_i) = \mathbb{E}\, e^{-\alpha\xi_i}. \tag{5.33}$$

See further Remark 5.23 below.

In order to extend Theorem 5.14 to the extended fringe, we first define the limiting random sin-tree $\widetilde{\mathcal{T}}$; this is the family tree of the *doubly infinite pedigree process* in [95] (*doubly infinite stable population process* in [65]). In the latter, we start with an individual o ("ego") born at time 0, and grow a branching process

starting with it as usual. We also give o an infinite line of ancestors $o^{(1)}, o^{(2)}, \dots$ having a modified distribution of their life histories defined below, and let each child x of each ancestor $o^{(k)}$, except $x = o^{(k-1)}$, start a new branching process where all individuals have the original distribution. We denote the (infinite) family tree of this branching process by $\widetilde{\mathcal{T}}_t$, $-\infty < t < \infty$. Finally, we stop the entire process at a random time $\overline{\tau} \sim \mathrm{Exp}(\alpha)$ as before, and let $\widetilde{\mathcal{T}} := \widetilde{\mathcal{T}}_{\overline{\tau}}$ be the resulting sin-tree, with distinguished node o. (Note that the subtree of $\widetilde{\mathcal{T}}$ rooted at o equals $\overline{\mathcal{T}}$ defined in Theorem 5.14.)

It remains to define the distribution of the life history of an ancestor. This is really a distribution of a life history with a distinguished child, which we call the *heir*. The heir may be any child, but the probability distribution is weighted by $e^{-\alpha\tau}$, where τ is the time the heir is born. Thus, recalling that the children are born at times $(\xi_i)_{i=1}^N$, for any event $\mathcal{E}$ in the life history,

$$\mathbb{P}(\mathcal{E}, \text{ and the heir is the } i\text{:th child}) = \int_{\mathcal{E}} e^{-\alpha\xi_i} dP, \tag{5.34}$$

where for $i > N$ we define $\xi_i = \infty$, so $e^{-\alpha\xi_i} = 0$. In particular,

$$q_i := \mathbb{P}(\text{the heir is the } i\text{:th child}) = \mathbb{E}\, e^{-\alpha\xi_i}. \tag{5.35}$$

Note that (5.34) defines a probability distribution, since the total probability equals

$$\sum_{i=1}^{\infty} \int_\Omega e^{-\alpha\xi_i} dP = \mathbb{E} \sum_{i=1}^{N} e^{-\alpha\xi_i} = 1 \tag{5.36}$$

by (5.8).

We may give the children of the ancestor another order as in Remark 5.1, still using (5.34). Note that then (5.34)–(5.35) hold also if we consider the i:th child in the final order and redefine ξ_i as the birth time of that child; this is seen by summing over all children and combinations of marks ν_j that put a certain child in place i at a given time.

The ancestors $o^{(k)}$ are given independent copies of this modified life history distribution, and are put together so that the heir of $o^{(k)}$ is $o^{(k-1)}$ (with $o^{(0)} = o$); this also defines recursively the birth times of all $o^{(k)}$.

Remark 5.22. Let $\widetilde{\xi}^*$ denote the age of an ancestor when its heir is born. Then $\widetilde{\xi}^*$ has by (5.34) the distribution $e^{-\alpha t}\mu(\mathrm{d}t)$, i.e., the distribution in Remark 5.6. Its Laplace transform is given by

$$\mathbb{E}\, e^{-s\widetilde{\xi}^*} = \int_0^\infty e^{-st-\alpha t}\mu(\mathrm{d}t) = \widehat{\mu}(\alpha+s) = \mathbb{E}\,\widehat{\Xi}(\alpha+s), \tag{5.37}$$

cf. (5.34) and (5.6)–(5.7). Assumption (A5) thus says that $\mathbb{E}\, e^{\varepsilon\widetilde{\xi}^*} < \infty$ for some $\varepsilon > 0$. In particular, $\widetilde{\xi}^*$ has a finite expectation

$$\beta := \mathbb{E}\,\widetilde{\xi}^* = \int_0^\infty t e^{-\alpha t}\mu(\mathrm{d}t) < \infty. \tag{5.38}$$

By (5.37), we also have the formula

$$\beta = \mathbb{E}\,\widetilde{\xi}^* = -\frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\,e^{-s\widetilde{\xi}^*}\,\Big|_{s=0} = -\widehat{\mu}'(\alpha), \tag{5.39}$$

and directly from (5.34), or by (5.38),

$$\beta = \mathbb{E}\sum_{i=1}^{N}\xi_i e^{-\alpha\xi_i}. \tag{5.40}$$

Remark 5.23. Let, as in (5.35), q_i be the probability that the heir in the ancestor distribution is child i (in birth order), and let $D = \Xi([0,\overline{\tau}])$ be the degree of the root in $\overline{\mathcal{T}}$, which by Corollary 5.20 is the limit in distribution of the outdegree of a random node in T_n. By (5.35) and (5.33),

$$q_i = \mathbb{E}\,e^{-\alpha\xi_i} = \mathbb{P}(D \geqslant i), \tag{5.41}$$

so the two distributions are closely related. Note also that (5.41) implies

$$\mathbb{E}\,D = \sum_{i=1}^{\infty}\mathbb{P}(D \geqslant i) = \sum_{i=1}^{\infty} q_i = 1, \tag{5.42}$$

so the average asymptotic outdegree is always 1. This should not be surprising; it just is an asymptotic version of the fact that in a tree with n nodes, there are together $n-1$ children, and thus the average outdegree is $1-1/n$; see also [1, Lemma 1].

Remark 5.24. Recall that we may regard the node set of $\overline{\mathcal{T}}$ as a subset of V_∞, the node set of the infinite Ulam–Harris tree. Let $v \in V_\infty$. By the recursive definition of the branching process $\mathcal{T}_t$ and the memoryless property of the exponential random variable $\overline{\tau}$, it follows that conditioned on $v \in \overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$, the subtree of $\overline{\mathcal{T}}$ rooted at v has the same distribution as $\overline{\mathcal{T}}$. In particular, conditioned on $v \in \overline{\mathcal{T}}$, the outdegree of v has the same distribution as D.

It follows from this and (5.42), by induction, that for every $k \geqslant 0$, the expected number of nodes in the k:th generation of $\overline{\mathcal{T}}$ is 1. In particular, the expected size $\mathbb{E}\,|\overline{\mathcal{T}}| = \infty$.

Note also that the outdegrees of two different nodes are not independent, since they both depend on the common stopping time $\overline{\tau}$; it is easy too see that for any $v, w \in V_\infty$, conditioned on $v, w \in \overline{\mathcal{T}}$, the outdegrees $\deg(v)$ and $\deg(w)$ are (strictly) positively correlated.

In fact, the properties in this remark except the last one hold for any fringe distribution in the sense of Aldous [1], see [1, Section 2.1]. However, the positive correlation of node degrees is not general; in particular, it makes the asymptotic fringe trees $\overline{\mathcal{T}}$ studied in this paper different from the ones obtained from conditioned Galton–Watson trees, since the latter are just unconditioned Galton–Watson trees, where all outdegrees are independent, see [1].

Theorem 5.25 (Jagers, Nerman, Aldous)**.** *Under the assumptions (A1)–(A5), as $n \to \infty$, $h(v) \xrightarrow{\mathrm{p}} \infty$ for a random node $v \in T_n$ and thus each $T_n^{*,-k}$ is well-defined w.h.p.; moreover, the following hold:*

(i) *(Annealed version.) The extended random fringe tree of T_n converges in distribution as $n \to \infty$ to the random sin-tree $\widetilde{\mathcal{T}}$ defined above, in the sense (4.14), see Section 4.1.*

(ii) *(Quenched version.) The convergence in (i) holds also conditioned on T_n, a.s. Equivalently, for every property $\mathcal{P}$ of nodes v that depends only on v, its descendants and the descendants of its ancestors at most a fixed number of generations back, we have if v is a uniformly random node in T_n,*

$$\mathbb{P}(v \text{ has } \mathcal{P} \mid T_n) = \frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\widetilde{\mathcal{T}}). \tag{5.43}$$

Proof. Again, this is a special case of the main results in Jagers and Nerman [65] and [95], and is at least implicit in Aldous [1], but we give the proof for completeness.

First consider the case of ordered trees (possibly with labels) with the children taken in order of birth. Fix a finite tree T with a distinguished node of depth $k \geqslant 0$, and let $v_0 \cdots v_k$ be the path in T from the root to the distinguished node; also, let v_i be the j_i:th child of v_{i-1}. Let $\mathcal{P} = \mathcal{P}(T)$ be the property of a node v that it has depth at least k and that, if w is its k:th ancestor, the subtree T^w, with v as distinguished node, is isomorphic to T. Then $n_{\mathcal{P}}(T_n)$, the number of $v \in T_n$ that have this property, equals the number of $w \in T_n$ such that $T_n^w \approx T$, i.e., $n_T(T_n)$. Thus, by Theorem 5.14,

$$\frac{n_{\mathcal{P}(T)}(T_n)}{|T_n|} = \frac{n_T(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(\overline{\mathcal{T}} \approx T) = \mathbb{P}(\mathcal{T}_{\overline{\tau}} \approx T). \tag{5.44}$$

Construct $\overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$ as above, and let $\xi_{(i)} = \xi_{v_i,j_i}$ be the age of v_i when its distinguished child v_{i+1} is born. The distinguished node v_k is thus born at time $V = \sum_{i=0}^{k-1} \xi_{(i)}$. If $\mathcal{T}_{\overline{\tau}} \approx T$, then necessarily $\overline{\tau} \geqslant V$. Moreover, conditioned on $\overline{\tau} \geqslant V$, $\overline{\tau}$ has the same distribution as $V + \tau'$ with $\tau' \sim \mathrm{Exp}(\alpha)$ and independent of everything else. Thus, by conditioning on V,

$$\mathbb{P}(\mathcal{T}_{\overline{\tau}} \approx T) = \mathbb{E}\bigl(e^{-\alpha V} \mathbf{1}\{\mathcal{T}_{V+\tau'} \approx T\}\bigr) = \mathbb{E}\Bigl(\mathbf{1}\{\mathcal{T}_{V+\tau'} \approx T\} \prod_{i=0}^{k-1} e^{-\alpha \xi_{(i)}}\Bigr), \tag{5.45}$$

By shifting the time parameter in $\mathcal{T}$ by V, so that the distinguished node v_k becomes born at time 0, and recalling that the subtree $\widetilde{\mathcal{T}}^{o^{(k)}}$ has the modified distribution (5.34) for the ancestors of the distinguished node, we see that (5.45) equals

$$\mathbb{P}\bigl(\widetilde{\mathcal{T}}^{o^{(k)}} \approx T\bigr) = p_{\mathcal{P}(T)}(\widetilde{\mathcal{T}}). \tag{5.46}$$

Consequently, (5.44)–(5.46) show that

$$\frac{n_{\mathcal{P}(T)}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}\bigl(\widetilde{\mathcal{T}}^{o^{(k)}} \approx T\bigr) = p_{\mathcal{P}(T)}(\widetilde{\mathcal{T}}) \tag{5.47}$$

for every finite tree with a distinguished node of depth k.

More generally, for any fixed $k \geqslant 0$ and any set A of finite trees, each having a distinguished node of depth k, let $\mathcal{P} = \mathcal{P}(A) := \bigcup_{T\in A} \mathcal{P}(T)$ be the property of a node v that it has depth at least k and that $T^w \in A$, where w is its k:th ancestor. Then as in (5.44), by Theorem 5.14 applied to the property $T^w \in A$, and using again (5.45)–(5.46),

$$\begin{aligned}\frac{n_{\mathcal{P}(A)}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(\mathcal{T}_\infty \in A) &= \sum_{T\in A} \mathbb{P}(\mathcal{T}_\infty \approx T) = \sum_{T\in A} \mathbb{P}\bigl(\widetilde{\mathcal{T}}^{o^{(k)}} \approx T\bigr) \\ &= \mathbb{P}\bigl(\widetilde{\mathcal{T}}^{o^{(k)}} \in A\bigr) = p_{\mathcal{P}(A)}(\widetilde{\mathcal{T}}).\end{aligned} \tag{5.48}$$

In particular, taking A to be the set of all finite trees, $\mathcal{P}(A)$ is the property that $h(v) \geqslant k$ and $p_{\mathcal{P}(A)}(\widetilde{\mathcal{T}}) = 1$, so (5.48) shows that for any k, $\mathbb{P}(h(v) \geqslant k) \to 1$ for a random node v in T_n. Since k is arbitrary, thus $h(v) \xrightarrow{\text{p}} \infty$. Moreover, every property $\mathcal{P}$ in (5.43) is of the form $\mathcal{P}(A)$ for some k and A, and thus the result (5.43) follows.

As in Theorem 5.14, the annealed case follows from the quenched case by taking expectations.

The case of unordered trees follows by ignoring the order.

Finally, if $\mathcal{T}_t$ is an ordered tree with the order of children defined by marks ν_i as in Remark 5.1, we first fix an integer M and consider $\mathcal{T}_t$ and $\widetilde{\mathcal{T}}$ ordered by birth order and with each node labelled with the sequence of marks $\boldsymbol{\nu}_M := (\nu_i)_{i=1}^{M\wedge N}$ (in addition to existing labels, if any). (We use a cut-off M in order to keep the space of labels countable.) We have just shown that (5.48) holds for any set A of ordered trees with such a label $\boldsymbol{\nu}_M$ on each node. Since the birth order and the marks define the true order in the trees, it follows immediately that (5.47) holds also with the true order in T_n and $\widetilde{\mathcal{T}}$, for any tree T with such marks and with maximum degree at most M. Since M is arbitrary, it holds with the true order for any T, and we may then forget the marks. (For an m-ary tree, we keep the marks.) Then (5.48) and (5.43) follow as above. □

5.1. An extension to some more general properties

In Theorem 5.25, we consider only properties of a node v that depend only on v, its ancestors at most a fixed number of generations back, and their descendants. (Theorem 5.14 is even more restrictive.) In this subsection, we show how this result can be extended to some properties that depend on all ancestors of v. A typical example is the property that v has no ancestor with outdegree 1; we consider this and some related examples in Section 11. (This section can be omitted at the first reading.)

Theorem 5.26. *Let $\mathcal{P}_0$ and $\mathcal{Q}$ be two properties of a node v in a tree, such that both $\mathcal{P}_0$ and $\mathcal{Q}$ depend only on v and its descendants. Let $\mathcal{P}$ be the property of a node v that v satisfies $\mathcal{P}_0$ but no ancestor of v satisfies $\mathcal{Q}$. Suppose, in addition to (A1)–(A5), that*

$$\mathbb{E}\,\widehat{\Xi}(\alpha)^2 < \infty \tag{5.49}$$

and that if $\Lambda := \sup\{t : \mathcal{T}_t \in \mathcal{Q}\}$, *then*

$$\mathbb{E}\, e^{\delta\Lambda} < \infty \tag{5.50}$$

for some $\delta > 0$.

Then, as $n \to \infty$, *if* v *is a uniformly random node in* T_n,

$$\mathbb{P}(v \text{ has } \mathcal{P} \mid T_n) = \frac{n_{\mathcal{P}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\widetilde{\mathcal{T}}). \tag{5.51}$$

In other words, (5.43) holds also for properties $\mathcal{P}$ of this type, although they are not covered by Theorem 5.25.

Remark 5.27. The assumption (5.50) entails $\Lambda < \infty$ a.s.; moreover, (5.50) is equivalent to assuming $\mathbb{P}(\Lambda \geqslant t) = O\bigl(e^{-\delta t}\bigr)$ for some (possibly different) $\delta > 0$, and thus to

$$\mathbb{P}\bigl(\mathcal{T}_u \in \mathcal{Q} \text{ for some } u \geqslant t\bigr) = O\bigl(e^{-\delta t}\bigr). \tag{5.52}$$

In the examples in Section 11, the property $\mathcal{Q}$ is (or can be taken as) decreasing in the sense that if it holds for some rooted tree T, then it holds also for every subtree with the same root; hence if $\mathcal{Q}$ holds for $\mathcal{T}_u$ with $u \geqslant t$, then it holds for $\mathcal{T}_t$, so (5.52) can be simplified to

$$\mathbb{P}\bigl(\mathcal{T}_t \in \mathcal{Q}\bigr) = O\bigl(e^{-\delta t}\bigr). \tag{5.53}$$

Before the proof, we give a lemma.

Lemma 5.28. *Suppose that* (A1)–(A5) *and* (5.49) *hold. Let* $\mathcal{Q}$ *be a property of rooted trees such that* (5.52) *holds for some* $\delta > 0$. *Then there exists* $\eta > 0$ *and* $a < \infty$ *such that*

$$\frac{1}{|\mathcal{T}_t|} \sum_{v \in \mathcal{T}_t} |\mathcal{T}_t^v|^{1+\eta} \mathbf{1}\{\mathcal{T}_t^v \in \mathcal{Q}\} \xrightarrow{\text{a.s.}} a. \tag{5.54}$$

Proof. The left-hand side of (5.54) equals $Z_t^{\phi}/|\mathcal{T}_t|$, where $\phi(t)$ is the characteristic given as in Remark 5.10 with

$$\phi_0(t) := |\mathcal{T}_t|^{1+\eta} \mathbf{1}\{\mathcal{T}_t \in \mathcal{Q}\}. \tag{5.55}$$

The result thus follows from (5.5), provided we can choose $\eta > 0$ such that (A6) holds for this ϕ.

To verify (A6), we first note that (5.49) (together with the other conditions) implies

$$\mathbb{E}\,|\mathcal{T}_t|^2 \leqslant C_1 e^{2\alpha t}, \tag{5.56}$$

for some $C_1 < \infty$, see [64, Theorem 6.4.3 with Note or Theorem 6.8.1] or [65, Theorem 3.5].

If $0<\eta<1$, we thus have by (5.55), Hölder's inequality, (5.56) and (5.52),

$$\begin{aligned}
\mathbb{E}\sup_{t\geqslant 0}\bigl(e^{-\theta t}\phi_0(t)\bigr) &\leqslant \sum_{n=0}^{\infty}\mathbb{E}\sup_{n\leqslant t\leqslant n+1}\bigl(e^{-\theta t}\phi_0(t)\bigr)\\
&\leqslant \sum_{n=0}^{\infty} e^{-\theta n}\,\mathbb{E}\Bigl(|\mathcal{T}_{n+1}|^{1+\eta}\mathbf{1}\{\mathcal{T}_u\in\mathcal{Q}\text{ for some } u\geqslant n\}\Bigr)\\
&\leqslant \sum_{n=0}^{\infty} e^{-\theta n}\bigl(\mathbb{E}\,|\mathcal{T}_{n+1}|^2\bigr)^{(1+\eta)/2}\bigl(\mathbb{P}\bigl(\mathcal{T}_u\in\mathcal{Q}\text{ for some } u\geqslant n\bigr)\bigr)^{(1-\eta)/2} \qquad (5.57)\\
&\leqslant \sum_{n=0}^{\infty} e^{-\theta n}\bigl(C_1e^{2\alpha(n+1)}\bigr)^{(1+\eta)/2}\bigl(C_2e^{-\delta n}\bigr)^{(1-\eta)/2}\\
&= C_3\sum_{n=0}^{\infty} e^{n(-\theta+\alpha(1+\eta)-\delta(1-\eta)/2)},
\end{aligned}$$

which is finite provided

$$(1+\eta)\alpha<\theta+\delta(1-\eta)/2. \qquad (5.58)$$

If $0<\eta<\min(\delta/4\alpha,1/2)$, then $\eta\alpha<\delta/4<\delta(1-\eta)/2$, and thus there exists $\theta<\alpha$ such that (5.58) holds; hence the sum in (5.57) is finite and (A6) holds, which completes the proof. □

Proof of Theorem 5.26. For each integer M, let $\mathcal{P}_M$ be the truncated property "v satisfies $\mathcal{P}_0$ but no ancestor at most M generations before v satisfies $\mathcal{Q}$." Then $\mathcal{P}_M$ is covered by Theorem 5.25, so

$$\frac{n_{\mathcal{P}_M}(T_n)}{|T_n|}\xrightarrow{\text{a.s.}} p_{\mathcal{P}_M}(\widetilde{\mathcal{T}}) \qquad (5.59)$$

as $n\to\infty$, for each M. Since $\mathcal{P}$ is the intersection of the decreasing sequence of properties $\mathcal{P}_M$, it is clear that $p_{\mathcal{P}_M}(\widetilde{\mathcal{T}})\to p_{\mathcal{P}}(\widetilde{\mathcal{T}})$ as $M\to\infty$. Furthermore, $n_{\mathcal{P}}(\mathcal{T}_t)\leqslant n_{\mathcal{P}_M}(\mathcal{T}_t)$ and for any $\eta>0$, writing $w\prec v$ when w is an ancestor of v,

$$\begin{aligned}
n_{\mathcal{P}_M}(\mathcal{T}_t)-n_{\mathcal{P}}(\mathcal{T}_t) &\leqslant \sum_{v\in\mathcal{T}_t}\sum_{w\in\mathcal{T}_t}\mathbf{1}\{w\prec v, h(w)<h(v)-M\text{ and }\mathcal{T}_t^w\in\mathcal{Q}\}\\
&\leqslant \sum_{w\in\mathcal{T}_t}\sum_{v\succ w}\mathbf{1}\{|\mathcal{T}_t^w|\geqslant M\text{ and }\mathcal{T}_t^w\in\mathcal{Q}\}\\
&= \sum_{w\in\mathcal{T}_t}(|\mathcal{T}_t^w|-1)\mathbf{1}\{|\mathcal{T}_t^w|\geqslant M\text{ and }\mathcal{T}_t^w\in\mathcal{Q}\}\\
&\leqslant M^{-\eta}\sum_{w\in\mathcal{T}_t}|\mathcal{T}_t^w|^{1+\eta}\mathbf{1}\{\mathcal{T}_t^w\in\mathcal{Q}\}. \qquad (5.60)
\end{aligned}$$

By Lemma 5.28 we can choose $\eta>0$ and a such that (5.54) holds. Then for any M, by (5.60) and (5.54), a.s.,

$$\limsup_{n\to\infty}\frac{|n_{\mathcal{P}_M}(T_n)-n_{\mathcal{P}}(T_n)|}{|T_n|}\leqslant\limsup_{t\to\infty}\frac{|n_{\mathcal{P}_M}(\mathcal{T}_t)-n_{\mathcal{P}}(\mathcal{T}_t)|}{|\mathcal{T}_t|}\leqslant M^{-\eta}a. \qquad (5.61)$$

Consequently, using also (5.59), a.s.,

$$\begin{aligned}\limsup_{n\to\infty}\left|\frac{n_{\mathcal{P}}(T_n)}{|T_n|}-p_{\mathcal{P}}(\widetilde{\mathcal{T}})\right| &\leqslant \limsup_{n\to\infty}\frac{\bigl|n_{\mathcal{P}_M}(T_n)-n_{\mathcal{P}}(T_n)\bigr|}{|T_n|}\\ &\quad+\limsup_{n\to\infty}\left|\frac{n_{\mathcal{P}_M}(T_n)}{|T_n|}-p_{\mathcal{P}_M}(\widetilde{\mathcal{T}})\right|+\bigl|p_{\mathcal{P}_M}(\widetilde{\mathcal{T}})-p_{\mathcal{P}}(\widetilde{\mathcal{T}})\bigr|\\ &\leqslant M^{-\eta}a+0+\bigl|p_{\mathcal{P}_M}(\widetilde{\mathcal{T}})-p_{\mathcal{P}}(\widetilde{\mathcal{T}})\bigr|. \qquad (5.62)\end{aligned}$$

The right-hand side tends to 0 as $M\to\infty$, and the theorem follows. □

6. Examples with uniform or preferential attachment

We begin with a few standard examples, where we repeat earlier results by other authors, together with some new results on the limiting sin-trees. In all examples in this section, $|T_n|=n$, so we stop the branching process using the weight $\psi(t)=1$ as in Example 5.3. Since this weight is bounded, (A6ψ) holds trivially.

Example 6.1 (random recursive tree)**.** An important example, considered already by Aldous [1], is the *random recursive tree*. This tree, usually considered as an unordered rooted tree, is constructed recursively by adding nodes one by one, with each new node attached as a child of a (uniformly) randomly chosen existing node, see [40, Section 1.3.1]. It is easy to see, by the memoryless property of the exponential distribution, that the random recursive tree with n nodes is the tree T_n defined in Section 5 for the branching process where each individual gives birth with constant intensity 1, i.e. with independent Exp(1) waiting times between births, and weight function $\psi(t)=1$ as in Example 5.3. In other words, the point process Ξ describing the births of the children of an individual is a Poisson process with intensity 1. This branching process (or just the sizes $(|\mathcal{T}_t|)_{t\geqslant0}$) is often called the *Yule process*, so the process $(\mathcal{T}_t)_t$ of trees is called the *Yule tree process* [1]. Note that Yule process formed by the size $|\mathcal{T}_t|$ is a pure birth process where the birth rate $\lambda_n=n$, see Example A.3.

We will need some notation. Let $X_i:=\xi_i-\xi_{i-1}$ (with $\xi_0:=0$) be the waiting times between the births of the children of a given individual. Thus X_i are i.i.d. Exp(1), and $\xi_i=\sum_{j=1}^i X_j\sim\Gamma(i,1)$ has a Gamma distribution.

The intensity measure μ is Lebesgue measure on $[0,\infty)$, so

$$\widehat{\mu}(\theta)=\int_0^\infty e^{-\theta t}\,\mathrm{d}t=\frac{1}{\theta},\qquad \theta>0, \qquad (6.1)$$

and (5.4) holds with $\alpha=1$. The conditions (A1)–(A5) are trivially verified.

As shown by Aldous [1], the limiting fringe tree $\overline{\mathcal{T}}=\mathcal{T}_{\overline{\tau}}$ can also be described as a random recursive tree with a random number M nodes, where

$$\mathbb{P}(M=n)=\frac{1}{n(n+1)},\qquad n\geqslant1. \qquad (6.2)$$

In fact, by symmetry, if $M = |\overline{\mathcal{T}}|$ and we condition $\overline{\mathcal{T}}$ on $M = n$, we get a random recursive tree on n nodes. Moreover, if we at some time have $n \geqslant 1$ individuals in the branching process, then a new child is born with intensity n, while the process stops (at $\overline{\tau}$) with intensity 1, so the probability that the process continues with at least one more individual is $n/(n+1)$. In other words, $\mathbb{P}(M \geqslant n+1) = \frac{n}{n+1}\,\mathbb{P}(M \geqslant n)$, and thus by induction $\mathbb{P}(M \geqslant n) = 1/n$ and (6.2) follows. (For an alternative argument, see Example 6.4 below.)

As noted by [1], various results for the random recursive tree T_n now follows from Theorem 5.14. For example, the asymptotic distribution of the size of a random fringe tree is given by (6.2). Furthermore, the asymptotic distribution of the outdegree of the nodes in T_n equals by Corollary 5.20 the distribution of the root degree D in $\overline{\mathcal{T}}$, which is geometric $\mathrm{Ge}_0(1/2)$ as an immediate consequence of (5.33). (See (6.3) below and (5.41).) See Section 10.2 for yet another example.

In order to construct the random sin-tree $\widetilde{\mathcal{T}}$, which enables applications of Theorem 5.25 on the extended fringe, we have to find the distribution of the life history of the ancestors, given by (5.34). Consider an ancestor and denote its successive birth times by $\widetilde{\xi}_i$, $i \geqslant 1$, and let $\widetilde{X}_i := \widetilde{\xi}_i - \widetilde{\xi}_{i-1}$ (with $\widetilde{\xi}_0 := 0$) be the succesive waiting times. Furthermore, let $\widetilde{\Xi}$ be the point process of all births of children of this ancestor (thus $\widetilde{\Xi} = \sum_i \delta_{\widetilde{\xi}_i}$) and let J be the number of the heir (in birth order). Then, by (5.35),

$$\mathbb{P}(J = j) = \mathbb{E}\, e^{-\xi_j} = \mathbb{E}\, e^{-\sum_1^j X_i} = \prod_{i=1}^{j} \mathbb{E}\, e^{-X_i} = 2^{-j}, \qquad j \geqslant 1. \tag{6.3}$$

Thus J has the (shifted) geometric distribution $\mathrm{Ge}_1(1/2)$. Moreover, conditioned on $J = j$, the joint density of $(\widetilde{X}_1, \dots, \widetilde{X}_m)$, for any $m \geqslant j$, is by (5.34)

$$\frac{1}{P(J=j)} e^{-\sum_1^j x_i} \prod_{i=1}^{m} e^{-x_i} = \prod_{i=1}^{j} 2e^{-2x_i} \prod_{i=j+1}^{m} e^{-x_i}. \tag{6.4}$$

Consequently, conditioned on $J = j$, the waiting times $\widetilde{X}_i$ between the births for an ancestor are independent, with $\widetilde{X}_i \sim \mathrm{Exp}(2)$ for $i \leqslant J$ and $\widetilde{X}_i \sim \mathrm{Exp}(1)$ for $i > J$.

We claim that we can describe $\widetilde{\Xi}$ in a simpler way as a Poisson process Ξ with intensity 1, plus an extra point $Z \sim \mathrm{Exp}(1)$, independent of Ξ, with Z the heir. To see this, note that with this description, the first point of $\widetilde{\Xi}$ is either the first point of Ξ or Z; these two first points are both $\mathrm{Exp}(1)$ and independent, so the first point $\widetilde{X}_1$, which is the smallest of these two points, will be $\mathrm{Exp}(2)$. Furthermore, with probability $1/2$, this point is the heir Z, so $J = 1$, and then the rest of the process is Ξ, with independent $\mathrm{Exp}(1)$ waiting times. And with probability $1/2$, $\widetilde{X}_1$ comes from Ξ, and then the whole process repeats from $\widetilde{X}_1$, so the next waiting time $\widetilde{X}_2 \sim \mathrm{Exp}(2)$, and so on. A simple induction shows that this yields both the distribution of J in (6.3) and the right conditional distribution of $(\widetilde{X}_i)_1^\infty$ given $J = j$ for each j, which proves the claim.

In particular, an ancestor's age $\widetilde{\xi}^*$ when its heir is born has distribution Exp(1), which also follows directly from Remark 5.22. As a consequence, $\beta = \mathbb{E}\,\widetilde{\xi}^* = 1$, cf. (5.38) and (5.39).

We can thus describe the random sin-tree $\widetilde{\mathcal{T}}$ as follows: First construct an infinite chain of ancestors $o^{(1)}, o^{(2)}, \dots$ of o (backwards in time), with the times between their births i.i.d. Exp(1); in other word, $(o^{(k)})_{k\geqslant 1}$ are born according to a Poisson process with intensity 1 on $(-\infty, 0)$. Then grow independent Yule tree processes from all $o^{(k)}$, $k \geqslant 0$. Finally, stop everything at $\overline{\tau} \sim \mathrm{Exp}(1)$. (Cf. Aldous [1, Section 4], where the description is less explicit.) For an application, see Theorem 11.6.

Example 6.2 (binary search tree)**.** Another important example studied by Aldous [1] is the (random) *binary search tree*. This is the case $m = 2$ of the m-ary search tree in Section 3, but it is simpler than the general case, so we treat it separately, using a slightly different but equivalent formulation. (Since each (internal) node has exactly one key, the number of keys equals the number of nodes, and we can ignore the keys completely.)

The binary search tree can be grown recursively as follows. (See e.g. [40] for other, equivalent, constructions.) Start with a single node. Since we grow a binary tree, each node may have a left child and a right child. When the tree has n nodes, there are $n+1$ empty places for children (these places are the external nodes in the description in Section 3). The tree grows by adding a node to one of these $n+1$ places, chosen uniformly at random. Similarly as in Example 6.1, it is easy to see that the binary search tree is the tree T_n produced by the branching process where each individual has two children, labelled left and right and born at age ξ_{L} and ξ_{R}, say, with ξ_{L} and ξ_{R} both Exp(1) and independent; furthermore we use again the weight function $\psi(t) = 1$ as in Example 5.3. (This continuous-time branching process seems to have been first used to study the binary search tree by Pittel [101], who considered the height and saturation level, see Section 13.)

We thus have $N = 2$. Since each child is born with the density function e^{-x}, the intensity measure μ of Ξ has density $2e^{-x}$. Thus

$$\widehat{\mu}(\theta) = \int_0^\infty 2e^{-\theta x - x}\,\mathrm{d}x = \frac{2}{1+\theta}, \qquad \theta > -1, \tag{6.5}$$

and (5.4) holds with the Malthusian parameter $\alpha = 1$. The conditions (A1)–(A5) are trivially verified.

Note that if we order the children in order of birth as usual, then $\xi_1 = \min(\xi_{\mathsf{L}}, \xi_{\mathsf{R}})$, and thus $\xi_1 \sim \mathrm{Exp}(2)$, while the waiting time $\xi_2 - \xi_1$ for the second child is Exp(1) and independent of ξ_1.

We see also that the size $|\mathcal{T}_t|$ grows as a pure birth process with birth rate $\lambda_n = n+1$, see Appendix A. Equivalently, $|\mathcal{T}_t| + 1$, which can be interpreted as the number of external nodes, is a pure birth process with rate $\lambda_n = n$, i.e., the Yule process in Example A.3, and in Example 6.1, but started at 2 instead of 1.

As shown by Aldous [1], the limiting fringe tree $\overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$ can be described as

a binary search tree with a random number M nodes, where

$$\mathbb{P}(M=n)=\frac{2}{(n+1)(n+2)},\qquad n\geqslant 1; \tag{6.6}$$

cf. the similar result (6.2) for the random recursive tree. To see this we argue as in Example 6.1; the difference is that when there are n individuals, there are now $n+1$ places to add a new node, and thus $n+1$ independent Exp(1) for these, competing with the random time $\overline{\tau}$ that stops the process; hence the probability of adding another node is $(n+1)/(n+2)$ and thus by induction $\mathbb{P}(M\geqslant n)=2/(n+1)$ and (6.6) follows. (For an alternative argument, see Example 6.4 below.)

By Theorem 5.14, the asymptotic distribution of the size of a random fringe tree is given by (6.6). Another simple calculation in [1] shows that the asymptotic distribution of the outdegree of the nodes in T_n, which by Corollary 5.20 equals the distribution of the root degree D in $\overline{\mathcal{T}}$, is uniform on $\{0,1,2\}$, see (5.33). This can also be seen without calculation: $\xi_{\mathsf{L}},\xi_{\mathsf{R}}$ and $\overline{\tau}$ are three i.i.d. Exp(1) random variables, so the three events that $\overline{\tau}$ is the smallest, the middle, or the largest of these three have by symmetry all the same probability 1/3. These events equal the events that the root in $\overline{\mathcal{T}}$ has degree $0,1,2$.

To find the random sin-tree $\widetilde{\mathcal{T}}$, note that by the comments after (5.34)–(5.35), (5.34) holds also when taking the children in order left–right. For an individual in $\overline{\mathcal{T}}$, the pair $(\xi_{\mathsf{L}},\xi_{\mathsf{R}})$ has the density function $e^{-x_{\mathsf{L}}-x_{\mathsf{R}}}$. For an ancestor, the probability that the heir is the left child is 1/2 (by symmetry or by (5.35)), and it follows that conditioned on the heir being the left child, the pair $(\xi_{\mathsf{L}},\xi_{\mathsf{R}})$ has the density function $2e^{-x_{\mathsf{L}}}e^{-x_{\mathsf{L}}-x_{\mathsf{R}}}=2e^{-2x_{\mathsf{L}}}e^{-x_{\mathsf{R}}}$. In other words, for an ancestor, given that the heir is the left child, the age $\widetilde{\xi}^*$ when the heir is born is Exp(2) and the age when the other child is born is Exp(1), and these two ages are independent. The same holds given that the heir is the right child. In particular, $\widetilde{\xi}^*\sim\mathrm{Exp}(2)$ and thus $\beta=\mathbb{E}\,\widetilde{\xi}^*=1/2$, cf. Remark 5.22.

Consequently, the random sin-tree $\widetilde{\mathcal{T}}$ can be described as follows, cf. the case of the random recursive tree in Example 6.1: First construct an infinite chain of ancestors $o^{(1)},o^{(2)},\dots$ of o (backwards in time), with the times between their births i.i.d. Exp(2); in other word, $(o^{(k)})_{k\geqslant1}$ are born according to a Poisson process with intensity 2 on $(-\infty,0)$. Moreover, make a random choice (uniform and independent of everything else) for each ancestor to decide whether its heir is the left or right child. Then grow independent binary tree processes at all empty places (external nodes), with independent Exp(1) waiting times for all new nodes. Finally, stop everything at $\overline{\tau}\sim\mathrm{Exp}(1)$. (Applications are given in Section 11.)

Example 6.3 (general preferential attachment trees)**.** We can generalise the preceding examples as follows, see Rudas, Tóth and Valkó [105] and Rudas and Tóth [104] where this example is studied using the branching process method described here; see also Bhamidi [10]. The branching process below was also earlier used by Biggins and Grey [15] to study the height of these trees. Moreover, the method has been generalized by Deijfen [28] to a class of random trees

where vertices also may die. We give only a summary and some complements, in particular on sin-trees. Some special cases are treated in Examples 6.4–6.8 below, see in particular Example 6.6; these cases have been studied by many authors, using various methods. (Further references are given below, but we do not attempt a complete history.)

Suppose that we are given a sequence of non-negative weights $(w_k)_{k=0}^{\infty}$, with $w_0 > 0$. Grow a random tree T_n (with n nodes) recursively, starting with a single node and adding nodes one by one. Each new node is added as a child of some randomly chosen existing node; when a new node is added to T_{n-1}, the probability of choosing a node $v \in T_{n-1}$ as the parent is proportional to $w_{d^+(v)}$, where $d^+(v)$ is the outdegree of v in T_{n-1}. (More formally, this is the conditional probability, given T_{n-1} and the previous history. The sequence $(T_n)_{n=1}^{\infty}$ thus constitutes a Markov process.) If we want the trees T_n to be ordered trees, we also insert the new child of v among the existing $d^+(v)$ children in a random position, uniformly chosen among the $d^+(v)+1$ possibilities.

The random recursive tree in Example 6.1 is the special case $w_k = 1$, $k \geqslant 0$, and the binary search tree in Example 6.2 is the special case with $w_0 = 2$, $w_1 = 1$ and $w_k = 0$, $k \geqslant 2$ (and, furthermore, each first child randomly assigned to be left or right).

Note that we require $w_0 > 0$ (and $w_1 > 0$ will be implicitly assumed, as a consequence of (6.12) below), but we allow $w_m = 0$ for some larger m, as in the example of the binary search tree. In this case, no individual will ever get more than m children; in fact (provided m is chosen minimal), $N = m$ a.s. In this case, the weights $w_{m+1}, w_{m+2}, \dots$ are irrelevant, so it suffices to prescribe w_k for $k \leqslant m$. (In this case, we interpret $1/w_m = \infty$ below, and the infinite sums in (6.11) become finite. We leave such obvious modifications to the reader.)

In some important examples, for example Example 6.6 below, w_k is a strictly increasing function of k, which means that nodes with a high degree are more likely to attract a new node than nodes with a low degree; hence the name *preferential attachment*, which comes from Barabási and Albert [6] where this type of model was introduced (in a more general version, in general yielding graphs and not trees), see Example 6.6. The tree version of their model had been studied earlier under a different name by Szymański [107] and others, see Example 6.5. The model with general w_k was considered by Móri [92].

As in the examples above, the tree T_n can be constructed by a branching process as in Section 5, again with weight $\psi(t) = 1$ and taking the birth times $\xi_i := \sum_{j=1}^{i} X_j$, now with the waiting times between births $X_j = \xi_j - \xi_{j-1} \sim \mathrm{Exp}(w_{j-1})$ and independent. In other words, the stochastic process $\Xi([0,t])$, $t \geqslant 0$, (i.e., the number of children of a given individual born up to age t) is a pure birth process, starting at 0 and with birth rate w_k when the state is k. (See Appendix A.)

Let $\xi_\infty := \lim_{n\to\infty} \xi_n = \sum_{j=1}^{\infty} X_j \in (0,\infty]$. Then

$$\mathbb{E}\,\xi_\infty = \sum_{j=1}^{\infty} \mathbb{E}\,X_j = \sum_{k=0}^{\infty} \frac{1}{w_k}. \tag{6.7}$$

We distinguish between two cases, depending on whether this sum is finite or not.

In the *explosive case*,

$$\mathbb{E}\,\xi_\infty = \sum_{k=0}^{\infty} \frac{1}{w_k} < \infty. \tag{6.8}$$

Thus $\xi_\infty < \infty$ a.s., i.e., an individual will have an infinite number of children in a finite time (the point process Ξ explodes). In this case, the branching process will explode in finite time, and several of the assumptions in Section 5 fail. Nevertheless, this case can be treated separately. It turns out that the random fringe tree T_n^* is asymptotically degenerate and w.h.p. consists of a single node only, i.e., $|T_n^*| \xrightarrow{\mathrm{p}} 1$, see Theorem 6.11 below. (Equivalently, the proportion of leaves in the tree T_n tends to 1.) The case $w_k = (k+1)^p$ for some $p > 1$ is studied by Krapivsky, Redner and Leyvraz [81], Krapivsky and Redner [80] and (rigorously and in detail) by Oliveira and Spencer [98], who show that if $p > 2$ (but not if $1 < p \leqslant 2$), the random tree process T_n, $n \geqslant 1$, is even more strongly degenerate: a.s. there exists a (random) n_0 and a node $v \in T_{n_0}$ such that all nodes added after time n_0 become children of v (and thus remain leaves forever). See also Athreya [3].

In the sequel we consider the *non-explosive case*

$$\sum_{k=0}^{\infty} \frac{1}{w_k} = \infty. \tag{6.9}$$

In this case, by (6.7), $\mathbb{E}\,\xi_\infty = \infty$; moreover, it is easy to see that $\xi_\infty = \lim_{n\to\infty} \xi_n = \infty$ a.s., for example by calculating

$$\mathbb{E}\,e^{-\lambda\xi_\infty} = \prod_{j=1}^{\infty} \mathbb{E}\,e^{-\lambda X_j} = \prod_{k=0}^{\infty} \frac{w_k}{\lambda + w_k} = \prod_{k=0}^{\infty} \frac{1}{1+\lambda/w_k} = 0 \tag{6.10}$$

for any $\lambda > 0$, see [3]. Hence, an individual has a.s. only a finite number of children in each finite interval, i.e., $\Xi([0,t]) < \infty$ for every $t < \infty$. Furthermore, using (5.7),

$$\widehat{\mu}(\theta) = \mathbb{E}\sum_{n=1}^{\infty} e^{-\theta\xi_n} = \sum_{n=1}^{\infty}\prod_{j=1}^{n} \mathbb{E}\,e^{-\theta X_j} = \sum_{n=1}^{\infty}\prod_{k=0}^{n-1} \frac{1}{1+\theta/w_k}. \tag{6.11}$$

We assume that there exists $\theta > 0$ such that $1 < \widehat{\mu}(\theta) < \infty$, i.e.,

$$1 < \sum_{n=0}^{\infty}\prod_{k=0}^{n} \frac{1}{1+\theta/w_k} < \infty. \tag{6.12}$$

(This is also easily seen to imply (6.9).) This implies, by dominated convergence, that $\lambda \mapsto \widehat{\mu}(\lambda)$ is continuous on $[\theta,\infty]$, with $\widehat{\mu}(\infty) = 0$, and thus there exists $\alpha > \theta$ such that $\widehat{\mu}(\alpha) = 1$. Hence the assumptions (A4) and (A5) hold. The

remaining assumptions are trivially satisfied, and thus Theorems 5.14 and 5.25 apply.

The asymptotic degree distribution is by Corollary 5.20 and (5.33) given by

$$\mathbb{P}(D \geqslant i) = \mathbb{E}\, e^{-\alpha\xi_i} = \prod_{j=1}^{i} \mathbb{E}\, e^{-\alpha X_j} = \prod_{k=0}^{i-1} \frac{w_k}{w_k+\alpha} \tag{6.13}$$

and thus

$$\begin{aligned}\mathbb{P}(D = i) &= \mathbb{P}(D \geqslant i) - \mathbb{P}(D \geqslant i+1) = \Bigl(1 - \frac{w_i}{w_i+\alpha}\Bigr)\mathbb{P}(D \geqslant i) \\ &= \frac{\alpha}{w_i+\alpha} \prod_{k=0}^{i-1} \frac{w_k}{w_k+\alpha}.\end{aligned} \tag{6.14}$$

This can also be seen as an example of Theorem A.4.

To describe the life of an ancestor, let $\mathcal{E}_i$ be the event that the heir of the ancestor is child i. We note first that if we fix $M < \infty$, then in the point process ξ, the waiting times $X_1, \dots, X_M$ have the joint density function $\prod_{j=1}^{M} w_{j-1} e^{-w_{j-1}x_j}$. It follows from (5.34) that for any i and M with $1 \leqslant i \leqslant M$, conditioned on $\mathcal{E}_i$, the waiting times $(\widetilde{X}_j)_{j=1}^{M}$ between the M first children of the ancestor have a joint density function that is proportional to

$$e^{-\alpha\sum_{j=1}^{i} x_j} \prod_{j=1}^{M} e^{-w_{j-1}x_j} = \prod_{j=1}^{i} e^{-(w_{j-1}+\alpha)x_j} \prod_{j=i+1}^{M} e^{-w_{j-1}x_j}. \tag{6.15}$$

Furthermore, by (5.41) and (6.13) (or by tracking constants in the argument just given),

$$q_i := \mathbb{P}(\mathcal{E}_i) = \prod_{k=0}^{i-1} \frac{w_k}{w_k+\alpha}, \qquad i \geqslant 1. \tag{6.16}$$

Consequently, the point process $\widetilde{\Xi}$ describing the births of the children of an ancestor can be constructed as follows: Select the number I of the heir at random, with the distribution (6.16). Then, conditioned on $I = i$, let the waiting times $\widetilde{X}_j$ be independent exponential variables, with $\widetilde{X}_j \sim \mathrm{Exp}(w_{j-1}+\alpha)$ for $j \leqslant i$ and $\widetilde{X}_j \sim \mathrm{Exp}(w_{j-1})$ for $j > i$.

The limiting random sin-tree $\widetilde{\mathcal{T}}$ then is constructed as in Section 5.

In Examples 6.1 and 6.2, we have seen alternative, simpler, constructions of $\widetilde{\Xi}$. This will be extended to the linear case in Example 6.4 and Theorem 6.9, but it does not seem possible to extend it further. In particular, we show in Theorem 6.10 that the age $\widetilde{\xi}^*$ when the heir is born has an exponential distribution only in the linear case. See also Example 6.8 for a simple non-linear example.

Example 6.4 (linear preferential attachment)**.** The simplest, and most studied, case of preferential attachment as in Example 6.3 is the linear case

$$w_k = \chi k + \rho, \tag{6.17}$$

for some real parameters χ and ρ, with $\rho = w_0 > 0$. Note that we obtain the same random trees T_n if we multiply all w_k by a positive constant. (In the branching processes, only the time scale changes.) Hence, only the quotient χ/ρ matters, and it suffices to consider $\chi \in \{1, 0, -1\}$.

The case $\chi = 0$ is the (non-preferential) random recursive tree in Example 6.1. (In this case ρ is irrelevant and we take $\rho = 1$.)

The case $\chi = 1$ (the increasing case) is studied in Example 6.6.

In the case $\chi = -1$, so $w_k = \rho - k$, w_k is eventually negative. This is impossible, and violates our basic assumption in Example 6.3. However, this is harmless if (and only if) $\rho = m$ is an integer; then $w_m = 0$ and, as said above in Example 6.3, the values w_k for $k > m$ do not matter. This is the m-ary case studied in Example 6.7; the binary search tree in Example 6.2 is the special case $\chi = -1$, $\rho = 2$.

We continue with some results valid for any linear weight (6.17), and refer to Examples 6.1, 6.6 and 6.7 for further results for the different cases $\chi = 0, 1, -1$.

Since $\Xi([0,t])$ is a pure birth process with a rate that is a linear function $\chi k + \rho$ of the current state k, and with initial value 0, it is easy to see, see Theorem A.6, that the expectation $\mathbb{E}\,\Xi([0,t]) = \mu(t)$ is given by

$$\mu(t) = \begin{cases} \frac{\rho}{\chi}\bigl(e^{\chi t} - 1\bigr), & \chi \neq 0, \\ \rho t, & \chi = 0. \end{cases} \tag{6.18}$$

Hence, μ has density $\rho e^{\chi t}$ (also when $\chi = 0$), cf. (A.15), and thus

$$\widehat{\mu}(\theta) = \int_0^\infty \rho e^{\chi t - \theta t}\,\mathrm{d}t = \frac{\rho}{\theta - \chi}, \qquad \theta > \chi. \tag{6.19}$$

It follows that (6.12) holds, and that (5.4) holds with

$$\alpha = \chi + \rho = w_1. \tag{6.20}$$

(Alternatively, (6.19) can be verified algebraically, see (6.30) and (6.41) below.)

By Remark 5.22 and (6.20), the age $\widetilde{\xi}^*$ when an heir is born to an ancestor has density $e^{-\alpha t}\mu(\mathrm{d}t) = e^{-\alpha t}\rho e^{\chi t}\,\mathrm{d}t = \rho e^{-\rho t}\,\mathrm{d}t$; thus $\widetilde{\xi}^*$ has an exponential distribution $\mathrm{Exp}(\rho)$. (This also follows from (6.19) and the formula for the Laplace transform in Remark 5.22.) As a consequence, generalizing the values of β found in Examples 6.1 and 6.2,

$$\beta = \mathbb{E}\,\widetilde{\xi}^* = \rho^{-1} = w_0^{-1}. \tag{6.21}$$

We claim that the life history $\widetilde{\Xi}$ of an ancestor can be described as follows (as a simpler alternative to the general construction in Example 6.3), cf. the special cases in Examples 6.1–6.2; we postpone the proof to Theorem 6.9 below: For an ancestor, the ordinary children are born according to a point process Ξ' which is a pure birth process, with birth rate $w_{k+1} = w_k + \chi$ when the state (number of ordinary children so far) is k, and the heir is born at an age $\widetilde{\xi}^* \sim \mathrm{Exp}(\rho)$, independent of Ξ'. Consequently, the limiting random sin-tree $\widetilde{\mathcal{T}}$

can be constructed as follows, generalising the constructions in Examples 6.1–6.2: First construct an infinite chain of ancestors $o^{(1)}, o^{(2)}, \dots$ of o (backwards in time), with the times between their births i.i.d. $\mathrm{Exp}(\rho)$; in other word, $(o^{(k)})_{k\geqslant 1}$ are born according to a Poisson process with intensity ρ on $(-\infty, 0)$. Give each ancestor additional children according to independent copies of Ξ' (where the intensities are shifted from Ξ, as said above). Then, every other individual gets children according to independent copies of Ξ. Finally, stop everything at $\overline{\tau} \sim \mathrm{Exp}(\alpha) = \mathrm{Exp}(\chi + \rho)$.

The linear case (6.17) treated in this example is simpler than the general case in Example 6.3 in several ways. For example, we have shown that the age $\widetilde{\xi}^*$ when the heir of an ancestor is born has an exponential distribution, and (as said earlier) it will be shown in Theorem 6.10 that this holds only in the linear case. An important reason (perhaps the main reason) that the linear case is simpler is that the total weight in a tree depends only on the size of the tree: if $|T| = n$, then the total weight of the nodes in T, which we may label by $1, \dots, n$, is

$$\sum_{i=1}^{n} w_{d^+(i)} = \sum_{i=1}^{n} (\chi d^+(i) + \rho) = \chi \sum_{i=1}^{n} d^+(i) + n\rho = \chi(n-1) + n\rho = n\alpha - \chi. \tag{6.22}$$

This property has several important consequences. First, it follows (as remarked for the random recursive tree and the binary search tree above) that if $M = |\overline{\mathcal{T}}|$ and we condition $\overline{\mathcal{T}}$ on $M = n$, we get the random tree T_n. (The property called *coherence* by Aldous [1, Section 2.6].) The distribution of M can be found by the same argument as for the random recursive tree in Example 6.1, which now, using (6.22) and (6.20), yields

$$\mathbb{P}(M \geqslant n) = \frac{\rho}{n\alpha - \chi} = \frac{\rho}{(n-1)\alpha + \rho} \tag{6.23}$$

and hence

$$\mathbb{P}(M = n) = \frac{\rho\alpha}{((n-1)\alpha + \rho)(n\alpha + \rho)} = \frac{\kappa}{(n + \kappa - 1)(n + \kappa)} \tag{6.24}$$

with

$$\kappa := \frac{\rho}{\alpha} = \frac{\rho}{\chi + \rho} = \frac{w_0}{w_1}. \tag{6.25}$$

Consequently, $\overline{\mathcal{T}}$ can be described as the random tree T_M with a random size M given by (6.24).

An alternative way to see (6.24) is to note that (6.22) implies that the size $Z_t = |\mathcal{T}_t|$ of the branching process is a pure birth process with birth rates $\lambda_n = n\alpha - \chi = n(\chi + \rho) - \chi$, and thus $|\mathcal{T}_t| - 1$ is a pure birth process with birth rates $\lambda_n = (n+1)\alpha - \chi = n\alpha + \rho$. (The special case $\chi = 0$, $\rho = 1$, when $|\mathcal{T}_t|$ is a Yule process, was noted in Remark 6.1.) Theorem A.5 shows that

$$|\overline{\mathcal{T}}| - 1 = |\mathcal{T}_{\overline{\tau}}| - 1 \sim \mathsf{HG}(\rho/\alpha, 1; (\rho + \alpha)/\alpha + 1) = \mathsf{HG}(\kappa, 1; \kappa + 2), \tag{6.26}$$

which by (B.3) and (B.5), or simpler by (B.10), yields (6.24). Note also that Theorem A.7 shows that the size $|\mathcal{T}_t|$ at a fixed time, minus 1, has a negative binomial distribution.

Furthermore, (6.22) implies that if we label the nodes of T_n by $1,\dots,n$ in the order they are added to the tree, so that T_n becomes an *increasing tree* (or *recursive tree* [40, Section 1.3]), then the probability that T_n equals a given ordered increasing tree T (with $|T| = n$) is, by the definition and a simple rearrangement,

$$\frac{\prod_{i=1}^{n}\bigl(\frac{1}{d^+(i)!}\prod_{k=0}^{d^+(i)-1} w_k\bigr)}{\prod_{j=1}^{n-1}(j\alpha-\chi)} \tag{6.27}$$

which is proportional to $\prod_{i=1}^{n}\phi_{d^+(i)}$ with $\phi_d := \frac{1}{d!}\prod_{k=0}^{d-1} w_k$. Hence T_n has the distribution of a *simply generated random increasing tree* [40, Section 1.3.3], with weight sequence $\phi_d = \frac{1}{d!}\prod_{k=0}^{d-1} w_k$. Conversely, a simply generated random increasing tree can be generated by a random evolution where nodes are added one by one only when its weight sequence is of this form, for some w_k of the form (6.17) [82], [100]. (Such trees are called *very simple increasing trees* in [82], [100].) In other words, the random increasing tree generated by a general sequence of weights w_k (as in Example 6.3) is a simply generated increasing tree if and only if the weights are of the linear type (6.17). (I.e., we are in the case of the present example.)

Finally, (6.22) is very useful when using martingale methods (which we do not do in the present paper).

Example 6.5 (plane oriented recursive tree)**.** A random *plane oriented recursive tree*, introduced by Szymański [107], is constructed similarly to the random recursive tree in Example 6.1, but we now consider the trees as ordered; an existing node with k children thus has $k+1$ position in which a new node can be added, and we give all possible positions of the new node the same probability. The probability of choosing a node v as the parent is thus proportional to $d^+(v)+1$, so the plane oriented recursive tree is the case $w_k = k+1$ of Example 6.3. This is the special case $\chi = \rho = 1$ of Example 6.4, and thus the special case $\rho = 1$ of the following example (Example 6.6), where some results and further references are given.

Example 6.6 (positive linear preferential attachment)**.** Consider the case $\chi = 1$ of (6.17), i.e.,

$$w_k = k+\rho, \qquad k \geqslant 0, \tag{6.28}$$

where $\rho > 0$ is a parameter.

Thus, w_k is a strictly increasing function of k, so this is a model with preferential attachment as mentioned in Example 6.3. This is a popular model, that has been studied by many authors (often by methods different from the branching processes used here). The original preferential attachment model by Barabási and Albert [6] was the case $\rho = 1$, so $w_k = k+1$; thus the probability of attaching a new node to an existing node v is proportional to $d^+(v)+1 = d(v)$, the total degree of the node (except for the root). As said above, trees of this

type had earlier been studied by Szymański [107]. (Barabási and Albert [6] considered a more general model where a new node may be attached to more than one existing node, thus creating graphs that are not trees. We only consider the tree case here.) Bollobás, Riordan, Spencer and Tusnády [17] made a precise formulation of the definition, and found (and proved rigorously) the asymptotic degree distribution (in the general, graph case). See also van der Hofstad [58, Chapter 8], with many details and references. The tree model with a general ρ was studied by Móri [92]. See also Athreya, Ghosh, and Sethuraman [4] for an extension with multiple edges, treated by an extension of the methods used here. Rudas, Tóth and Valkó [105] and Rudas and Tóth [104] also used the branching process method described here.

In the case (6.28), (6.11) becomes a hypergeometric series

$$\widehat{\mu}(\theta) = \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{k+\rho}{k+\rho+\theta} = F(\rho, 1; \rho+\theta; 1) - 1, \tag{6.29}$$

where F is a hypergeometric function, see (B.1) in Appendix B; the series converges for $\theta > 1$, and then (6.29) and (B.2) yield

$$\widehat{\mu}(\theta) = \frac{\Gamma(\rho+\theta)\Gamma(\theta-1)}{\Gamma(\rho+\theta-1)\Gamma(\theta)} - 1 = \frac{\rho+\theta-1}{\theta-1} - 1 = \frac{\rho}{\theta-1}, \tag{6.30}$$

as we have seen by another method in (6.19). Consequently, or by (6.20), the Malthusian parameter is

$$\alpha = \rho + 1. \tag{6.31}$$

The asymptotic degree distribution is by (6.13)–(6.14) and (6.31) given by

$$\mathbb{P}(D \geqslant i) = \prod_{k=0}^{i-1} \frac{k+\rho}{k+2\rho+1} = \frac{\Gamma(2\rho+1)\Gamma(i+\rho)}{\Gamma(\rho)\Gamma(i+2\rho+1)}, \qquad i \geqslant 0, \tag{6.32}$$

and thus

$$\mathbb{P}(D = i) = \mathbb{P}(D \geqslant i) - \mathbb{P}(D \geqslant i+1) = \frac{(\rho+1)\Gamma(2\rho+1)\Gamma(i+\rho)}{\Gamma(\rho)\Gamma(i+2\rho+2)}, \quad i \geqslant 0. \tag{6.33}$$

This is the hypergeometric distribution $\mathsf{HG}(\rho, 1; 2\rho+2)$, see Definition B.1. (This also follows from Theorem A.5, using Example A.1 and (6.31).) This degree distribution has a power-law tail: (6.33) implies, see Theorem B.7,

$$\mathbb{P}(D = i) \sim c(\rho) i^{-\rho-2}, \qquad \text{as } i \to \infty, \tag{6.34}$$

for the constant $c(\rho) = (\rho+1)\Gamma(2\rho+1)/\Gamma(\rho)$, as shown by Móri [92]. (Note that this power-law is quite sensitive to the choice of w_k, with the exponent depending on the constant term ρ in (6.28).)

In the special case $\rho = 1$, $D \sim \mathsf{HG}(1, 1; 4)$ and (6.33) becomes

$$\mathbb{P}(D = i) = \frac{4}{(i+1)(i+2)(i+3)}, \qquad i \geqslant 0, \tag{6.35}$$

found by Szymański [107] (showing the annealed version); see also Mahmoud, Smythe and Szymański [89], Lu and Feng [84], Bollobás, Riordan, Spencer and Tusnády [17] (quenched version), and Janson [68].

By (5.41) and (6.32),

$$q_i = \mathbb{P}(D \geqslant i) = \prod_{k=0}^{i-1} \frac{k+\rho}{k+2\rho+1} = \frac{\rho}{i+2\rho} \prod_{k=0}^{i-2} \frac{k+\rho+1}{k+2\rho+1} \qquad i \geqslant 1, \tag{6.36}$$

which in the special case $\rho = 1$ simplifies to

$$q_i = \frac{2}{(i+1)(i+2)}, \qquad i \geqslant 1. \tag{6.37}$$

Again using Definition B.1, this says that if I is the index of the heir of an ancestor, then $I - 1 \sim \mathsf{HG}(\rho+1, 1; 2\rho+2)$.

The limiting random sin-tree is given by the construction in Example 6.4.

Example 6.7 (m-ary increasing tree, negative linear preferentialattachment)**.** We may generalise the binary case Example 6.2 and grow a random m-ary tree as follows, for any $m \geqslant 2$. This is sometimes called an *m-ary increasing tree*. Note that for $m > 2$, this will *not* give the m-ary search tree defined in Section 3. (One difference is that we here fix the number of nodes to be n, while the m-ary search tree has a random number of nodes, but this is a minor technicality, see Remark 7.1. A more essential difference is seen in the asymptotic degree distribution D, see Theorem 7.14)

Start with a single node. Let each node have m positions for children, labelled $1, \dots, m$. Add each new node to an empty child position in the tree, chosen uniformly at random. (We may, as in Section 3, regard the empty child positions as external nodes.)

Since a node with outdegree d has $m - d$ empty positions for children, this is an instance of the general preferential attachment in Example 6.4, with

$$w_k = m - k, \qquad k = 0, \dots, m. \tag{6.38}$$

This is thus the case $\chi = -1$ of the linear case in Example 6.4 (with $\rho = m$), so all results there hold. In particular, by (6.19)–(6.20), μ has density me^{-t},

$$\widehat{\mu}(\theta) = \frac{m}{\theta+1} \tag{6.39}$$

and

$$\alpha = m - 1. \tag{6.40}$$

Also in the case (6.38), (6.11) becomes a hypergeometric series; in this case we obtain, cf. (6.29) and (B.1),

$$\widehat{\mu}(\theta) = \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{m-k}{m-k+\theta} = {}_2F_1(-m, 1; -m-\theta; 1) - 1. \tag{6.41}$$

(This is a case where the hypergeometric series is finite.) Gauss' formula (B.2) yields another proof of (6.39).

The asymptotic degree distribution is by (6.13)–(6.14) and (6.40) given by

$$\mathbb{P}(D \geqslant i) = \prod_{k=0}^{i-1} \frac{m-k}{2m-1-k} = \frac{m!\,(2m-1-i)!}{(2m-1)!\,(m-i)!} = \frac{\binom{2m-1-i}{m-1}}{\binom{2m-1}{m-1}} \tag{6.42}$$

and thus

$$\mathbb{P}(D = i) = \frac{(m-1)m!\,(2m-2-i)!}{(2m-1)!\,(m-i)!} = \frac{\binom{2m-2-i}{m-2}}{\binom{2m-1}{m-1}}, \qquad i = 0, \dots, m. \tag{6.43}$$

Thus, using Definition B.1 or Remark B.2, D has the hypergeometric distribution $\mathsf{HG}(-m, 1, 2-2m)$. (Again, this also follows by Theorem A.5 and Example A.1.) Note that the distribution (6.43) is the same as the distribution of balls in a given box when m indistinguishable balls are distributed in m distinguishable boxes, with all distributions having the same probability (Bose–Einstein statistics, see Example B.12); equivalently, it is the distribution of the first term in a random composition of m into m (possibly empty) parts.

It follows easily from (6.43) that as $m \to \infty$, the distribution of D converges to the Geometric distribution $\mathrm{Ge}_0(1/2)$, which is the distribution of D for the random recursive tree. (Cf. [69, Example 12.2 and Theorem 11.7].)

The point process Ξ contains m points, with successive exponential waiting times with rates m, $m-1$, …,1. As is well-known, this process can also be constructed by taking m i.i.d. $\bar{\xi}_1, \dots, \bar{\xi}_m \sim \mathrm{Exp}(1)$ and ordering them as $\xi_1 < \cdots < \xi_m$. Since the construction of the m-ary tree also involves randomly labelling the children, it follows that if $\bar{\xi}_i$ denotes the age when child at position i is born, then $\bar{\xi}_1, \dots, \bar{\xi}_m$ are i.i.d. Exp(1). The growing tree $\mathcal{T}_t$ is thus the subtree of the (rooted) infinite m-ary tree, where each child of each node is born after an Exp(1) waiting time (with all these waiting times independent).

Similarly, for an ancestor, the process Ξ' of its ordinary children described in Example 6.4 and Theorem 6.9 simply consists of $m-1$ i.i.d. Exp(1) points. Furthermore, the age $\widetilde{\xi}^*$ when the heir is born is $\mathrm{Exp}(m)$ and independent of Ξ'. Consequently, the description of the limiting random sin-tree $\widetilde{\mathcal{T}}$ in Example 6.4 can be simplified as follows, cf. the binary case in Example 6.2: Construct an infinite chain of ancestors $o^{(1)}, o^{(2)}, \dots$ of o (backwards in time), with the times between their births i.i.d. $\mathrm{Exp}(m)$; in other word, $(o^{(k)})_{k\geqslant 1}$ are born according to a Poisson process with intensity m on $(-\infty, 0)$. Moreover, make a random choice (uniform and independent of everything else) for each ancestor to decide which of its m children that is its heir. Then grow independent m-ary tree processes at all empty places (external nodes), with independent Exp(1) waiting times for all new nodes. Finally, stop everything at $\overline{\tau} \sim \mathrm{Exp}(m-1)$.

The examples above are all cases of Example 6.3, and all except the general Example 6.3 itself are special cases of Example 6.4. We have seen that in the latter cases, the age $\widetilde{\xi}^*$ when the heir is born to an ancestor has an expo-

nential distribution, and is independent of the births of the other children. We give a simple example showing that this is not always the case.

Example 6.8 (Binary pyramids)**.** Let $w_0 = w_1 = 1$ and $w_k = 0$ for $k > 1$. Thus no node ever gets more than 2 children, and we can regard the result as a binary tree by randomly labelling children as left or right as in Example 6.2, but the difference is that we here have $w_0 = w_1$; hence, when adding a new node, the parent of the new node is chosen uniformly among all existing nodes with less than 2 children. (I.e., as in Example 6.1 but with a cut-off at 2 children.) This random tree was called a *binary pyramid* by Mahmoud [86] who studied their height. (The name comes from pyramid schemes for chain letters, see Gastwirth and Bhattacharya [51]. As said in [86], the definition can be generalized to an arbitrary cut-off $m \geqslant 2$; we leave this case to the reader.)

We have $\xi_1 = X_1 \sim \mathrm{Exp}(1)$ with density e^{-x} and $\xi_2 = X_1 + X_2 \sim \Gamma(2,1)$ with density xe^{-x}. Hence, the intensity μ has density $(1+x)e^{-x}$ and Laplace transform, by (5.7) or (6.11),

$$\widehat{\mu}(\theta) = \frac{1}{1+\theta} + \frac{1}{(1+\theta)^2}, \qquad \theta > -1. \tag{6.44}$$

Hence (5.4) is satisfied with $\alpha = (\sqrt{5}-1)/2$ (the inverse golden ratio). By Remark 5.22, the age $\widetilde{\xi}^*$ when an heir is born has the density $(1+t)e^{-(1+\alpha)t} = (1+t)e^{-\frac{1}{2}(\sqrt{5}+1)t}$, and thus, or by (5.39), $\beta = \mathbb{E}\,\widetilde{\xi}^* = (3\sqrt{5}-5)/2$.

Furthermore, by (5.35),

$$q_1 = 1/(1+\alpha) = \alpha = (\sqrt{5}-1)/2, \tag{6.45}$$

$$q_2 = 1/(1+\alpha)^2 = \alpha^2 = 1-\alpha = (3-\sqrt{5})/2. \tag{6.46}$$

Thus, by (5.34), $\widetilde{\Xi}$, describing the life history of an ancestor, can be described as a mixture: with probability $q_1 = (\sqrt{5}-1)/2$, an heir is born at age $\widetilde{\xi}_1 \sim \mathrm{Exp}(1+\alpha)$, and then another child is born after an independent waiting time $\widetilde{\xi}_2 - \widetilde{\xi}_1 \sim \mathrm{Exp}(1)$; with probability $1-q_1$, first another child is born at age $\widetilde{\xi}_1 \sim \mathrm{Exp}(1+\alpha)$, and then an heir is born after an independent waiting time $\widetilde{\xi}_2 - \widetilde{\xi}_1 \sim \mathrm{Exp}(1+\alpha)$. We obtain also, by this or directly from (5.34), the joint density $f(x,y)$ of the ages when the ordinary child and the heir is born as

$$f(x,y) = \begin{cases} e^{-x-\alpha y}, & 0 < y < x, \\ e^{-(1+\alpha)y}, & 0 < x < y. \end{cases} \tag{6.47}$$

Consequently, the two births are not independent (unlike the linear case in Example 6.4).

Since $\widetilde{\xi}^*$ is not exponential, the times of births of the ancestors $o^{(1)}, o^{(2)}, \dots$ do not form a Poisson process on $(-\infty, 0)$.

The asymptotic degree distribution is by (5.41) given by $\mathbb{P}(D=0) = \mathbb{P}(D=2) = 1-\alpha = (3-\sqrt{5})/2$, $\mathbb{P}(D=1) = 2\alpha - 1 = \sqrt{5}-2$.

We end this section by proving some claims made above. First the ancestor in the linear case.

Theorem 6.9. *For the linear preferential attachment in Example 6.4, with weights $w_k = \chi k + \rho$, the life history $\widetilde{\Xi}$ of an ancestor consists of an heir born at age $\widetilde{\xi}^* \sim \mathrm{Exp}(\rho)$ together with ordinary children born according to a pure birth process Ξ', with rate w_{k+1} when there are k ordinary children, with $\widetilde{\xi}^*$ and Ξ' independent.*

Proof. Consider an ancestor, let $\widetilde{\xi}^*$ be its age when the heir is born, and denote its age at the births of the other children by $\xi'_1 < \xi'_2 < \dots$. (Also, let $\xi'_0 = 0$.) Let $\mathcal{E}_i$ be the event that the heir is child i. Thus $\mathcal{E}_i$ a.s. equals the event $\{\xi'_{i-1} < \widetilde{\xi}^* < \xi'_i\}$; furthermore, if this event holds, then $\widetilde{\xi}^* = \widetilde{\xi}_i$ and

$$\xi'_j = \begin{cases} \widetilde{\xi}_j, & j < i, \\ \widetilde{\xi}_{j+1}, & j \geqslant i. \end{cases} \tag{6.48}$$

Fix i and $M > i$. For an ordinary individual, the joint distribution of $(\xi_1, \dots, \xi_{M+1})$ has density, on $\{0 < x_1 < \dots < x_{M+1}\}$ and with $x_0 = 0$,

$$\begin{aligned} \prod_{j=0}^{M} w_j e^{-w_j(x_{j+1}-x_j)} &= \prod_{k=0}^{M} w_k \cdot e^{\sum_{j=1}^{M}(w_j - w_{j-1})x_j - w_M x_{M+1}} \\ &= \prod_{k=0}^{M} w_k \cdot e^{\sum_{j=1}^{M} \chi x_j - w_M x_{M+1}}. \end{aligned} \tag{6.49}$$

Hence, for an ancestor, (5.34) shows that restricted to the event $\mathcal{E}_i$, the joint density of $(\widetilde{\xi}_1, \dots, \widetilde{\xi}_{M+1})$ is, on $\{0 < x_1 < \dots < x_{M+1}\}$,

$$e^{-\alpha x_i} \prod_{k=0}^{M} w_k \cdot e^{\sum_{j=1}^{M} \chi x_j - w_M x_{M+1}} = \prod_{k=0}^{M} w_k \cdot e^{\chi \sum_{j=1}^{M} x_j - w_M x_{M+1} - \alpha x_i} \tag{6.50}$$

and, using (6.48), the joint distribution of $(\xi'_1, \dots, \xi'_M, \widetilde{\xi}^*)$ on the set $\{0 < x_1 < \dots < x_{i-1} < y < x_i < \dots < x_M\}$ is, recalling (6.20) and $w_0 = \rho$,

$$\begin{aligned} \prod_{k=0}^{M} w_k \cdot e^{\chi\left(\sum_{j=1}^{M-1} x_j + y\right) - w_M x_M - \alpha y} &= w_0 \prod_{k=1}^{M} w_k \cdot e^{\sum_{j=1}^{M-1} \chi x_j - w_M x_M - \rho y} \\ &= \rho e^{-\rho y} \prod_{k=1}^{M} w_k \cdot e^{\sum_{j=1}^{M-1}(w_{j+1} - w_j)x_j - w_M x_M} \\ &= \rho e^{-\rho y} \prod_{j=0}^{M-1} w_{j+1} e^{-w_{j+1}(x_{j+1} - x_j)}. \end{aligned} \tag{6.51}$$

This equals the joint density of the first M points of the birth process Ξ' defined in the statement, together with an independent $\widetilde{\xi}^* \sim \mathrm{Exp}(\rho)$. The result follows, since M is arbitrary. □

We have shown in Example 6.4 that the age $\widetilde{\xi}^*$ when the heir is born to an ancestor has an exponential distribution in the linear case. We now show the converse: this happens only in the linear case. (Recall that if $w_m = 0$ for some m, the weights w_k for $k > m$ are irrelevant.)

Theorem 6.10. *Consider a general preferential attachment tree defined as in Example 6.3 by a sequence $(w_k)_0^\infty$ of weights. If the age $\widetilde{\xi}^*$ when an ancestor gets an heir has an exponential distribution, then $w_k = \chi k + \rho$ for some $\chi \in \mathbb{R}$ and $\rho > 0$ (at least until $w_k = 0$, if that ever happens).*

Proof. The Laplace transform of $\widetilde{\xi}^*$ is by (5.37) $\widehat{\mu}(s+\alpha)$. Hence, if $\widetilde{\xi}^* \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$, then for all $s > 0$,

$$\widehat{\mu}(s+\alpha) = \mathbb{E}\, e^{-s\widetilde{\xi}^*} = \frac{\lambda}{\lambda + s}. \tag{6.52}$$

Consequently, by (6.11),

$$\frac{w_0}{w_0+s} + \frac{w_0}{w_0+s}\cdot\frac{w_1}{w_1+s} + \cdots = \widehat{\mu}(s) = \frac{\lambda}{\lambda - \alpha + s}, \tag{6.53}$$

at least for $s > \alpha$.

Consider, more generally, the equation

$$\frac{w_0}{w_0+s} + \frac{w_0}{w_0+s}\cdot\frac{w_1}{w_1+s} + \cdots = \frac{a}{b+s}. \tag{6.54}$$

for some real a and b, and all large s. Multiply (6.54) by $(w_0+s)/w_0$. This yields

$$1 + \frac{w_1}{w_1+s} + \frac{w_1}{w_1+s}\cdot\frac{w_2}{w_2+s} + \cdots = \frac{a(w_0+s)}{w_0(b+s)}. \tag{6.55}$$

Now let $s \to \infty$. On the left-hand side, each term except the first decreases to 0, and by dominated convergence, the sum converges to $1+0+\ldots$; thus (6.55) implies

$$1 = 1 + 0 + \cdots = \lim_{s\to\infty} \frac{a(w_0+s)}{w_0(b+s)} = \frac{a}{w_0}. \tag{6.56}$$

Consequently, (6.54) implies $w_0 = a$. Use this in (6.55) and subtract 1 to obtain

$$\frac{w_1}{w_1+s} + \frac{w_1}{w_1+s}\cdot\frac{w_2}{w_2+s} + \cdots = \frac{w_0+s}{b+s} - 1 = \frac{w_0-b}{b+s}. \tag{6.57}$$

If $w_1 = 0$, then the left-hand side vanishes, so $w_0 - b = 0$. Otherwise, (6.57) is of the same type as (6.54), with the weights (w_k) shifted to (w_{k+1}), and a replaced by $w_0 - b$. Hence, the argument above yields $w_1 = w_0 - b$.

Thus, in both cases $w_1 = w_0 - b$. Moreover, if $w_1 \neq 0$, we can iterate the argument, and find $w_2 = w_1 - b$, $w_3 = w_2 - b$, and so on, as long as the weights are non-zero. Thus $w_k = w_0 - kb = \chi k + \rho$, with $\chi = -b$ and $\rho = w_0$. □

Finally, we prove the result claimed above in the explosive case (6.8).

Theorem 6.11. *Let T_n be a general preferential attachment tree, defined by a sequence w_k, and assume that the explosion condition (6.8) holds. Then* $\mathbb{P}(|T_n^*| = 1) \xrightarrow{\text{a.s.}} 1$.

Proof. Let $T_\infty := \mathcal{T}_{\tau(\infty)}$, the (infinite) tree obtained by stopping when the process explodes. Thus $T_n \subset T_\infty$ for every n. Let, for $1 \leqslant i \leqslant n \leqslant \infty$, $I_{i,n} := \mathbf{1}\{d^+_{T_n}(v_i) > 0\}$, the indicator of the event that the i:th node (in order of appearance) v_i has at least one child in T_n.

Fix $\delta > 0$, and let $\mathcal{E}_{i,\delta}$ be the event that the i:th individual (in order of birth) in the branching process gets at least one child before age δ, i.e., that it has $\xi_1 < \delta$. Further, let $J_{i,\delta} := \mathbf{1}\{\mathcal{E}_{i,\delta}\}$. The events $\mathcal{E}_{i,\delta}$ are independent and have the same probability $\mathbb{P}(\xi_1 < \delta) = \mathbb{P}(X_1 < \delta)$. Thus, by the law of large numbers,

$$\frac{\sum_{i=1}^n J_{i,\delta}}{n} \xrightarrow{\text{a.s.}} \mathbb{P}(X_1 < \delta) = 1 - e^{-w_0\delta} < w_0\delta. \tag{6.58}$$

Furthermore, a.s. $\tau(\infty) < \infty$, and then $\sigma_i > \tau(\infty) - \delta$ for all but a finite number of i, i.e., all but a finite number of individuals have age less than δ when the process explodes. Hence, $I_{i,\infty} \leqslant J_{i,\delta}$ for all but a finite number of i and, a.s.,

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n I_{i,\infty}}{n} \leqslant \lim_{n\to\infty} \frac{\sum_{i=1}^n J_{i,\delta}}{n} < w_0\delta. \tag{6.59}$$

Since $\delta > 0$ is arbitrary, this shows $\limsup_{n\to\infty} \sum_{i=1}^n I_{i,\infty}/n = 0$ a.s. Furthermore, the finite tree T_n is a subtree of T_∞; hence, a.s.,

$$\mathbb{P}\big(|T_n^*| > 1\big) = \frac{\sum_{i=1}^n I_{i,n}}{n} \leqslant \frac{\sum_{i=1}^n I_{i,\infty}}{n} \to 0 \tag{6.60}$$

□

7. *m*-ary search trees and branching processes

In this section, as always when we discuss m-ary search trees, m is a fixed integer with $m \geqslant 2$. We apply the general theory in Section 5 to the m-ary search tree in Section 3. Recall from Section 3 that besides the m-ary search tree, we may also consider the extended m-ary search tree (including external nodes). It turns out that both versions can be described by stopped branching processes. It is easy to go between the two versions, but we find it instructive to treat them separately, and describe the two related but different branching processes connected to them. The reader is recommended to compare the two versions, even when we do not explicitly do so.

Remark 7.1. The random m-ary search tree is defined as in Section 3 to have a given number of keys, which makes the number of nodes random (in general). We can also define a random m-ary search tree with a given number of nodes, by adding keys until the desired number of nodes is obtained. This is obtained by the branching processes below, stopping when the number of nodes is a given number n; we thus use the weight $\psi(t) = 1$ in Example 5.3 (as in Section 6).

The asymptotics are the same for this version, see Remark 5.17. We therefore ignore this version in the sequel, and consider only the standard version with a given number of keys.

7.1. Extended m-ary search tree

Recall from Section 3 that we can grow an extended m-ary search tree by starting with an empty tree (a single external node) and then adding keys, each new key added with equal probability to each existing gap. Hence, we can also grow the extended m-ary search tree in continuous time by adding a key to each gap after an exponential Exp(1) waiting time (independent of everything else). By the construction of the extended m-ary search tree in Section 3, this is a Crump–Mode–Jagers branching process, where the life of each individual is as follows (Pittel [102]):

An individual is born as an external node with no keys. It acquires $m-1$ keys after successive independent waiting times $Y_1,\dots,Y_{m-1}$, where $Y_i\sim\mathrm{Exp}(i)$ (since the node has i gaps when there are $i-1$ keys). When the $(m-1)$:th key arrives, the individual immediately gets m children.

We let $\psi(t)$ be the number of keys stored at the individual at age t. Thus Z_t^ψ is the total number of keys at time t and $\tau(n)$ is the time the n:th key is added. Hence T_n is a random m-ary search tree with n keys, as we want.

Let $S_k:=\sum_{i=1}^k Y_i$, $k=0,\dots,m-1$; for $1\leqslant k\leqslant m-1$, this is the time the k:th key arrives. Let further $S_m:=\infty$. Then $\psi(t)=k$ for $S_k\leqslant t<S_{k+1}$. For $\theta\geqslant 0$ (in fact, for $\theta>-1$) and $k\leqslant m-1$,

$$\mathbb{E}\,e^{-\theta S_k}=\prod_{i=1}^k\mathbb{E}\,e^{-\theta Y_i}=\prod_{i=1}^k\frac{i}{i+\theta}=\frac{k!}{\prod_{i=1}^k(i+\theta)}. \tag{7.1}$$

(See also Theorem C.1, which furthermore gives the distribution of S_k; in the notation used in Appendix C, $S_k\overset{\mathrm{d}}{=}V_{k,k}$.)

Furthermore, all children are born at the same time with $\xi_1=\dots=\xi_m=S_{m-1}$, and thus the random variable $\widehat{\Xi}(\theta)$ in Remark 5.7 equals $me^{-\theta S_{m-1}}$. Hence, see (5.7) and (7.1),

$$\widehat{\mu}(\theta)=m\,\mathbb{E}\,e^{-\theta S_{m-1}}=\frac{m!}{\prod_{i=1}^{m-1}(i+\theta)},\qquad \theta>-1. \tag{7.2}$$

In particular, we see that $\widehat{\mu}(1)=1$, so the Malthusian condition (5.4) is satisfied with $\alpha=1$. It is easy to see that all other conditions (A1)–(A5) are satisfied. (Note that in this case, $N=m$ is non-random. Furthermore, ψ is bounded, so (A6ψ) holds too.) Consequently, Theorem 5.14 applies, and shows (in particular) that the random fringe tree T_n^* converges in distribution to $\overline{\mathcal{T}}$, which is obtained by running the branching process above and stopping it after a random time $\overline{\tau}\sim\mathrm{Exp}(1)$.

Similarly, Theorem 5.25 applies. In order to find the sin-tree $\widetilde{\mathcal{T}}$, note that since all children of an individual are born at the same time, so $\xi_1=\dots=\xi_m=S_{m-1}$,

it does not matter which one is the heir. It thus follows from (5.34), that if we let $\widetilde{Y}_1,\dots,\widetilde{Y}_{m-1}$ be the successive waiting times between the arrival of keys for an ancestor, so all m children are born at time $\widetilde{\xi}=\sum_{i=1}^{m-1}\widetilde{Y}_i$, then $\widetilde{Y}_1,\dots,\widetilde{Y}_{m-1}$ have joint density

$$me^{-(y_1+\cdots+y_{m-1})}\prod_{i=1}^{m-1} ie^{-iy_i} = m!\prod_{i=1}^{m-1} e^{-(i+1)y_i} = \prod_{i=1}^{m-1}(i+1)e^{-(i+1)y_i}. \tag{7.3}$$

Thus $\widetilde{Y}_1,\dots,\widetilde{Y}_{m-1}$ are independent with $\widetilde{Y}_i\sim\operatorname{Exp}(i+1)$. (Cf. (6.4) and the proof of Theorem 6.9, with similar calculations in different but related situations.) The m children are numbered $1,\dots,m$, with the heir chosen uniformly at random among them.

Remark 7.2. The age $\widetilde{\xi}^*$ when the heir is born is $\widetilde{\xi}^*=\sum_{i=1}^{m-1}\widetilde{Y}_i$, with mean

$$\beta=\mathbb{E}\,\widetilde{\xi}^*=\sum_{i=1}^{m-1}\mathbb{E}\,\widetilde{Y}_i=\sum_{i=1}^{m-1}\frac{1}{i+1}=H_m-1, \tag{7.4}$$

where $H_m:=\sum_1^m\frac{1}{i}$ denotes the m:th harmonic number. (See also (5.39) and (7.2).) The distribution of $\widetilde{\xi}^*$ is given by Theorem C.1; using the notation there

$$\widetilde{\xi}^*=\sum_{i=1}^{m-1}\widetilde{Y}_i\overset{\mathrm{d}}{=}\sum_{i=2}^{m}Y_i\overset{\mathrm{d}}{=}V_{m,m-1}. \tag{7.5}$$

In particular, $\widetilde{\xi}^*$ is *not* exponentially distributed unless $m=2$.

In the construction of $\mathcal{T}_t$ above, the number of gaps is always 1 + the number of keys, and we add keys (and thus gaps) with an intensity equal to the number of gaps. Hence, the number of gaps at time t forms a pure birth process with birth rates $\lambda_k=k$, starting at 1 (this is again the Yule process in Example 6.1, see Example A.3), and thus the number of keys at time t forms a pure birth process with birth rates $\lambda_k=k+1$, starting at 0. (Note that this is independent of the choice of m.)

Since $\overline{\tau}$ has the same distribution Exp(1) here as in Example 6.1, it follows that the number of gaps in $\overline{\mathcal{T}}=\mathcal{T}_{\overline{\tau}}$ has the same distribution as the number of nodes M in $\mathcal{T}_{\overline{\tau}}$ in Example 6.1, given by (6.2). Moreover, by symmetry, conditioned on the number of keys $K=k$ in $\mathcal{T}_{\overline{\tau}}=\overline{\mathcal{T}}$, $\overline{\mathcal{T}}$ has the same distribution as the random extended m-ary search tree T_k with k keys. Hence, we get the following result:

Theorem 7.3. *The number K of keys in the asymptotic fringe tree $\overline{\mathcal{T}}$ has the distribution*

$$\mathbb{P}(K=k)=\mathbb{P}(M=k+1)=\frac{1}{(k+1)(k+2)},\qquad k\geqslant 0. \tag{7.6}$$

Furthermore, $\overline{\mathcal{T}}$ can also be described as an extended m-ary search tree with a random number K keys, where K has the distribution (7.6). □

Remark 7.4. Using the notation in Definition B.1, $K \sim \mathsf{HG}(1,1;3)$. (This also follows from Theorem A.5, with $\chi = \rho = \alpha = 1$.)

The property in the second part of the theorem, describing the asymptotic fringe tree $\overline{\mathcal{T}}$ as an extended m-ary search tree with a random number K keys is called *coherence* by Aldous [1, Section 2.6], and was seen also in Example 6.4. (In the present case with respect to the number of keys; we might call this *key-coherent*.)

We proceed to derive some properties of the random extended m-ary search tree T_n. Note that, unlike the examples in Section 6, T_n does not have n nodes; n is the number of keys, while the number of nodes is random for $m \geqslant 3$. (For $m = 2$, the number of nodes is $2n+1$, of which n are internal, see Section 3.) To find the asymptotic number of nodes, we use Theorem 5.12 and obtain the following result.

Theorem 7.5. *For the extended m-ary search tree T_n with n keys,*

$$\frac{|T_n|}{n} \xrightarrow{\text{a.s.}} \frac{1}{m_\psi} = \frac{1}{H_m - 1}. \tag{7.7}$$

The asymptotic value of the expectation $\mathbb{E}\,|T_n|/n$ was found by Baeza-Yates [5]. We do not know any reference where (7.7) is stated explicitly, but closely related results for the number of internal nodes have been shown in several papers, see Remark 7.12; the result follows also immediately from the main result by Kalpathy and Mahmoud [76].

Proof. This follows from Theorem 5.12(ii), except for the value of m_ψ, which we calculate as follows. Since $\psi(t) = \sum_{i=1}^{m-1} \mathbf{1}\{S_i \leqslant t\}$,

$$\hat{\psi}(1) = \int_0^\infty e^{-t}\psi(t)\,\mathrm{d}t = \sum_{i=1}^{m-1} \int_0^\infty e^{-t}\mathbf{1}\{S_i \leqslant t\}\,\mathrm{d}t = \sum_{i=1}^{m-1} e^{-S_i}. \tag{7.8}$$

By (7.1),

$$\mathbb{E}\,e^{-S_i} = \frac{i!}{(i+1)!} = \frac{1}{i+1}, \qquad i = 0,\dots,m-1, \tag{7.9}$$

and thus,

$$m_\psi = \mathbb{E}\,\hat{\psi}(1) = \sum_{i=1}^{m-1} \mathbb{E}\,e^{-S_i} = \sum_{i=1}^{m-1} \frac{1}{i+1} = H_m - 1. \tag{7.10}$$

□

Theorem 7.6. *Let $N_k(T_n)$ be the number of nodes in T_n with k keys, for $k = 0,\dots,m-1$. Then,*

$$\frac{N_k(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \begin{cases} \frac{1}{(k+1)(k+2)}, & 0 \leqslant k < m-1, \\ \frac{1}{m}, & k = m-1. \end{cases} \tag{7.11}$$

Again, we do not know any reference where this is stated explicitly; the asymptotic values of the expectations $\mathbb{E}\,N_k(T_n)/n$ were found by Baeza-Yates [5]; see also the references in Remark 7.12. The result can also easily be shown using Pólya urns, see [67, Example 7.8], [76] and [60].

Proof. $N_k(T_n) = n_{\mathcal{P}}(T_n)$, where $\mathcal{P}$ is the property of a node v that it contains k keys. Hence, $p_{\mathcal{P}}(\overline{\mathcal{T}})$ is the probability that the root of $\overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$ contains k keys, i.e., that $\psi(\overline{\tau}) = k$ or, equivalently, $S_k \leqslant \overline{\tau} < S_{k+1}$. We apply Theorem 5.14, and note that the characteristic ϕ there is $\phi(t) = \mathbf{1}\{\psi(t) = k\} = \mathbf{1}\{S_k \leqslant t < S_{k+1}\}$. By (5.24), arguing similarly to (7.8)–(7.10) and in particular using (7.9), and recalling that $S_m := \infty$,

$$\begin{aligned} p_{\mathcal{P}}(\overline{\mathcal{T}}) = \mathbb{E}\,\hat{\phi}(1) = \mathbb{E}\int_0^\infty e^{-t}\mathbf{1}\{S_k \leqslant t < S_{k+1}\}\,\mathrm{d}t = \mathbb{E}\bigl(e^{-S_k} - e^{-S_{k+1}}\bigr) \\ = \begin{cases} \frac{1}{k+1} - \frac{1}{k+2}, & 0 \leqslant k < m-1, \\ \frac{1}{k+1}, & k = m-1. \end{cases} \end{aligned} \tag{7.12}$$

The result follows by Theorem 5.14. (Alternatively, one can use Theorem A.4.) □

Remark 7.7. In particular, the fraction of external nodes

$$\frac{N_0(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \frac{1}{2}. \tag{7.13}$$

and thus the same holds for the number of internal nodes; the numbers of external and internal nodes are thus asymptotically the same. (Perhaps surprisingly, the asymptotic fractions of external and internal nodes are thus independent of m.)

Remark 7.8. The asymptotic degree distribution D is not very interesting for the extended m-ary search tree, since every internal node has outdegree m and every external node has outdegree 0; thus, as a corollary of (7.13), $\mathbb{P}(D = 0) = \mathbb{P}(D = m) = 1/2$.

7.2. m-ary search tree, internal nodes only

Usually, we consider an m-ary search tree as consisting only of the internal nodes. This can be obtained from the tree with external nodes in Section 7.1 by deleting all external nodes, but it may also be constructed directly as follows, using a different Crump–Mode–Jagers process.

We now start with a node containing a single key. Thus each individual is born as a node with 1 key. It acquires more keys after successive waiting times $Y_2, \dots, Y_{m-1}$, where $Y_i \sim \mathrm{Exp}(i)$. At the arrival of the $(m-1)$:th key, at time $S := \sum_{i=2}^{m-1} Y_i$, the individual becomes fertile; it then gets m children, marked by $1, \dots, m$, with child i born after a further waiting time X_i, i.e., at time $\xi_i = S + X_i$, where $X_1, \dots, X_m$ are independent and $\mathrm{Exp}(1)$. (The children are here marked by their final positions in the tree, see Remark 5.1.)

Alternatively, taking the children in order of birth, we may say that after the $(m-1)$:th key, there are m children born after successive waiting times $X'_1, \dots, X'_m$, with $X'_i \sim \mathrm{Exp}(m+1-i)$, all waiting times independent.

We let again the weight $\psi(t)$ be the number of keys at time t in an individual. It is easy to see that then T_n is a random m-ary search tree with n keys, as defined in Section 3.

The random variable $\widehat{\Xi}(\theta)$ in Remark 5.7 is now given by

$$\widehat{\Xi}(\theta) = \sum_{j=1}^{m} e^{-\theta(S+X_j)} = \sum_{j=1}^{m} e^{-\theta(\sum_{i=2}^{m-1} Y_i + X_j)}. \tag{7.14}$$

Its distribution is not the same as in Section 7.1, but the mean $\mathbb{E}\,\widehat{\Xi}(\theta) = \widehat{\mu}(\theta)$ is easily seen to be the same as in (7.2), and thus we still have $\alpha = 1$; similarly, by (5.39), β is the same as in (7.4), i.e.,

$$\beta = H_m - 1. \tag{7.15}$$

(That α has to be the same for the m-ary search tree with and without external nodes is rather obvious, since the number of internal nodes in $\mathcal{T}_t$ is the same for both versions, and grows like $e^{\alpha t}$ by (5.10) and (7.13).) The conditions (A1)–(A5) and (A6ψ) are satisfied, and Theorem 5.14 shows that the random fringe tree T_n^* converges in distribution to $\overline{\mathcal{T}}$, which is obtained by running this branching process and stopping it after a random time $\overline{\tau} \sim \mathrm{Exp}(1)$.

Moreover, the random sin-tree $\widetilde{\mathcal{T}}$ is constructed by the general procedure in Section 5. To find the distribution of an ancestor, we note that by symmetry, each child has the same probability $1/m$ of being the heir. Furthermore, using $\widetilde{Y}_i$ and $\widetilde{X}_j$ to denote the waiting times (corresponding to Y_i and X_j above) for an ancestor, it follows from (5.34) that conditioned on the heir being the child marked k, the joint distribution of $\widetilde{Y}_2, \dots, \widetilde{Y}_{m-1}, \widetilde{X}_1, \dots, \widetilde{X}_m$ has density

$$\begin{aligned} &m e^{-\sum_{i=2}^{m-1} y_i - x_k} \prod_{i=2}^{m-1} i e^{-iy_i} \prod_{j=1}^{m} j e^{-jx_j} \\ &\qquad = \prod_{i=2}^{m-1} (i+1) e^{-(i+1)y_i} \prod_{j\neq k} j e^{-jx_j} \cdot 2e^{-2x_k}. \end{aligned} \tag{7.16}$$

Consequently, $\widetilde{Y}_i \sim \mathrm{Exp}(i+1)$ and, given that the heir is child k, $\widetilde{X}_k \sim \mathrm{Exp}(2)$ while $\widetilde{X}_j \overset{\mathrm{d}}{=} X_j \sim \mathrm{Exp}(1)$ for $j \neq k$, all waiting times independent (conditioned on k).

Remark 7.9. The distributions of the birth times can be obtained from Theorem C.1; it follows that using the notation there, $S \overset{\mathrm{d}}{=} V_{m-1,m-2}$, and $\xi_i \overset{\mathrm{d}}{=} V_{m-1,m-1}$ (with marks as above), while the i:th child in birth order is born at time $\xi_i' = S + V_{m,i}$, with S and $V_{m,i}$ independent.

Similarly, an ancestor becomes fertile at time $\widetilde{S} = \sum_{i=2}^{m-1} \widetilde{Y}_i \sim V_{m,m-2}$ and the heir is born at time

$$\widetilde{\xi}^* = \widetilde{S} + \widetilde{X}_k \overset{\mathrm{d}}{=} \sum_{j=2}^{m} Y_j \sim V_{m,m-1}. \tag{7.17}$$

Thus $\widetilde{\xi}^*$ has the same distribution as for the extended m-ary search tree, see (7.5). (This is not surprising since we really construct the same trees in two somewhat different ways.)

As for the extended m-ary search tree in Section 7.1, the number of gaps in the process (= 1 + the number of keys, i.e., $1 + Z_t^\psi$) forms a Yule process (see Example A.3), but in the present case it starts at 2, while it starts with 1 for the extended m-ary search tree in Section 7.1. (In other words, the number of gaps is the sum of two independent standard Yule processes.) The number of keys in $\mathcal{T}_t$ thus evolves in exactly the same way for every $m \geqslant 2$, and hence is the same as for the binary case $m = 2$ treated in Example 6.2. In particular, since also $\overline{\tau} \sim \mathrm{Exp}(1)$ is the same for all m, the number K of keys in $\mathcal{T}_{\overline{\tau}}$ has the distribution (6.6). Moreover, as for the extended m-ary search tree, if we condition on $K = k$, then $\overline{\mathcal{T}}$ has the same distribution as the random m-ary search tree T_k with k keys. Hence, we get the following result:

Theorem 7.10. *The number K of keys in the asymptotic fringe tree $\overline{\mathcal{T}}$ has the distribution*

$$\mathbb{P}(K = k) = \frac{2}{(k+1)(k+2)}, \qquad k \geqslant 1. \tag{7.18}$$

Furthermore, $\overline{\mathcal{T}}$ can also be described as an m-ary search tree T_K with a random number K keys, where K has the distribution (7.18). □

Cf. (7.6), the similar result for the extended m-ary search tree, and note that the distribution (7.18) equals the distribution (7.6) conditioned on $K \geqslant 1$. Furthermore, the number of keys thus grows as e^t, so the number of nodes has to grow at the same rate, which again shows that $\alpha = 1$. Note that the second part of the theorem is another instance of key-coherence.

As for the extended m-ary search tree in Section 7.1, the number of nodes is random for $m \geqslant 3$. We can again find the asymptotics from Theorem 5.12, yielding the following theorem. (Alternatively, we can obtain the result from (7.7) and (7.11) for the extended m-ary search tree.)

Theorem 7.11. *For the m-ary search tree T_n with n keys,*

$$\frac{|T_n|}{n} \xrightarrow{\text{a.s.}} \frac{1}{m_\psi} = \frac{1}{2(H_m - 1)}. \tag{7.19}$$

In other words, the average number of keys per node converges a.s. to $2(H_{m-1} - 1)$.

Remark 7.12. This result was first shown, in a weaker form with convergence in probability, by Mahmoud and Pittel [87]; convergence of the expectation had been shown earlier by Knuth [78, answer to exercise 6.2.4-8] and Baeza-Yates [5].

For the variance and asymptotic distribution (which we do not consider in the present paper), there is an interesting phase transition: the variance is linear in n and the distribution asymptotically normal if $m \leqslant 26$ but not if $m \geqslant 27$, see [87], [83], [26], [25].

Proof. For the present branching process, $\psi(t) = 1 + \sum_{i=2}^{m-1} \mathbf{1}\{S'_i \leqslant t\}$, where $S'_i := \sum_{j=2}^{i} Y_j$ is the time the i:th key comes to the node. Arguing as in (7.8)–(7.10) we find (omitting some details)

$$\hat{\psi}(1) = \int_0^\infty e^{-t}\psi(t)\,\mathrm{d}t = 1 + \sum_{i=2}^{m-1} e^{-S'_i} \tag{7.20}$$

and thus,

$$m_\psi = \mathbb{E}\,\hat{\psi}(1) = 1 + \sum_{i=2}^{m-1}\prod_{j=2}^{i} \mathbb{E}\,e^{-Y_j} = 1 + \sum_{i=2}^{m-1} \frac{2}{i+1} = 2(H_m - 1). \tag{7.21}$$

Hence, Theorem 5.12(ii) yields

$$\frac{|T_n|}{n} \xrightarrow{\text{a.s.}} \frac{1}{m_\psi} = \frac{1}{2(H_m-1)}. \tag{7.22}$$

□

The asymptotic number of nodes with a given number of keys can be found similarly. Note that the tree is constructed so that each node contains at least one key. (This theorem is also an immediate corollary of results by Kalpathy and Mahmoud [76], shown using a Pólya urn, see also [67, Example 7.8].)

Theorem 7.13. *Let $N_k(T_n)$ be the number of nodes in T_n with k keys, for $k = 1, \dots, m-1$. Then,*

$$\frac{N_k(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \begin{cases} \frac{2}{(k+1)(k+2)}, & 1 \leqslant k \leqslant m-2, \\ \frac{2}{m}, & k = m-1. \end{cases} \tag{7.23}$$

Proof. This follows either from (7.11) for the extended m-ary search tree or by a similar argument as in the proof of Theorem 7.6 (which we omit). □

Finally, we give the asymptotic degree distribution D. (This was found, using a Pólya urn, by Kalpathy and Mahmoud [76], generalizing the special case of leaves ($k = 0$) given in [60].)

Theorem 7.14. *Let $n_k(T_n)$ be the number of nodes of outdegree k in T_n. Then*

$$\frac{n_k(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(D = k) = \begin{cases} \frac{m-1}{m+1}, & k = 0, \\ \frac{2}{m(m+1)}, & 1 \leqslant k \leqslant m. \end{cases} \tag{7.24}$$

The asymptotic degree distribution is thus uniform on $\{1, \dots, m\}$, but with a large proportion of the nodes being leaves (outdegree 0). (For $m = 2$, the distribution is uniform on $\{0, 1, 2\}$.) Note that $\mathbb{E}\,D = 1$, as always, see (5.42).

Proof. This follows by straightforward calculations from (5.33) and Remark 7.9, for example using Theorem C.1. However, we find it illuminating to instead give a less computational proof, using the properties of the exponential distributions. Recalling that D is the degree of the root of $\overline{\mathcal{T}}$, we consider the life of an

individual (the root), stopped at $\overline{\tau}$; we regard $\overline{\tau}$ as an exponential clock (the *doomsday clock*) that strikes at a random time, and then stops the process.

After the creation (at $t = 0$, and with a single key), the next thing that happens is either the arrival of the second key, or that the doomsday clock strikes. Since the second key arrives with intensity 2 and the clock strikes with intensity 1, the probability is $2/3$ that the second key will arrive before the clock strikes. Conditioned on this event, the same argument shows that the probability that also the third key arrives before the clock strikes is $3/4$, and so on. It follows that the probability that the node acquires all $m-1$ keys before the clock strikes is

$$\frac{2}{3}\frac{3}{4}\cdots\frac{m-1}{m} = \frac{2}{m}. \tag{7.25}$$

(Note that this argument also yields another proof of Theorem 7.13.)

After the arrival of all $m-1$ keys, assuming that the doomsday clock still has not struck, we wait for the m children. Each child arrives with intensity 1, and the clock strikes with the same intensity, so by symmetry (and independence), the order of the m births and the strike of the clock is uniform among all $(m+1)!$ possibilities. In particular, the position of the clock strike is uniform among these $m+1$ events, i.e., the number of children born before the clock strikes is uniform on $\{0,\dots,m\}$. Combining this and (7.25) we obtain, for $1 \leqslant k \leqslant m$,

$$\mathbb{P}(D = k) = \frac{2}{m}\frac{1}{m+1} \tag{7.26}$$

and, including the cases where less than $m-1$ keys arrive before the clock strikes,

$$\mathbb{P}(D = 0) = 1 - \frac{2}{m} + \frac{2}{m}\frac{1}{m+1}. \tag{7.27}$$

The result follows. □

Remark 7.15. Note that the degree distribution in (7.24) differs from the degree distribution (6.43) for the random m-ary tree defined in Example 6.7; as said there, the two different types of random m-ary trees are thus not even asymptotically equivalent.

8. Median-of-$(2\ell+1)$ binary search tree

Let $\ell \geqslant 1$ be a fixed integer. The random *median-of-*$(2\ell+1)$ *binary search tree*, see e.g. [33], is a modification of the binary search tree in Example 6.2, where each internal node still contains exactly one key, but each external node can contain up to 2ℓ keys. (We can also include the case $\ell = 0$; this is just the extended binary search tree, i.e., the special case $m = 2$ of Section 7.1.)

The tree is grown recursively, starting with a single external node without any keys. The first 2ℓ keys are placed in this node. When the $(2\ell+1)$:th key arrives to the node (or to another external node later in the process), the node becomes an internal node with two new external nodes as children, say v_L and

v_R; moreover, the median of the $2\ell+1$ keys now at the node is found and put in the internal node, while the ℓ keys that are smaller than the median are put in the left child v_L and the ℓ keys that are larger than the median are put in the right child v_R.

In order to model this by a branching process, we start the tree with ℓ keys in the root. (This is no restriction, since the first ℓ keys always go there.) Then each external node will contain between ℓ and 2ℓ keys, throughout the process, and the median-of-$(2\ell+1)$ binary search tree is produced by a branching process with the following life histories: An individual is born as an external node with ℓ keys. It acquires $\ell+1$ additional keys after successive independent waiting times $Y_1,\dots,Y_{\ell+1}$, where $Y_i\sim\operatorname{Exp}(\ell+i)$ (since the node has $\ell+i$ gaps when there are $\ell+i-1$ keys). When the $(\ell+1)$:th key arrives, the individual immediately gets 2 children.

We let the weight $\psi(t)$ be the number of keys stored at the individual at age t. Thus Z_t^ψ is the total number of keys at time t and $\tau(n)$ is the time the n:th key is added. Hence, assuming $n\geqslant\ell$, T_n is a random median-of-$(2\ell+1)$ binary search tree with n keys.

Note that this construction is very similar to the one for the extended m-ary search tree in Section 7.1, and we analyse it in the same way. Let $S_k:=\sum_{i=1}^k Y_i$, $k=0,\dots,\ell+1$; this is the time the node gets its $(\ell+k)$:th key. Then

$$\psi(t)=\begin{cases}\ell+k, & S_k\leqslant t<S_{k+1},\ 0\leqslant k\leqslant\ell,\\ 1, & S_{k+1}\leqslant t.\end{cases}\tag{8.1}$$

For $\theta>-\ell-1$ and $k\leqslant\ell+1$,

$$\mathbb{E}\,e^{-\theta S_k}=\prod_{i=1}^k\mathbb{E}\,e^{-\theta Y_i}=\prod_{i=1}^k\frac{\ell+i}{\ell+i+\theta}=\frac{\langle\ell+1\rangle_k}{\langle\ell+1+\theta\rangle_k}.\tag{8.2}$$

(See also Theorem C.1; in the notation used in Appendix C, $S_k\overset{\mathrm{d}}{=}V_{\ell+k,k}$.)

Furthermore, $\xi_1=\xi_2=S_{\ell+1}$, and thus the random variable $\widehat{\Xi}(\theta)$ in Remark 5.7 equals $2e^{-\theta S_{\ell+1}}$. Hence, see (5.7) and (7.1),

$$\widehat{\mu}(\theta)=2\,\mathbb{E}\,e^{-\theta S_{\ell+1}}=\frac{2\langle\ell+1\rangle_{\ell+1}}{\langle\ell+1+\theta\rangle_{\ell+1}}=\frac{\langle\ell+2\rangle_{\ell+1}}{\langle\ell+1+\theta\rangle_{\ell+1}}.\tag{8.3}$$

In particular, we see that $\widehat{\mu}(1)=1$, so the Malthusian condition (5.4) is satisfied with $\alpha=1$. (Again, $\alpha=1$ has to hold since the number of keys is a Yule process, although now started with ℓ keys.) It is easy to see that all other conditions (A1)–(A5) are satisfied. Consequently, Theorem 5.14 applies; the asymptotic random fringe tree $\overline{\mathcal{T}}$ is obtained by running the branching process above and stopping it after a random time $\overline{\tau}\sim\operatorname{Exp}(1)$.

Theorems 7.5 and 7.6 can be adapted with minor modifications as follows; we omit the proofs which are similar to the ones in Section 7.1, now using (8.2).

Theorem 8.1. *For the median-of-$(2\ell+1)$ binary search tree T_n with n keys,*

$$\frac{|T_n|}{n} \xrightarrow{\text{a.s.}} \frac{1}{m_\psi} = \frac{1}{(\ell+1)(H_{2\ell+2}-H_{\ell+1})}. \qquad \square \quad (8.4)$$

Theorem 8.2. *Let $N_k^{\mathrm{e}}(T_n)$ be the number of external nodes in T_n with k keys, for $k=\ell,\dots,2\ell$, and let $N_k^{\mathrm{i}}(T_n)$ be the number of internal nodes (all having one key). Then, for the median-of-$(2\ell+1)$ binary search tree,*

$$\frac{N_k^{\mathrm{e}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \frac{\ell+1}{(k+1)(k+2)}, \qquad \ell \leqslant k \leqslant 2\ell, \quad (8.5)$$

$$\frac{N_k^{\mathrm{i}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \frac{1}{2}. \qquad \square \quad (8.6)$$

Remark 8.3. Chern, Hwang and Tsai [27] consider (using different methods) a more general class of trees, where an external node has up to $r-1$ keys; when the r:th key arrives to the node, a pivot is selected among them at random, such that its rank R (i.e., its number if the r keys are ordered) has some fixed distribution on $\{1,\dots,r\}$. (The case above is thus $r=2\ell+1$ and $R=\ell+1$; in this case R is deterministic.) The pivot is put in the internal node, and its children get $R-1$ and $r-R$ keys. Translated to the branching process, this means (in general) that the individuals start with different number of keys, which would require a multi-type version of the results above (see Remark 5.9). However, it is possible to modify the branching process by including the external nodes in the life of their parent. Thus the individuals now are the internal nodes. (Properties of external nodes can be found using suitable characteristics.) The life of an individual starts with r keys; these are immediately split up with a random R as above, and we regard the individual as carrying two unborn children (fetuses) with initially $R-1$ and $r-R$ keys. The fetuses get new keys, independently of each other and each with rate 1 + the number of existing keys, and each is born when it has got r keys. We omit the details.

Example 8.4. An m-ary generalisation, introduced by Hennequin [57] and further studied by e.g. [26], [27], [24] and [44], has internal nodes with $m-1$ keys and external nodes with up to $m\ell+m-2$ keys. (Here $m\geqslant 2$ and $\ell\geqslant 0$.) When an external node gets $m\ell+m-1$ keys, it is converted to an internal node with m external children; the $m\ell+m-1$ keys are ordered and keys number $\ell+1, 2(\ell+1),\dots,(m-1)(\ell+1)$ are put in the internal node, while the external children get ℓ keys each. The m-ary search tree in Section 7.1 is the case $\ell=0$.

This version can be treated as above; again each individual starts with ℓ keys, but now it acquires $(m-1)(\ell+1)$ more keys, after waiting times $Y_i\sim\mathrm{Exp}(\ell+i)$, $i=1,\dots,(m-1)(\ell+1)$. At time $S_{(m-1)(\ell+1)}$, m children are born. Note that (7.1)–(7.2) generalize to, cf. the special case (8.3),

$$\widehat{\mu}(\theta) = m\,\mathbb{E}\,e^{-\theta S_{(m-1)(\ell+1)}} = m\prod_{i=1}^{(m-1)(\ell+1)}\mathbb{E}\,e^{-\theta Y_i}$$

$$= m \prod_{i=1}^{(m-1)(\ell+1)} \frac{i+\ell}{i+\ell+\theta} = \frac{m\langle \ell+1\rangle_{(m-1)(\ell+1)}}{\langle \ell+1+\theta\rangle_{(m-1)(\ell+1)}} \tag{8.7}$$
$$= \frac{\langle \ell+2\rangle_{(m-1)(\ell+1)}}{\langle \ell+1+\theta\rangle_{(m-1)(\ell+1)}}$$

and hence again $\alpha = 1$. It then follows from (8.7) and (5.39) that

$$\beta = \sum_{i=1}^{(m-1)(\ell+1)} \frac{1}{i+\ell+1} = H_{m\ell+m} - H_{\ell+1}, \tag{8.8}$$

cf. (7.4) for the case $\ell = 0$. Results for this model can be derived as above, but we leave this to the readers.

More generally, one can similarly make an m-ary version of the model with random pivot in Remark 8.3, see [27]; a corresponding Crump–Mode–Jagers branching process (with the internal nodes as individuals) can be constructed as there.

9. Fragmentation trees

Another type of example is provided by the following fragmentation process, introduced by Kolmogorov [79], see also Bertoin [9, Chapter 1] and Janson and Neininger [73], and the further references given there. Fix $b \geqslant 2$ and the law for a random vector $\mathbf{V} = (V_1, \dots, V_b)$; this is commonly called the *dislocation law*. We assume that $0 \leqslant V_j \leqslant 1$, $j = 1, \dots, b$, and

$$\sum_{j=1}^{b} V_j = 1, \tag{9.1}$$

i.e., that $(V_1, \dots, V_b)$ belongs to the standard simplex. For simplicity we also assume that each $V_j < 1$ a.s. (We allow $V_j = 0$.)

Starting with an object of mass $x_0 \geqslant 1$, break it into b pieces with masses $V_1 x_0, \dots, V_b x_0$. For a given threshold $x_1 \in (0, x_0]$, continue recursively with each piece of mass $\geqslant x_1$, using new (independent) copies of the random vector $(V_1, \dots, V_b)$ each time. The process terminates a.s. after a finite number of steps, leaving a finite set of fragments of masses $< x_1$. We regard the fragments of mass $\geqslant x_1$ that occur during this process as the (internal) nodes of a (random) tree, the *fragmentation tree*; the resulting fragments of mass $< x_1$ can be added as external nodes.

Obviously, the fragmentation tree depends only on the ratio x_0/x_1, so we denote it by $\widehat{T}_{x_0/x_1}$. (We may assume either $x_0 = 1$ or $x_1 = 1$ without loss of generality, but we prefer to be more flexible.)

We can translate the fragmentation process to a Crump–Mode–Jagers branching process by regarding a fragment of mass x as born at time $\log(x_0/x)$; an individual will have b children, born at ages $\xi_1, \dots, \xi_b$ with $\xi_i := -\log V_i$. (If

some $V_i = 0$, we get $\xi_i = \infty$, meaning that this child is not born at all, so there are fewer than b children. Note also that in this section, we do not require that $\xi_1, \xi_2, \dots$ are ordered in increasing order.) It is easy to see that the fragmentation tree $\widehat{T}_{x_0/x_1}$ defined above for a threshold x_1 is the same as the family tree $\mathcal{T}_{\log(x_0/x_1)}$ of this branching process at time $\log(x_0/x_1)$.

The relation (9.1) can be written as

$$\sum_{j=1}^{b} e^{-\xi_i} = 1. \tag{9.2}$$

Taking the expectation we find, see (5.7), $\widehat{\mu}(1) = 1$, so the Malthusian parameter $\alpha = 1$. It is easy to see that the assumptions (A1)–(A5) hold, except possibly (A2); we say that the fragmentation process is *non-lattice* if (A2) holds, i.e., if not every V_i is concentrated on $\{r, r^2, r^3 \dots\}$ for some $r \in (0, 1)$. (A sufficient condition for (A2) is thus that V_1 has a continuous distribution.)

Furthermore, (9.2) and (5.6) say that $\widehat{\Xi}(\alpha) = 1$ is non-random. This has the consequence that the random variable W in Remark 5.11 also is deterministic; more precisely, see [64, Theorem (6.8.1)],

$$W = 1/\beta, \tag{9.3}$$

where by (5.40),

$$\beta = \mathbb{E}\sum_{j=1}^{b} \xi_j e^{-\xi_j} = \mathbb{E}\sum_{j=1}^{b} V_j \log(1/V_j). \tag{9.4}$$

Remark 9.1. The Laplace transform $\Psi(s) := \mathbb{E}\, e^{-sW}$ of the limit W in Remark 5.11 satisfies the functional equation

$$\Psi(s) = \mathbb{E}\prod_{i=1}^{N} \Psi\bigl(se^{-\alpha\xi_i}\bigr), \tag{9.5}$$

and this equation together with (5.15) (provided (5.11) holds) determines W uniquely, see [39]. Assuming $\alpha = 1$ (which can be regarded as a normalisation of the time scale), it is easy to see that a constant W satisfies (9.5) if (9.2) holds, which gives an alternative proof of (9.3). Yet another proof of (9.3) is obtained by noting that when (9.2) holds, the martingales R_n and Y_t in [94] are constant 1, and then (9.3) follows from [94, Corollary 3.2]. (The related intrinsic martingale in [9] is constant 1 too, see [9, Section 1.2.2].) Note also that the converse holds: W is constant (and then $1/\beta$) if and only if (9.2) holds; this too follows from the functional equation (9.5).

Remark 9.2. Note that unlike the trees studied in the previous sections, we consider the family tree $\mathcal{T}_t$ at a fixed time $t = \log(x_0/x_1)$ instead of stopping when some weight Z_t^{ψ} reaches a given value. However, since W is constant, this makes a very small difference. In fact, by (5.9) and (9.3), $Z_t \sim \beta^{-1}e^t$ a.s., and

thus, if we use the characteristic $\psi(t) = 1$ again, the stopping time $\tau(n)$ when the tree has n nodes satisfies a.s.

$$\tau(n) = \log(\beta n) + o(1) = \log n + \log\beta + o(1). \tag{9.6}$$

We may define a fragmentation tree T_n of fixed size n by stopping at $\tau(n)$; in the original formulation this means that we choose the threshold x_1 to be the size of the n:th largest fragment in the process, so that there will be exactly n fragment of size $\geqslant x_1$ (unless there is a tie). We see from (9.6) that asymptotically, this is almost the same as taking a constant time $t = \log n + \log\beta$. For more precise results on $|Z_t|$, and thus on $\tau(n)$, see [73].

Theorem 9.3. *Let $\widehat{T}_{x_0/x_1}$ be a random fragmentation tree defined as above, for a non-lattice fragmentation process. Then Theorem 5.14 holds also (as $x_0/x_1 \to \infty$, and with other obvious notational modifications) for the random fringe tree $\widehat{T}^*_{x_0/x_1}$.*

The limiting random fringe tree $\overline{\mathcal{T}}$ can be constructed by the fragmentation process above, starting at $x_0 = 1$ and with a random threshold $x_1 = U \sim U(0,1)$, with U independent of the fragmentation.

Proof. By the equivalence above of the fragmentation process and the Crump–Mode–Jagers branching process, $\widehat{T}_{x_0/x_1} = \mathcal{T}_{\log(x_0/x_1)}$, and the first part follows from Theorem 5.14 (and its proof).

The limiting fringe tree $\overline{\mathcal{T}}$ is obtained by stopping the branching process $\mathcal{T}_t$ at a random time $\overline{\tau} \sim \mathrm{Exp}(1)$; by the equivalence above, this is equivalent to starting the fragmentation process at $x_0 = 1$ and stopping at a threshold $x_1 = \exp(-\overline{\tau})$. This completes the proof, since $\exp(-\overline{\tau}) \sim U(0,1)$. □

Similarly, Theorem 5.25 holds, and the random sin-tree $\widetilde{\mathcal{T}}$ can be defined by a suitable extension of this random fragmentation process; we leave the details of the general case to the reader, and discuss only one case in the example below.

Example 9.4 (Binary splitting)**.** Let $b = 2$ and $\mathbf{V} = (V_1, V_2) = (V_1, 1 - V_1)$ with $V_1 \sim U(0,1)$. Thus, at each fragmentation event, the object is split into two parts, with uniformly random sizes.

In the corresponding Crump–Mode–Jagers branching process, each individual gets two children, born at ages ξ_1 and ξ_2, where $\xi_1, \xi_2 \sim \mathrm{Exp}(1)$ and one of them determines the other by

$$e^{-\xi_1} + e^{-\xi_2} = 1. \tag{9.7}$$

Note the similarities with the Crump–Mode–Jagers branching process for the binary search tree in Example 6.2; the difference is that there ξ_1 and ξ_2 are independent, while here they are dependent. For properties that depend only on the individual (marginal) distributions of ξ_1, ξ_2 and not on their joint distribution, we thus have the same results for both processes; some examples are the intensity μ, the distribution of $\widetilde{\xi}^* \sim \mathrm{Exp}(2)$ and its mean $\beta = 1/2$, and the expected size of the population $\mathbb{E}\, Z_t = 2(e^t - 1)$. However, many properties really depend on the joint distribution of the times of birth of the children, and

are thus in general different for the two processes. For example, although $\mathbb{E}\, Z_t$ is the same for both processes, the distributions of Z_t are not: for the present process, there is by (9.7) always one child of the root born before time $\log 2$, so $Z_{\log 2} \geqslant 2$, while for the process in Example 6.2, $\mathbb{P}(Z_t = 1) = e^{-2t} > 0$ for every $t \geqslant 0$. Also the fringe tree distributions will be different, as is seen below.

Let us first consider the asymptotic outdegree distribution in the fragmentation tree, which equals the distribution of the root D in $\overline{\mathcal{T}}$. We have, using the construction of $\overline{\mathcal{T}}$ in Theorem 9.3,

$$\begin{cases} D = 0 \iff U \geqslant \max(V_1, 1 - V_1) \\ D = 1 \iff \min(V_1, 1 - V_1) < U \leqslant \max(V_1, 1 - V_1) \\ D = 2 \iff U < \min(V_1, 1 - V_1) \end{cases} \tag{9.8}$$

and simple calculations yield

$$\mathbb{P}(D = 0) = \mathbb{P}(1 - U \leqslant V_1 \leqslant U) = \mathbb{E}(2U - 1)_+ = 1/4 \tag{9.9}$$

and similarly $\mathbb{P}(D = 1) = 1/2$ and $\mathbb{P}(D = 2) = 1/4$. Consequently, $D \sim \mathrm{Bi}(2, 1/2)$ has a binomial distribution. (This differs from Example 6.2, where D has a uniform distribution.)

Furthermore, let $X_1 := V_1/U$ and $X_2 := (1 - V_1)/U$ be the masses of the two children of the root, relative to the threshold U. Then $X_1, X_2 > 0$ and $X_1 + X_2 \geqslant 1$, and a calculation of the Jacobian of the mapping $(U, V_1) \mapsto (X_1, X_2)$ shows that in this region, (X_1, X_2) has the density $f(x_1, x_2) = (x_1 + x_2)^{-3}$. This enables us to again compute the distribution of D; for example, $D = 0 \iff X_1, X_2 < 1$. Moreover, we can now easily find the distribution of nodes in the second generation too; we give a few examples.

Denote the children of the root by v_1 and v_2. Then $\overline{\mathcal{T}}$ contains v_1 but not v_2 if and only if $X_1 \geqslant 1 > X_2$; denote this event by $\mathcal{E}_1$. Conditioned on $\mathcal{E}_1$, the density of X_1 is, by a small calculation, $2\bigl(x^{-2} - (x+1)^{-2}\bigr)$, $x \geqslant 1$. Furthermore, the outdegree of v_1 is given by (9.8) with U replaced by $1/X_1$ (and V_1 by an independent copy V_{11}); hence, in analogy to (9.9),

$$\begin{aligned} \mathbb{P}\bigl(\deg(v_1) = 0 \mid \mathcal{E}_1\bigr) &= \mathbb{E}\Bigl(\Bigl(\frac{2}{X_1} - 1\Bigr)_+ \Bigm| \mathcal{E}_1\Bigr) \\ &= \int_1^2 \Bigl(\frac{2}{x} - 1\Bigr)\Bigl(\frac{2}{x^2} - \frac{2}{(x+1)^2}\Bigr)\,\mathrm{d}x \\ &= \frac{3}{2} - 8\log 2 + 4\log 3 \approx 0.34927, \end{aligned} \tag{9.10}$$

and, similarly,

$$\begin{aligned} \mathbb{P}\bigl(\deg(v_1) = 1 \mid \mathcal{E}_1\bigr) &= 2\,\mathbb{E}\Bigl(\min\Bigl(\frac{1}{X_1}, 1 - \frac{1}{X_1}\Bigr) \Bigm| \mathcal{E}_1\Bigr) \\ &= 1 + 12\log 2 - 8\log 3 \approx 0.52887, \end{aligned} \tag{9.11}$$

$$\begin{aligned}\mathbb{P}\bigl(\deg(v_1)=2 \mid \mathcal{E}_1\bigr) &= \mathbb{E}\Bigl(\Bigl(1-\frac{2}{X_1}\Bigr)_+ \Bigm| \mathcal{E}_1\Bigr)\\ &= -\frac{3}{2}-4\log 2+4\log 3 \approx 0.12186.\end{aligned} \tag{9.12}$$

Recall that if we just condition on $v_1 \in \overline{\mathcal{T}}$, its outdegree distribution equals the unconditional distribution of D, i.e., $\mathrm{Bin}(2,1/2)$; hence, (9.10)–(9.12) illustrate the dependencies between the outdegrees of different nodes (in this case, o and v_1), see Remark 5.24.

We obtain also $\mathbb{P}(|\overline{\mathcal{T}}|=1)=\mathbb{P}(D=0)=1/4$ and, from (9.10),

$$\begin{aligned}\mathbb{P}(|\overline{\mathcal{T}}|=2) &= 2\,\mathbb{P}\bigl(\mathcal{E}_1 \text{ and } \deg(v_1)=0\bigr) = 2\,\mathbb{P}\bigl(\deg(v_1)=0 \mid \mathcal{E}_1\bigr)\,\mathbb{P}(\mathcal{E}_1)\\ &= \frac{3}{4}-4\log 2+2\log 3 \approx 0.17464.\end{aligned} \tag{9.13}$$

Again, this differs from the binary search tree in Example 6.2. We do not know any general formula for the probability distribution of the size $|\overline{\mathcal{T}}|$.

The irrational probabilities in (9.10)–(9.12) and (9.13) seem to exclude any simple combinatorial construction or interpretation of the asymptotic fringe tree $\overline{\mathcal{T}}$.

To construct the limiting sin-tree $\widetilde{\mathcal{T}}$ in Theorem 5.25, we note that the heir $\widetilde{\xi}^*$ of an ancestor has distribution $\mathrm{Exp}(2)$ by (5.34) (as in Example 6.2). Going back to the mass scale, we note that $Y := \exp(\widetilde{\xi}^*)$ has the Pareto(2) distribution

$$\mathbb{P}(Y>x)=\mathbb{P}(\widetilde{\xi}^*>\log x)=x^{-2}, \qquad x>1. \tag{9.14}$$

The random sin-tree $\widetilde{\mathcal{T}}$ can thus be constructed as follows: Start with a root o of mass 1 (as in Theorem 9.3) and give it an infinite sequence of ancestors of mass Y_1, Y_1Y_2, $Y_1Y_2Y_3$, …, where Y_i are i.i.d. with the distribution (9.14); the other child of the ancestors thus has mass Y_1-1, $Y_1(Y_2-1)$, $Y_1Y_2(Y_3-1)$, …. Grow independent fragmentation trees from these other children and from o, using uniformly random binary splittings, and stop at a common threshold $x_1=U\sim U(0,1)$.

Remark 9.5. We have, for simplicity, assumed that the branching factor b is a constant finite integer. (Although we may allow fewer than b fragments by letting some $V_i=0$.) We can also allow $b=\infty$, or a random b (which can be reduced to $b=\infty$ by adding variables V_i that are 0). The results above extend, provided (A5) holds.

Remark 9.6. As noted in Section 5, the results extend also to the lattice case, with minor modifications, but for simplicity we ignore that case. Only very special fragmentation processes are lattice; one trivial example is the deterministic symmetric binary splitting $V_1=V_2=1/2$. More generally, the deterministic binary splitting $V_1=p$, $V_2=q=1-p$ is lattice if and only if $\log p/\log q$ is rational. For a random example, let $r=(\sqrt{5}-1)/2$, take $b=3$ and let (V_1,V_2,V_3) be either $(r,r^2,0)$ or (r^2,r^2,r^3) with probability $1/2$ each.

Remark 9.7. The *split trees* defined by Devroye [35] are related to fragmentation trees. A split tree is a b-ary tree defined using a number of balls that enter the root and are distributed (randomly and recursively) to the subtrees of the root and further down in the tree according to certain rules that are based on a splitting law $\mathbf{V} = (V_1, \dots, V_b)$ satisfying (9.1), see [35] for details. (A splitting law is thus the same as a dislocation law.) Far away from the fringe, where there are many balls and the law of large numbers applies, the numbers of balls in different subtrees are distributed asymptotically as the masses in the corresponding fragmentation tree, so there are many similarities between the two types of random trees. However, at the fringe, the details differ, and the asymptotic fringe distributions are in general not the same. For example, the binary search tree in Example 6.2 can be defined as a split tree, where the splitting law $\mathbf{V} = (V_1, V_2) = (V_1, 1 - V_2)$ with $V_1 \sim U(0, 1)$ uniform. The corresponding fragmentation tree is thus the tree studied in Example 9.4, and as noted there the asymptotic fringe tree distribution is not the same as for the binary search tree; for example, the degree distributions differ. Fringe distributions of split trees will be studied in another paper.

10. Rank

Define, following Bóna and Pittel [19], the *rank* of a node in a rooted tree to be the smallest distance to a descendant that is a leaf. Thus a leaf has rank 0, while a non-leaf has rank $\geqslant 1$. A node with rank $\geqslant k$ is also said to be *k-protected*. (For example, 1-protected = non-leaf; 2-protected = non-leaf and no child is a leaf.) The simplest "non-trivial" case is 2-protected, which sometimes is called just *protected*. There has in recent years been a number of papers on the number of 2-protected nodes in various random trees, or (equivalently) the probability that a random node is 2-protected, and a few papers on k-protected nodes for higher k; see e.g. Devroye and Janson [37] and the references therein. Such results can equivalently be described as results on the distribution of the rank of a random node.

For a tree T (deterministic or random), let $R(T)$ be the rank of a uniformly random node in T, and let $R_0(T)$ be the rank of the root of T. (Thus $R(T)$ is a random variable, while $R_0(T)$ is deterministic if T is.) Since the rank of v depends only on the subtree T^v, $R(T) = R_0(T^*)$, the rank of the root of the random fringe tree T^*. This reduces the study of rank and k-protected nodes to the study of random fringe trees. (This was the method by Devroye and Janson [37], there applied to several classes of random trees, including random recursive trees and binary search trees but also conditioned Galton–Watson trees which are not of the type considered in the present paper.)

For the random trees considered here, Theorem 5.14 applies, for any fixed $k \geqslant 0$, to the property that a node has rank $\geqslant k$ (i.e., is k-protected); we denote this property by $\mathcal{P}_k$ (in this section) and deduce the following. (Note that, depending on one's point of view, (10.1) can be seen both as a limit result for the distribution of the rank, and as a limit result for the proportion of k-protected nodes.)

Theorem 10.1. *Suppose that (A1)–(A5) hold. Then, for any $k \geqslant 0$, as $n \to \infty$,*

$$\mathbb{P}\bigl(R(T_n) \geqslant k \mid T_n\bigr) = \frac{n_{\mathcal{P}_k}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} P_k := \mathbb{P}\bigl(R_0(\overline{\mathcal{T}}) \geqslant k\bigr). \tag{10.1}$$

In other words, the conditioned random variables $\bigl(R(T_n) \mid T_n\bigr)$ converge in distribution a.s. as $n \to \infty$,

$$\mathcal{L}\bigl(R(T_n) \mid T_n\bigr) \xrightarrow{\text{a.s.}} \mathcal{L}\bigl(R_0(\overline{\mathcal{T}})\bigr). \tag{10.2}$$

In particular, the same holds for the unconditioned random variables, i.e.,

$$R(T_n) \xrightarrow{\text{d}} R_0(\overline{\mathcal{T}}). \tag{10.3}$$

Proof. An immediate application of Theorem 5.14. □

10.1. m-ary search tree

We consider here the rank and k-protected nodes in the m-ary search tree in Section 7.2. (The binary case $m = 2$ has been studied by Mahmoud and Ward [90], Bóna [18], Devroye and Janson [37], Bóna and Pittel [19], Holmgren and Janson [59]; the case $m = 3$ by Holmgren and Janson [60] and some higher m by Heimbürger [56].)

We let as in (10.1) $P_k = P_k(m) := \mathbb{P}\bigl(R_0(\overline{\mathcal{T}}) \geqslant k\bigr)$. Thus, by (10.1), the fraction of k-protected nodes in an m-ary search tree T_n converges a.s. to $P_k(m)$. Recall that (for $m \geqslant 3$) the number of nodes $|T_n|$ is random. Hence it is interesting to study not only the fraction of k-protected nodes, but also the (random) number $n_{\mathcal{P}_k}(T_n)$ of them in T_n. (The results are formulated in this way in some of the references above.) We note that, as an immediate consequence of (10.1) and (7.19), see Remark 5.18,

$$\frac{n_{\mathcal{P}_k}(T_n)}{n} \xrightarrow{\text{a.s.}} \frac{P_k(m)}{2(H_m - 1)}. \tag{10.4}$$

Note also that this implies asymptotics for the expectation $\mathbb{E}\, n_{\mathcal{P}_k}(T_n)$, see Remark 5.19.

We proceed to the calculation of the numbers $P_k(m)$. It will be convenient to use the extended m-ary search tree in Section 7.1, but note that we really are interested in the subtree of internal nodes; to emphasize this we say *internally k-protected* for the k-protected nodes in the tree of internal nodes. As usual, m is fixed and will often be omitted from the notation.

With this in mind, define, for $k \geqslant 0$,

$$h_k(t) := \mathbb{P}(\text{the root of } \mathcal{T}_t \text{ is internal and internally } k\text{-protected}). \tag{10.5}$$

The root becomes an internal node at time $S_1 \sim \mathrm{Exp}(1)$, when it receives its first key. Every node is 0-protected, so $h_0(t)$ is the probability that the root is internal; thus

$$h_0(t) = \mathbb{P}(S_1 \leqslant t) = 1 - e^{-t}. \tag{10.6}$$

Recall that all m children of the root are born at time $\xi = S_{m-1}$. Let $f(t)$ be the density function of $\xi = S_{m-1}$. For $k \geqslant 1$, the root of $\mathcal{T}_t$ is internal and internally k-protected if and only if $t > \xi$ and the m children of the root either are external or internally $(k-1)$-protected, but not all external. Conditioned on ξ, with $\xi < t$, the m subtrees of the root of $\mathcal{T}_t$ are independent, and distributed as $\mathcal{T}_{t-\xi}$. Hence, the conditional probability (given ξ) that a given child is internally $(k-1)$-protected is $h_{k-1}(t-\xi)$, while the probability that it is external is $1 - h_0(t-\xi) = e^{-(t-\xi)}$. Consequently,

$$h_k(t) = \int_0^t f(s)\bigl((h_{k-1}(t-s) + e^{-(t-s)})^m - e^{-m(t-s)}\bigr)\,\mathrm{d}s. \tag{10.7}$$

This can be written as a convolution

$$h_k(t) = f(t) * \bigl((h_{k-1}(t) + e^{-t})^m - e^{-mt}\bigr). \tag{10.8}$$

Furthermore, $f(t)$ is the density function of the sum $S_{m-1} = \sum_{i=1}^{m-1} Y_i$ of independent random variables, and thus $f(t)$ is itself a convolution of their densities ie^{-it}. Hence, (10.8) can be written

$$h_k(t) = \bigl((h_{k-1}(t) + e^{-t})^m - e^{-mt}\bigr) * \mathop{*}_{i=1}^{m-1} ie^{-it}. \tag{10.9}$$

This, with the initial (10.6), makes it possible to calculate any $h_k(t)$ by recursion (preferably using computer algebra).

By Theorem 5.14 and (10.5), the fraction of nodes in T_n that are internal and internally k-protected converges a.s. to

$$\mathbb{P}(\text{the root of } \overline{\mathcal{T}} \text{ is internal and internally } k\text{-protected}) = \int_0^\infty h_k(t)e^{-t}\,\mathrm{d}t. \tag{10.10}$$

Recall that we are really interested in the internal nodes only. By (7.13), asymptotically half of the nodes in the extended m-ary search tree are internal, and thus the fraction of k-protected nodes in the (internal) m-ary search tree T_n converges a.s. to

$$P_k = P_k(m) := 2\int_0^\infty h_k(t)e^{-t}\,\mathrm{d}t. \tag{10.11}$$

This can be regarded as a value of a Laplace transform. Since the Laplace transform transforms convolutions to multiplications, and $\int_0^\infty f(t)e^{-t}\,\mathrm{d}t = \mathbb{E}\,e^{-S_{m-1}} = 1/m$ by (7.9), we have by (10.8) also the alternative formula (that might be better for numerical calculations)

$$P_k = \frac{2}{m}\int_0^\infty e^{-t}\bigl((h_{k-1}(t) + e^{-t})^m - e^{-mt}\bigr)\,\mathrm{d}t. \tag{10.12}$$

In the binary case $m = 2$, these formulas are equivalent to the formulas derived by a very similar argument in Devroye and Janson [37]. (The function

$r_k(t)$ in [37] equals $1-e^{-t}-h_k(t)$.) For a different method to find $P_k(2)$, see Bóna [18] and Bóna and Pittel [19].

Consequently, Theorem 10.1 says that for an m-ary search tree the asymptotic distribution of the rank is

$$\mathbb{P}(R(T_n)=k \mid T_n) \xrightarrow{\text{a.s.}} \mathbb{P}(R_0(\overline{T})=k) = P_k - P_{k+1}, \tag{10.13}$$

where P_k is given by (10.11)–(10.12).

Note that (by induction), each $h_k(t)$ is a polynomial in e^{-t} and t with rational coefficients. Hence, each $P_k(m)$ is a rational number.

Trivially, $P_0=1$ for every m by (10.1).

For $k=1$, $h_0(t)+e^{-t}=1$ by (10.6), and thus by (10.12),

$$P_1 = \frac{2}{m}\int_0^\infty e^{-t}(1-e^{-mt})\,\mathrm{d}t = \frac{2}{m}\Bigl(1-\frac{1}{m+1}\Bigr) = \frac{2}{m+1}, \tag{10.14}$$

in accordance with (7.24) (recall that the 1-protected nodes are precisely the non-leaves).

Also for P_2, we may deduce a rather simple formula.

Lemma 10.2. *For the m-ary search tree, with $m \geqslant 2$,*

$$h_1(t) = (1-e^{-t})^m. \tag{10.15}$$

Proof. We extend the notation of Section 7.1 and let $S_j := \sum_{i=1}^j Y_i$ for any integer $j \geqslant 0$, where $Y_i \sim \mathrm{Exp}(i)$ are independent. (We thus change the earlier special definition of S_m.) By Theorem C.1,

$$S_j \overset{\mathrm{d}}{=} V_{j,j} = \max_{1\leqslant i\leqslant j} E_i, \tag{10.16}$$

where $E_1, E_2, \dots, E_j \sim \mathrm{Exp}(1)$ are i.i.d.

Since $h_1(t)+e^{-t}=1$ by (10.6), (10.8) yields, recalling that f is the density function of S_{m-1} and using (10.16),

$$\begin{aligned} h_1(t) &= f(t) * (1-e^{-mt}) = f(t) * \mathbb{P}(Y_m \geqslant t) = \int_0^t f(s)\,\mathbb{P}(Y_m \leqslant t-s)\,\mathrm{d}s \\ &= \mathbb{P}(S_{m-1}+Y_m \leqslant t) = \mathbb{P}(S_m \leqslant t) = \mathbb{P}\bigl(\max_{1\leqslant i\leqslant m} E_i \leqslant t\bigr) \\ &= \prod_{i=1}^m \mathbb{P}(E_i \leqslant t) = (1-e^{-t})^m. \end{aligned}$$

□

Theorem 10.3. *The asymptotic probability that a random node in an m-ary search tree T_n is 2-protected is*

$$P_2 = P_2(m) = \frac{2}{m}\sum_{\ell=0}^{m-1} \frac{m!}{(m-\ell)!}\cdot\frac{(m(m-\ell))!}{(m(m-\ell)+\ell+1)!}. \tag{10.17}$$

Proof. By (10.12), a binomial expansion, (10.15), the change of variables $x = e^{-t}$ and a standard evaluation of a beta integral,

$$\begin{aligned}
P_2 &= \frac{2}{m}\int_0^\infty e^{-t}\sum_{\ell=0}^{m-1}\binom{m}{\ell}h_1(t)^{m-\ell}e^{-\ell t}\,\mathrm{d}t \\
&= \frac{2}{m}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\int_0^\infty (1-e^{-t})^{m(m-\ell)}e^{-(\ell+1)t}\,\mathrm{d}t \\
&= \frac{2}{m}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\int_0^1 (1-x)^{m(m-\ell)}x^{\ell}\,\mathrm{d}t \qquad (10.18)\\
&= \frac{2}{m}\sum_{\ell=0}^{m-1}\binom{m}{\ell}B\bigl(m(m-\ell)+1,\ell+1\bigr) \\
&= \frac{2}{m}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\frac{(m(m-\ell))!\,\ell!}{(m(m-\ell)+\ell+1)!}.
\end{aligned}$$

□

Remark 10.4. We can also prove this result using a more combinatorial proof with balls and boxes. We recall from Theorem 7.10 that the asymptotic fringe tree $\overline{\mathcal{T}}$ can be constructed as an m-ary search tree with a random number K of keys, where by (7.18) $\mathbb{P}(K=k) = \frac{2}{(k+1)(k+2)} = 1/\binom{k+2}{2}$, $k \geqslant 1$. We condition on $K = k$ and find the probability that the root of T_k is 2-protected.

Recall that a node is 2-protected if it is not a leaf and has no child that is a leaf. Thus, the root of T_k is 2-protected if and only if it is filled with $m-1$ keys and each of the m subtrees of the root has the property that it is either empty or contains at least m keys, and at least one of the subtrees is nonempty; in particular, we must have $k \geqslant 2m-1$.

For $k \geqslant 2m-1$ we order the keys in increasing order and represent the $k-m+1$ keys that are distributed to the m subtrees of the root by 0's and the $m-1$ keys that stay in the root by 1's. We also add two additional 1's first and last. This gives a string of length $k+2$, beginning and ending with 1, and with $m-1$ additional 1's. There are $\binom{k}{m-1}$ such strings, and all occur with the same probability.

Furthermore, the corresponding tree T_k is 2-protected if and only if between every pair of 1's in the string, there is either no 0's, or at least m 0's. In other words, the 1's appear in clusters, separated by at least m 0's. Let the number of clusters be $r+1$, and note that $1 \leqslant r \leqslant m$. To count the number of strings of length $k+2$ such that these properties are satisfied for a given r, we first distribute the $m+1$ 1's into $r+1$ boxes, such that no box is empty. This gives $\binom{m}{r}$ different choices. We then distribute the $k-m+1$ 0's into the r gaps between the clusters; it is required that there should be at least m 0's in each gap, but the remaining $k-m+1-mr$ 0's can be distributed arbitrarily into the r gaps. This can be done in $\binom{k+r-m(r+1)}{r-1}$ ways. Hence, summing over r and

k and using (7.18), we obtain

$$P_2 = P_2(m) = \sum_{r=1}^{m} \sum_{k=m(r+1)-1}^{\infty} \frac{\binom{m}{r}\binom{k+r-m(r+1)}{r-1}}{\binom{k+2}{2}\binom{k}{m-1}}. \tag{10.19}$$

The sum over k can be written as a hypergeometric sum (B.1), which using Gauss' formula (B.2) simplifies and yields

$$\begin{aligned} P_2(m) &= \sum_{r=1}^{m} 2\binom{m}{r} \frac{(m-1)!\,(mr)!}{(mr+m+1)!} F(mr+1, r; mr+m+2; 1) \\ &= 2\sum_{r=1}^{m} \frac{(m-1)!\,(mr)!}{r!\,(mr+m-r+1)!}, \end{aligned} \tag{10.20}$$

which is equivalent to (10.17) by letting $r = m - \ell$.

We can find the asymptotics of $P_2(m)$ as $m \to \infty$ from (10.17).

Theorem 10.5. *As $m \to \infty$, the probability $P_2 = P_2(m)$ in Theorem 10.3 is*

$$P_2(m) = \frac{2}{m^3} + O\bigl(m^{-4}\bigr). \tag{10.21}$$

Proof. Write (10.17) as

$$P_2(m) = \frac{2}{m} \sum_{\ell=0}^{m-1} a_\ell(m) \tag{10.22}$$

with

$$a_\ell(m) = \frac{m!}{(m-\ell)!} \cdot \frac{(m(m-\ell))!}{(m(m-\ell)+\ell+1)!}. \tag{10.23}$$

In particular,

$$a_0(m) = \frac{(m^2)!}{(m^2+1)!} = \frac{1}{m^2+1}. \tag{10.24}$$

For each fixed ℓ, as $m \to \infty$,

$$a_\ell(m) \sim m^\ell (m(m-\ell))^{-\ell-1} \sim m^\ell m^{-2(\ell+1)} = m^{-\ell-2}. \tag{10.25}$$

Furthermore, rather crudely, if $\ell \leqslant m/2$, then

$$a_\ell(m) \leqslant m^\ell \frac{1}{(m(m-\ell))^{\ell+1}} \leqslant \frac{m^\ell}{(m^2/2)^{\ell+1}} = \frac{2^{\ell+1}}{m^{\ell+2}} \tag{10.26}$$

and thus

$$\sum_{\ell=1}^{m/2} a_\ell(m) \leqslant \frac{2^2}{m^3}\Bigl(1 - \frac{2}{m}\Bigr)^{-1} = O\bigl(m^{-3}\bigr). \tag{10.27}$$

If $m/2 < \ell \leqslant m-1$, we instead note that $\ell - m/4 + O(1) \geqslant m/4 + O(1)$ of the factors $(m-i)$, $i = 0, \dots, \ell-1$, are less than $\frac{3}{4}m$, and thus, similarly to (10.26),

$$a_\ell(m) \leqslant \Big(\frac{3}{4}\Big)^{m/4+O(1)} m^\ell \frac{1}{(m(m-\ell))^{\ell+1}} \leqslant \Big(\frac{3}{4}\Big)^{m/4+O(1)} \tag{10.28}$$

and thus

$$\sum_{\ell=\lfloor m/2\rfloor+1}^{m-1} a_\ell(m) = O\left(m\Big(\frac{3}{4}\Big)^{m/4}\right) = O\big(m^{-3}\big). \tag{10.29}$$

Consequently, by (10.24), (10.27) and (10.29),

$$\sum_{\ell=0}^{m-1} a_\ell(m) = a_0(m) + O\big(m^{-3}\big) = \frac{1}{m^2} + O\big(m^{-3}\big) \tag{10.30}$$

and the result follows from (10.22). □

Remark 10.6. The proof shows that $\sum_{\ell=k+1}^{m-1} a_\ell(m) = O\big(m^{-k-3}\big)$ for any fixed k, and thus

$$P_2(m) = \frac{2}{m}\sum_{\ell=0}^{k} a_\ell(m) + O\big(m^{-k-4}\big) \tag{10.31}$$

which together with (10.23) gives an asymptotic expansion of $P_2(m)$ in powers of m^{-1} to arbitrary precision. For example, taking $k = 2$,

$$\begin{aligned} P_2(m) &= \frac{2}{m}\Big(\frac{1}{m^2+1} + \frac{m}{\langle m^2-m+1\rangle_2} + \frac{m(m-1)}{\langle m^2-2m+1\rangle_3}\Big) + O\big(m^{-6}\big) \\ &= \frac{2}{m^3} + \frac{2}{m^4} + \frac{4}{m^5} + O\big(m^{-6}\big). \end{aligned} \tag{10.32}$$

We give some numerical examples for small m and k, calculated by `Maple` using (10.11), (10.12) or (for $k = 2$) (10.17). Recall that $P_0(m) = 1$ and $P_1(m) = 2/(m+1)$ by (10.14). The value $P_2(2)$ was first found by Mahmoud and Ward [90]. Bóna [18] found also $P_3(2)$ and $P_4(2)$ in an equivalent form; in our notation he computed $\mathbb{P}(R_0(\overline{\mathcal{T}}) = k)$ for $k \leqslant 4$; this was extended to $k \leqslant 6$ by Bóna and Pittel [19], see also Devroye and Janson [37]. For $m > 2$, $P_2(m)$ was found using Pólya urns for $m = 3$ by Holmgren and Janson [60] and for $m = 4, 5, 6, 7$ by Heimbürger [56]. The values of $P_3(3)$ and $P_3(4)$ are new. (We have also calculated e.g. $P_4(3)$ and $P_3(5)$, but they have too many digits to fit on a line.)

$$P_2(2) = \tfrac{11}{30}$$
$$P_3(2) = \tfrac{1249}{8100}$$
$$P_4(2) = \tfrac{103365591157608217}{2294809143026400000}$$
$$P_5(2) = \tfrac{2898843211947012642874550347245023104911370489425501083914785267 7}{335337702502244919985290072567096006741828080379723178828800000000}$$
$$P_2(3) = \tfrac{19}{140}$$
$$P_3(3) = \tfrac{1550707922167467531619}{1091712188392817191200000}$$

$$P_2(4) = \frac{54731}{1021020}$$
$$P_3(4) = \frac{1866929360927367184832939100272472707820442065482742871065157101180 1}{10647619779410811926633805505705710276401405786992612983909862200000000}$$
$$P_2(5) = \frac{3491}{145860}$$
$$P_2(6) = \frac{760687429}{61618399920}$$
$$P_2(7) = \frac{30706935422593}{4254556530624400}$$
$$P_2(8) = \frac{351290243384177}{75964420635503400}$$
$$P_2(9) = \frac{4396518809902327}{1391843538882476680}$$
$$P_2(10) = \frac{2342020375280850167304355177}{10364894888849114175198337 10040}$$
$$P_2(13) = \frac{26484106080648861012732023951265705943439}{26631299618409351014679549721918492033886700}$$
$$P_2(15) = \frac{33452413685832796193050304618778894634092293794795491 49}{52374998662930022993158026783022806524589397098678503376 00}$$
$$P_2(18) = \frac{291057072374356381926366502629609543999495527334257770304419581448683067}{79843850145351896473219324099483440967778321003331265016051521391306531640 0}$$

The numerators and denominators of these rational numbers evidently grow very rapidly with k; Bóna and Pittel [19] note that (in our notation) the denominator of $P_6(2)$ has 274 digits, but the largest prime factor is only 61, and they show that in general, the largest prime factor of the denominator of $P_k(2)$ is at most $2^k + 1$. This can be generalized to arbitrary m, using the recursion above; this also gives a new and simpler proof for $m = 2$. (Nothing similar seems to hold for the numerators; they typically have only a few and often large prime factors in these examples. The numerator of $P_2(13)$ happens to be a prime with 41 digits.)

Theorem 10.7. *The largest prime factor of the denominator of $P_k(m)$ is at most $m^k + 1$, for any $k \geqslant 1$ and $m \geqslant 2$.*

Proof. A simple calculation shows that for integers $j, a, b \geqslant 0$ with $a \neq b$, the convolution $t^j e^{-at} * e^{-bt}$ is of the form $\sum_{i=0}^{j} c_i t^i e^{-at} + c' e^{-bt}$ with coefficients $c_i, c' \in (b-a)^{-j-1}\mathbb{Z}$; if $a = b$, we obtain instead $\frac{1}{j+1} t^{j+1} e^{-at}$. It follows by (10.9), (10.15) and induction that for $k \geqslant 1$, $h_k(t)$ is a polynomial in e^{-t} and t of degree m^k in e^{-t} and of degree (at most) $1 + m + \cdots + m^{k-2} = (m^{k-1} - 1)/(m - 1)$ in t, with rational coefficients whose denominators have all their prime factors $< m^k$. The result then follows from (10.11). □

For the binary case $m = 2$, the probabilities $P_k(2)$ where shown to have an exponential decay by Bóna and Pittel [19]. We conjecture that this holds for $m \geqslant 3$ too, but leave that as an open problem.

10.2. Random recursive tree

Consider the random recursive tree in Example 6.1. This has been studied by Mahmoud and Ward [91] and Devroye and Janson [37]; we follow here [37]. Let $\mathcal{T}_t$ be the Yule tree process in Example 6.1 and define

$$p_k(t) := \mathbb{P}(R_0(\mathcal{T}_t) \geqslant k), \tag{10.33}$$

the probability that the root of $\mathcal{T}_t$ is k-protected. By the construction of the fringe tree $\overline{\mathcal{T}} = \mathcal{T}_{\overline{\tau}}$, with $\overline{\tau} \sim \mathrm{Exp}(1)$, the limit P_k in (10.1) is given by

$$P_k = \int_0^\infty p_k(t) e^{-t} \, \mathrm{d}t. \tag{10.34}$$

The functions $p_k(t)$ can, in principle, be found by recursion. The children of the root in $\mathcal{T}_t$ arrive according to a Poisson process with intensity 1, and a child that is born at time $s \leqslant t$ is not $(k-1)$-protected at time t with probability $1 - p_{k-1}(t-s)$. Hence, for any $k \geqslant 1$, the number of children of the root in $\mathcal{T}_t$ that are not $(k-1)$-protected at time t is Poisson distributed with mean $\int_0^t \bigl(1 - p_{k-1}(t-s)\bigr) \, \mathrm{d}s = \int_0^t \bigl(1 - p_{k-1}(s)\bigr) \, \mathrm{d}s$. Since the root is k-protected if and only if there is no such child, but there is at least one child, and the probability that there is no child at all is e^{-t}, we obtain the recursion

$$\begin{aligned} p_k(t) &= \exp\left(- \int_0^t \bigl(1 - p_{k-1}(s)\bigr) \, \mathrm{d}s \right) - e^{-t} \\ &= e^{-t} \left(\exp\left(\int_0^t p_{k-1}(s) \, \mathrm{d}s \right) - 1 \right), \qquad k \geqslant 1, \, t \geqslant 0, \end{aligned} \tag{10.35}$$

with $p_0(t) = 1$.

Taking $k = 1$ in (10.35) we obtain the obvious $p_1(t) = 1 - e^{-t}$. Taking $k = 2$ in (10.35), we then find

$$p_2(t) = \exp\bigl(e^{-t} - 1\bigr) - e^{-t} \tag{10.36}$$

and thus by (10.34)

$$\begin{aligned} P_2 &= \int_0^\infty \exp\bigl(e^{-t} - 1\bigr) e^{-t} \, \mathrm{d}t - \int_0^\infty e^{-2t} \, \mathrm{d}t \\ &= \int_0^1 \exp(x - 1) \, \mathrm{d}x - \frac{1}{2} = \frac{1}{2} - e^{-1}, \end{aligned} \tag{10.37}$$

in accordance with Mahmoud and Ward [91].

In principle, the recursion (10.35) yields $p_k(t)$ and P_k for larger k too, but we do not know any closed form for $k \geqslant 3$.

11. Maximal clades in m-ary search trees

We define a *clade* in an m-ary tree to be a node with less than m children. (In the formulation using extended m-ary search trees with external nodes, a clade is thus a node with at least one external child.) A *maximal clade* is a clade such that no ancestor is a clade.

Remark 11.1. The reason for this somewhat strange terminology comes from applications of the binary case $m = 2$ to mathematical biology, where the clade is regarded as a set of external nodes, see e.g. Blum and François [16], Durand,

Blum and François [45], Chang and Fuchs [23], Durand and François [46], Drmota, Fuchs and Lee [42] and (for the elementary equivalence with the definition here) Janson [71]. We consider here the natural extension to m-ary trees. (As a mathematically interesting example; we do not claim any biological applications.)

The number of clades is thus the number of nodes with outdegree less than m, and the fraction of such nodes is by Theorem 5.14 asymptotically given by the probability that the root of the asymptotic fringe tree $\overline{\mathcal{T}}$ has outdegree less than m. (This is found to be $1-\frac{2}{m(m+1)}$ in Theorem 7.14.)

The property that a clade is maximal, however, depends also on its ancestors, and therefore we need the extended fringe and the random sin-tree $\widetilde{\mathcal{T}}$; moreover, we have to consider all ancestors, so Theorem 5.25 does not apply and we use Theorem 5.26.

Theorem 11.2. *Let $n_{\mathsf{mc}}(T)$ be the number of maximal clades in T. If T_n is a random m-ary search tree with n keys, then*

$$\frac{n_{\mathsf{mc}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} P_{\mathsf{mc}} = P_{\mathsf{mc}}(m) = \mathbb{P}\bigl(o \text{ is a maximal clade in } \widetilde{\mathcal{T}}\bigr). \tag{11.1}$$

Proof. We apply Theorem 5.26 with $\mathcal{P}_0 = \mathcal{Q} =$ "the outdegree is $< m$", i.e., the property that a node is a clade. Then $\mathcal{P}$ in Theorem 5.26 is the property that a node is a maximal clade. The assumption (5.49) holds trivially, since $\widehat{\Xi}(\alpha) \leqslant m$. The random variable Λ is the time the root of $\mathcal{T}_t$ gets it final child; by Remark 7.9, this can be written as a sum of a number of exponential variables (with different rates), and thus (5.50) holds for some small $\delta > 0$. (In fact, for all $\delta < 1$, by Remark 7.9 and Theorem C.1.)

Hence, Theorem 5.26 applies and the result follows. □

The constant $P_{\mathsf{mc}}(2)$, i.e., the asymptotic proportion of maximal clades in a binary search tree, was found to be $(1-e^{-2})/4$ by Durand and François [46], see also [42] and [71]. We give a different proof of this, using the properties of the sin-tree $\widetilde{\mathcal{T}}$ in Example 6.2.

Theorem 11.3.

$$P_{\mathsf{mc}}(2) = \frac{1-e^{-2}}{4}. \tag{11.2}$$

Proof. Recall the general construction of the sin-tree $\widetilde{\mathcal{T}}$ in Section 5 and the specific version for the binary search tree in Example 6.2. In the construction, we stop the tree at $\overline{\tau} \sim \mathrm{Exp}(1)$, but we first consider the tree $\widetilde{\mathcal{T}}_t$ at a fixed time $t \geqslant 0$. (Equivalently, we condition on $\overline{\tau} = t$.)

We thus want to compute the probability $\mathbb{P}\bigl(o \text{ is a maximal clade in } \widetilde{\mathcal{T}}_t\bigr)$. We first note that o is a clade unless it already has got its two children; each child has appeared with probability $1-e^{-t}$ and thus

$$\mathbb{P}\bigl(o \text{ is a clade in } \widetilde{\mathcal{T}}_t\bigr) = 1-(1-e^{-t})^2 = 2e^{-t}-e^{-2t}. \tag{11.3}$$

We also require that no ancestor is a clade, i.e., that each ancestor has two children. Note that each ancestor has an heir, so it is not a clade if and only if the other child is not yet born. Suppose that the ancestors are born at times $-\eta_1, -\eta_2, \dots$, and condition on these times. Ancestor $o^{(i)}$ thus has age $\eta_i + t$ at time t, so the probability that it is not a clade is $e^{-(\eta_i+t)}$. Consequently, using the independence of different parts of the sin-tree,

$$\mathbb{P}\bigl(o \text{ is a maximal clade in } \widetilde{\mathcal{T}}_t \mid \eta_1, \dots\bigr) = \bigl(2e^{-t} - e^{-2t}\bigr) \prod_{i=1}^{\infty} \bigl(1 - e^{-(\eta_i+t)}\bigr). \tag{11.4}$$

The next step is to find the expectation of (11.4) over all $\{\eta_i\}$. In the present case, this is not difficult since, by Example 6.2, $\{-\eta_i\}$ is a Poisson process with intensity 2 on $(-\infty, 0)$, and thus $\{\eta_i\}$ is a Poisson process with intensity 2 on $(0, \infty)$. For any Poisson process $\Xi = \{\xi_i\}$ on some space $\mathcal{S}$, with intensity measure λ, and any function f on $\mathcal{S}$ with $0 \leqslant f(x) \leqslant 1$, there is a standard formula

$$\mathbb{E} \prod_{\xi \in \Xi} f(\xi) = e^{-\int_{\mathcal{S}} (1 - f(x)) \lambda(\mathrm{d}x)}. \tag{11.5}$$

(See e.g. [75, Lemma 12.2]. Or note that (11.5) follows easily if f takes only a finite number of values, and the general case follows by monotone convergence.) Consequently, taking $f(x) = 1 - e^{-x-t}$, (11.4) yields

$$\begin{aligned} \mathbb{P}\bigl(o \text{ is a maximal clade in } \widetilde{\mathcal{T}}_t\bigr) &= \bigl(2e^{-t} - e^{-2t}\bigr)\, \mathbb{E} \prod_{i=1}^{\infty} f(\eta_i) \\ &= \bigl(2e^{-t} - e^{-2t}\bigr) e^{-\int_0^\infty e^{-x-t} 2\, \mathrm{d}x} \\ &= \bigl(2e^{-t} - e^{-2t}\bigr) e^{-2e^{-t}}. \end{aligned} \tag{11.6}$$

Finally, recalling that $\widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}_{\overline{\tau}}$ with $\overline{\tau} \sim \mathrm{Exp}(1)$,

$$\begin{aligned} \mathbb{P}\bigl(o \text{ is a maximal clade in } \widetilde{\mathcal{T}}\bigr) &= \int_0^\infty \bigl(2e^{-t} - e^{-2t}\bigr) e^{-2e^{-t}} e^{-t}\, \mathrm{d}t \\ &= \int_0^1 \bigl(2x - x^2\bigr) e^{-2x}\, \mathrm{d}x = \frac{1}{4} - \frac{1}{4} e^{-2}. \end{aligned} \tag{11.7}$$

□

For further, somewhat surprising, results on the number of maximal clades in the binary case (moments and asymptotic distribution), see Drmota, Fuchs and Lee [42] and Janson [71].

Problem 11.4. Unfortunately, we do not know how to compute $P_{\mathsf{mc}}(m)$ for $m > 2$, and we leave this as an open problem. Using the description in Section 7.2 of $\widetilde{\mathcal{T}}$, it is straightforward to modify (11.3)–(11.4) (although the result is more complicated since the birth times do not have exponential distributions, see Remark 7.9), but the birth times of the ancestors do not form a Poisson process so (11.5) does not apply and we do not know how compute the expectation.

We can use the same method for other, related, problems. We give two examples. Let us first consider again the binary search tree, but we simplify the property of being a maximal clade studied above by considering only the condition for the ancestors but ignoring the number of children. Thus, let $\mathcal{P}_{\mathsf{x}}$ be the property of a node that none of its ancestors has only one child, and let $n_{\mathsf{x}}(T)$ be the number of nodes in T with this property.

Theorem 11.5. *If T_n is a random binary search tree with n keys, then*

$$\frac{n_{\mathsf{x}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \frac{1-e^{-2}}{2}. \tag{11.8}$$

Proof. We apply Theorem 5.26 as in the proof of Theorem 11.2, but with $\mathcal{P}_0$ the trivial property "true". This yields convergence almost surely, to the limit $\mathbb{P}(o \text{ has } \mathcal{P}_{\mathsf{x}} \text{ in } \widetilde{\mathcal{T}})$. This probability is computed as in the proof of Theorem 11.3, replacing the factor (11.3) by 1, which yields the result

$$\int_0^\infty e^{-2e^{-t}} e^{-t}\,\mathrm{d}t = \int_0^1 e^{-2x}\,\mathrm{d}x = \frac{1}{2} - \frac{1}{2}e^{-2}. \tag{11.9}$$

□

The property $\mathcal{P}_{\mathsf{x}}$, as formulated above, can be studied also in other trees. We consider the random recursive tree as a different simple example.

Theorem 11.6. *If T_n is a random recursive tree with n keys, then*

$$\frac{n_{\mathsf{x}}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} 1-e^{-1}. \tag{11.10}$$

Proof. We argue as in the proof of Theorem 11.5, now using the description of the sin-tree $\widetilde{\mathcal{T}}$ in Example 6.1. In this sin-tree, the ancestors form a Poisson process with intensity 1 on $(-\infty, 0)$, and, as in the binary search tree case, for an ancestor, the time until birth of the first non-heir is Exp(1). Hence the limit $\mathbb{P}(o \text{ has } \mathcal{P}_{\mathsf{x}} \text{ in } \widetilde{\mathcal{T}})$ can be calculated by the method above, now yielding, cf. (11.9),

$$\int_0^\infty e^{-e^{-t}} e^{-t}\,\mathrm{d}t = \int_0^1 e^{-x}\,\mathrm{d}x = 1-e^{-1}. \tag{11.11}$$

□

12. Restricted sampling and sampling by a random key

We have so far considered the properties of a random node in the tree T_n. As pointed out by Jagers and Nerman [65], one can similarly obtain results for a random node sampled with some restriction. (For example, a random leaf, a random non-leaf, a random node with no sibling, ...)

In general, let $\mathcal{Q}$ be a property of the type in Theorem 5.14 or 5.25 and sample v uniformly among all nodes in T_n that satisfy $\mathcal{Q}$. If $\mathcal{P}$ is another such property, then, by Theorem 5.25,

$$\mathbb{P}(v \text{ has } \mathcal{P} \mid T_n) = \frac{n_{\mathcal{P}\wedge\mathcal{Q}}(T_n)}{n_{\mathcal{Q}}(T_n)} \xrightarrow{\text{a.s.}} \frac{p_{\mathcal{P}\wedge\mathcal{Q}}(\widetilde{\mathcal{T}})}{p_{\mathcal{Q}}(\widetilde{\mathcal{T}})}. \tag{12.1}$$

If we let $\widetilde{\mathcal{T}}_{\mathcal{Q}}$ denote $\widetilde{\mathcal{T}}$ conditioned on $o \in \mathcal{Q}$, then we can write (12.1) as

$$\mathbb{P}(v \text{ has } \mathcal{P} \mid T_n) = \frac{n_{\mathcal{P}\wedge\mathcal{Q}}(T_n)}{n_{\mathcal{Q}}(T_n)} \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\widetilde{\mathcal{T}}_{\mathcal{Q}}). \tag{12.2}$$

If Theorem 5.14 applies, we can replace $\widetilde{\mathcal{T}}$ by $\overline{\mathcal{T}}$ in (12.1)–(12.2).

Example 12.1. We have already seen an example of this in the first suggested proof of Theorem 7.13, where we note that sampling a node uniformly in an m-ary search tree is the same as sampling an internal node uniformly in the corresponding extended m-ary search tree. Thus T_n is the extended m-ary search tree and $\mathcal{Q}$ is "internal". Furthermore, $\mathcal{P} = \mathcal{P}_k$ is the property of having exactly k keys. (In this example, $\mathcal{Q}$ is the complement of $\mathcal{P}_0$, so $\mathcal{P}_0 \wedge \mathcal{Q}$ is the empty property while $\mathcal{P}_k \wedge \mathcal{Q} = \mathcal{P}_k$ for $k \geqslant 1$.)

Let us consider the example of sampling a random leaf v in more detail. Of course, the fringe tree T^v rooted at v is trivial, so the interest is in the extended fringe and in properties of the type in Theorem 5.25. For example, Drmota, Gittenberger, Panholzer, Prodinger and Ward [43] study the number of internal nodes (and the number of leaves) in the subtree rooted at the father of a randomly chosen leaf, for a variety of different types of random trees.

We have the following general result.

Theorem 12.2. *Suppose that (A1)–(A5) hold, and that $\mathcal{P}$ is a property as in Theorem 5.25. If v is a uniformly random leaf in T_n, then*

$$\mathbb{P}(v \text{ has } \mathcal{P} \mid T_n) \xrightarrow{\text{a.s.}} p_{\mathcal{P}}(\widetilde{\mathcal{T}}_{\mathrm{leaf}}), \tag{12.3}$$

where $\widetilde{\mathcal{T}}_{\mathrm{leaf}}$ is $\widetilde{\mathcal{T}}$ conditioned on o being a leaf.

The random sin-tree $\widetilde{\mathcal{T}}_{\mathrm{leaf}}$ may be constructed directly from the tree process $(\widetilde{\mathcal{T}}_t)$ in Section 5 by removing all descendants of o and stopping at a random time τ_o with the density function

$$\frac{e^{-\alpha t}\,\mathbb{P}(\xi_1 > t)}{\int_0^\infty e^{-\alpha t}\,\mathbb{P}(\xi_1 > t)\,\mathrm{d}t}, \qquad t > 0, \tag{12.4}$$

where ξ_1 is the time of birth of the first child of an individual in the branching process. In particular, if $\xi_1 \sim \mathrm{Exp}(a)$ for some $a > 0$, then $\tau_o \sim \mathrm{Exp}(a + \alpha)$.

Proof. Let $\mathcal{Q}$ be the property of a node that it is a leaf. Then (12.3) is the same as (12.2), with $\widetilde{\mathcal{T}}_{\mathrm{leaf}} = \widetilde{\mathcal{T}}_{\mathcal{Q}}$, i.e., $\widetilde{\mathcal{T}}$ conditioned on o being a leaf.

To see that $\widetilde{\mathcal{T}}_{\mathrm{leaf}}$ can be constructed as stated, note that in the construction of the tree process $(\widetilde{\mathcal{T}}_t)$ in Section 5, the descendants of o and the rest of the tree are independent. Since $\widetilde{\mathcal{T}}$ is obtained by stopping $\widetilde{\mathcal{T}}_t$ at $\overline{\tau}$, it follows that if we ignore descendants of o, $\widetilde{\mathcal{T}}_{\mathrm{leaf}}$ is obtained by stopping $\widetilde{\mathcal{T}}_t$ at an independent random time τ_o having the distribution of $\overline{\tau}$ conditioned on o being a leaf in $\widetilde{\mathcal{T}}_{\overline{\tau}}$. Moreover, if the first child of o is born at ξ_1, then o is a leaf in $\widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}_{\overline{\tau}}$ if and only if $\xi_1 > \overline{\tau}$. Since $\overline{\tau}$ has the density function $\alpha e^{-\alpha t}$, it follows that conditioned on the event $\xi_1 > \overline{\tau}$, $\overline{\tau}$ has the density function (12.4). □

Example 12.3. Let T_n be an extended binary search tree with n internal nodes, let v be a (uniformly) randomly chosen external node in T_n and let X_n be the number of internal nodes in $T_n^{v,-1}$, i.e., the number of internal nodes that are descendants of the parent of v. Equivalently, if v' denotes the sister of v, X_n is 1 plus the number of internal nodes in the subtree T_n^v. It follows from Theorem 12.2 that if X similarly is the number of internal nodes of $\widetilde{\mathcal{T}}_{\mathrm{leaf}}$ that are descendants of the parent $o^{(1)}$ of o, then $\mathbb{P}(X_n = k \mid T_n) \xrightarrow{\text{a.s.}} \mathbb{P}(X = k)$ for every $k \geqslant 1$, i.e., $X_n \xrightarrow{\text{d}} X$, also conditioned on T_n in the sense $\mathcal{L}(X_n \mid T_n) \xrightarrow{\text{a.s.}} \mathcal{L}(X)$, cf. Remark 4.1.

For the extended binary search tree, we use the branching process in Section 7.1 (with $m = 2$), where each individual gets 2 children at the same time, at age $\xi_1 \sim \mathrm{Exp}(1)$. Hence, the last statement of Theorem 12.2 applies with $a = 1$. Furthermore, $\alpha = 1$, and thus $\tau_o \sim \mathrm{Exp}(2)$.

Since X equals 1 plus the number of internal nodes in the subtree $\widetilde{\mathcal{T}}_{\mathrm{leaf}}^{o'}$ rooted at the sister o' of o, we do not have to consider the ancestor $o^{(1)}$; we just note that o' and o are twins, and thus o' too is born at time 0. The number of internal nodes in $\widetilde{\mathcal{T}}_t^{o'}$ is a pure birth process with birth rates $\lambda_k = k + 1$, started at 0. (In other words, it is $\mathcal{Y}_t - 1$, where $\mathcal{Y}_t$ is a Yule process, see Example A.3.) Stopping this at $\tau_o \sim \mathrm{Exp}(2)$ we find by Theorem A.5 (with $\chi = \rho = 1$, $\alpha = 2$) $X - 1 \sim \mathsf{HG}(1, 1; 4)$. In other words, using (B.3)–(B.5),

$$\mathbb{P}(X = k) = \frac{4}{(k+1)(k+2)(k+3)}, \qquad k \geqslant 1. \tag{12.5}$$

This limit distribution was found by Drmota, Gittenberger, Panholzer, Prodinger and Ward [43] (in a slightly different setting, keeping track of the position of the nodes).

Example 12.4. Let T_n be a random recursive tree with n nodes, as in Example 6.1, let v be a uniformly random leaf in T_n and let $\widehat{T}_n^1 := T_n^{v,-1} \setminus \{v\}$, the random tree consisting of the parent $v^{(1)}$ of v and all its decendants except v. Further, let $X_n := |\widehat{T}_n^1|$, the number of nodes that are descendants of the parent of v, including the parent but excluding v.

As in Example 12.3, it follows from Theorem 12.2 that if X similarly is the number of nodes in $\widehat{T}^1 := \widetilde{\mathcal{T}}_{\mathrm{leaf}}^{o,-1} \setminus \{o\}$, then $\mathbb{P}(X_n = k \mid T_n) \xrightarrow{\text{a.s.}} \mathbb{P}(X = k)$ for every $k \geqslant 1$, and thus $\mathcal{L}(X_n \mid T_n) \xrightarrow{\text{a.s.}} \mathcal{L}(X)$, cf. Remark 4.1.

By the description at the end of Example 6.1, the tree process $\widetilde{\mathcal{T}}_{\mathrm{leaf},t}^{o,-1}$, minus o and its descendants, is a Yule tree process, starting at time $-\widetilde{\xi}^*$ when the parent $o^{(1)}$ is born. Furthermore, $\widetilde{\xi}^* \sim \mathrm{Exp}(1)$. We stop this tree process at time τ_o, where by the last statement in Theorem 12.2 (with $a = \alpha = 1$), $\tau_o \sim \mathrm{Exp}(2)$. Consequently, $X \overset{\text{d}}{=} \mathcal{Y}_{\widetilde{\xi}^* + \tau_o}$, a Yule process (started at time 0 as usual) stopped at $\widetilde{\xi}^* + \tau_o$, with $\widetilde{\xi}^* \sim \mathrm{Exp}(1)$ and $\tau_o \sim \mathrm{Exp}(2)$ independent.

A simple calculation shows that $\widetilde{\xi}^* + \tau_o$ has the density function $2e^{-t}\bigl(1 - e^{-t}\bigr)$, $t > 0$, while $\mathcal{Y}_t \sim \mathrm{Ge}_1(e^{-t})$, see Example A.8. Hence, for any $k \geqslant 1$, with

$x = e^{-t}$,

$$\begin{aligned}\mathbb{P}(X=k) &= 2\int_0^\infty \mathbb{P}(\mathcal{Y}_t = k)e^{-t}\bigl(1-e^{-t}\bigr)\,\mathrm{d}t \\ &= 2\int_0^\infty e^{-t}(1-e^{-t})^{k-1}e^{-t}\bigl(1-e^{-t}\bigr)\,\mathrm{d}t = 2\int_0^1 x(1-x)^k\,\mathrm{d}x \\ &= \frac{2}{(k+1)(k+2)}.\end{aligned} \tag{12.6}$$

We have here considered only the number of nodes in $\widehat{T}^1$. However, it is furthermore clear from the symmetry of the Yule tree process that given $X = |\widehat{T}^1|$, the random tree $\widehat{T}^1$ is distributed as a random recursive tree of order X, i.e., $\widehat{T}^1$ is a random recursive tree with random order X given by (12.6). This describes, at least in principle, any properties of $\widehat{T}^1$. For example, we may as in [43] count leaves and non-leaves separately in $\widehat{T}^1$. It is easy to see by induction that if $k \geqslant 2$, then

$$\mathbb{P}(T_k \text{ has } i \text{ leaves}) = \mathbb{P}(T_k \text{ has } i \text{ non-leaves}) = \frac{\left\langle {k-1 \atop i-1} \right\rangle}{(k-1)!}, \tag{12.7}$$

where $\left\langle {k-1 \atop i-1} \right\rangle$ denotes the Eulerian number (see e.g. [53] or [97]); i.e., the number of leaves in T_k is distributed as 1 + the number of ascents in a random permutation of length $k-1$. (In fact, both random vectors (#leaves, #non-leaves) and (1 + #ascents, 1 + #descents) evolve when k is increased as generalized Pólya urns with balls of two colours where we draw a ball and return it together with a ball of the opposite colour.) Consequently we find, for $k \geqslant 2$ and $1 \leqslant i \leqslant k$,

$$\begin{aligned}&\mathbb{P}\bigl(\widehat{T}^1_n \text{ has } i \text{ non-leaves and } k-i \text{ leaves}\bigr) \\ &\qquad \xrightarrow{\text{a.s.}} \mathbb{P}\bigl(\widehat{T}^1 \text{ has } i \text{ non-leaves and } k-i \text{ leaves}\bigr) \\ &\qquad = \frac{2}{(k+1)(k+2)}\,\frac{\left\langle {k-1 \atop i-1} \right\rangle}{(k-1)!} = \frac{2k\left\langle {k-1 \atop i-1} \right\rangle}{(k+2)!}.\end{aligned} \tag{12.8}$$

Let $p_i := \mathbb{P}\bigl(\widehat{T}^1 \text{ has } i \text{ non-leaves}\bigr)$. Summing (12.8) over k we find for example, after short calculations (partly assisted by `Maple`), $p_1 = 6 - 2e \approx 0.563$, $p_2 = 11 - 4e \approx 0.127$, $p_3 = \frac{857}{54} - 5e - \frac{1}{2}e^2 + \frac{2}{27}e^3 \approx 0.072$.

Using [97, (26.14.6)], it is easy to see that p_i is a polynomial in e with rational coefficients, of degree at most i, but we do not know any simple general formula for p_i.

12.1. Sampling a random key

Similarly, in an m-ary search tree, one might sample a *key* uniformly at random and consider the properties of the node containing that key.

Theorem 12.5. *Let T_n be a random m-ary search tree, and let $\mathcal{P}$ be a property as in Theorem 5.14. Sample a random key uniformly, and let v be the node containing that key. Then, as $n \to \infty$, letting $R(T)$ denote the number of keys in the root of T,*

$$\mathbb{P}\bigl(v \text{ has } \mathcal{P}\bigr) \xrightarrow{\text{a.s.}} \frac{\mathbb{E}\bigl(R(\overline{\mathcal{T}})\mathbf{1}\{\overline{\mathcal{T}} \in \mathcal{P}\}\bigr)}{\mathbb{E}\, R(\overline{\mathcal{T}})} = \frac{\mathbb{E}\bigl(R(\overline{\mathcal{T}})\mathbf{1}\{\overline{\mathcal{T}} \in \mathcal{P}\}\bigr)}{2(H_m - 1)}. \tag{12.9}$$

Proof. Let Q_k be the property of a node v that it contains k keys. Then, by Theorem 5.14,

$$\begin{aligned}\mathbb{P}\bigl(v \text{ has } \mathcal{P}\bigr) = \frac{\sum_k k n_{\mathcal{P}\wedge\mathcal{Q}_k}(T_n)}{\sum_k k n_{\mathcal{Q}_k}(T_n)} &\xrightarrow{\text{a.s.}} \frac{\sum_k k\,\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{P} \wedge \mathcal{Q}_k)}{\sum_k k\,\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{Q}_k)} \\ &= \frac{\mathbb{E}\bigl(\sum_k k\mathbf{1}\{\overline{\mathcal{T}} \in \mathcal{P} \wedge \mathcal{Q}_k\}\bigr)}{\mathbb{E}\bigl(\sum_k k\mathbf{1}\{\overline{\mathcal{T}} \in \mathcal{Q}_k\}\bigr)},\end{aligned} \tag{12.10}$$

which equals the second term in (12.9) because $\overline{\mathcal{T}} \in \mathcal{Q}_k \iff R(\overline{\mathcal{T}}) = k$. Furthermore, for the same reason,

$$\frac{\sum_k k n_{\mathcal{Q}_k}(T_n)}{|T_n|} \xrightarrow{\text{a.s.}} \sum_k k\,\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{Q}_k) = \mathbb{E}\, R(\overline{\mathcal{T}}). \tag{12.11}$$

Since $\sum_k k n_{\mathcal{Q}_k}(T_n) = n$, the total number of keys, (12.11) and Theorem 7.11 imply

$$\mathbb{E}\, R(\overline{\mathcal{T}}) = 2(H_m - 1) \tag{12.12}$$

which completes the proof. (Alternatively, (12.12) follows from Theorem 7.13, noting that the limits in (7.23) are the probabilities $\mathbb{P}(R = k)$.) □

Remark 12.6. Theorem 12.5 extends to properties as in Theorem 5.25 (or Theorem 5.26) with only notational changes: replace $R(\overline{\mathcal{T}})$ by the number of keys in the distinguished node o and $\mathbf{1}\{\overline{\mathcal{T}} \in \mathcal{P}\}$ by $\mathbf{1}\{o \text{ has } \mathcal{P}\}$.

Example 12.7. Let K' be the number of keys in the node containing a random key in an m-ary search tree T_n. Theorems 12.5 and 7.13 imply that

$$\mathbb{P}(K' = k) \xrightarrow{\text{a.s.}} \begin{cases} \frac{1}{H_m - 1}\frac{k}{(k+1)(k+2)}, & 1 \leqslant k \leqslant m-2, \\ \frac{m-1}{m(H_m-1)}, & k = m-1. \end{cases} \tag{12.13}$$

For $m = 3, 4, 5$, this yields the limit distributions $\bigl(\frac{1}{5}, \frac{4}{5}\bigr)$, $\bigl(\frac{2}{13}, \frac{2}{13}, \frac{9}{13}\bigr)$, and $\bigl(\frac{10}{77}, \frac{10}{77}, \frac{9}{77}, \frac{48}{77}\bigr)$.

Example 12.8. Let D' be the number of children (the outdegree) of the node containing a random key in an m-ary search tree T_n. It follows from Theorems 12.5 and 7.14, noting that a node with outdegree different from 0 always contains $m-1$ keys,

$$\mathbb{P}(D' = k) \xrightarrow{\text{a.s.}} \begin{cases} 1 - \frac{(m-1)}{(m+1)(H_m-1)}, & k = 0, \\ \frac{(m-1)}{m(m+1)(H_m-1)}, & 1 \leqslant k \leqslant m. \end{cases} \tag{12.14}$$

For $m = 3, 4, 5$, this yields the limit distributions $\left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$, $\left(\frac{29}{65}, \frac{9}{65}, \frac{9}{65}, \frac{9}{65}, \frac{9}{65}\right)$ and $\left(\frac{37}{77}, \frac{8}{77}, \frac{8}{77}, \frac{8}{77}, \frac{8}{77}, \frac{8}{77}\right)$.

13. Height

We have in this paper so far considered fringe properties of random trees. However, the connection with Crump–Mode–Jagers branching processes has also been used very fruitfully to study properties related to the distance to the root, in particular the height of the tree. This was pioneered by Devroye [29] using results by Kingman [77] and Biggins [11, 12] for branching random walks with discrete time (based on Galton–Watson processes), see also Devroye [30], Mahmoud [85], the survey Devroye [34], and Broutin and Devroye [20]. (Partial results for the binary search tree had been proved earlier by Pittel [101], using the same continuous-time branching process as [29] in a somewhat different way.) The method was further developed by Biggins [13, 14] using the continuous-time Crump–Mode–Jagers branching processes used in the present paper. We give in this section a description of the method and apply it to the height; further applications to the saturation level, profile, typical depth and total path length are given in a preliminary version of the present paper [61]. See also the papers just mentioned for further details and results; in particular, note the second order results in [30, 34].

Recall that the *depth* $h(v)$ of a node v is its distance from the root. The *height* $H(T)$ of a tree T is defined as $\max_{v\in T} h(v)$, the maximum depth of a node.

The key idea that makes it possible to apply results on branching random walks is to plot the individuals in a branching process in the plane, using two coordinates that we call *time* and *position*; time is the usual time of birth in the branching process and position is an additional variable. We assume that for each individual is defined, besides the sequence $(\xi_i)_{i=1}^N$ of birth times of the children (relative to the birth of the parent), also a sequence $(\eta_i)_{i=1}^N$ (of the same length N) of random displacements, with $-\infty < \eta_i < \infty$; if the parent is born at time and position (σ, y), then child i is born at time and position $(\sigma + \xi_i, y + \eta_i)$. (The general results in [13], [14] allow also a further random component, describing a random motion of each individual during its life. For our purposes, we put that motion equal to 0 and let each individual be static.)

Results for branching random walks have been applied to the height (and other properties) of random trees in two different ways. In the original application of Devroye [29], see also [30, 34], the "position" is what we have called time in the Crump–Mode–Jagers branching process, while "time" is the number of the generation, i.e., the depth in the family tree $\mathcal{T}_t$. This means that "time" is discrete and that we consider a Galton–Watson process where each individual has a position that is its time of birth in the Crump–Mode–Jagers process studied elsewhere in the present paper. (Furthermore, in this application, the Galton–Watson process is deterministic; in the original application to binary search trees, we consider an infinite binary tree.) Note that $H(\mathcal{T}_t) \geqslant n$ if and only if the minimum position of an individual in generation n is $\leqslant t$,

which gives the required connection with the theorems on branching random walks.

The alternative approach, described by Biggins [14], reverses the two coordinates and lets "time" be time in the Crump–Mode–Jagers branching process while "position" is the generation number, i.e., the depth in the family tree. The offsets η_i are thus non-random with $\eta_i = 1$. (We sometimes reverse signs and take $\eta = -1$.) We use this approach in the present section, referring to [14] for further details on branching random walks and to [13] for proofs of the theorems used here.

13.1. Almost sure limits

Let B_t be the maximum of the positions y_x of all individuals x that are born before or on time t, i.e., with birth time $\sigma_x \leqslant t$. In our case with $\eta_i \equiv 1$, y_x is the generation number of x, and thus $B_t = H(\mathcal{T}_t)$, the height of $\mathcal{T}_t$.

Define the two-dimensional Laplace transform, for $\zeta, \theta \in (-\infty, \infty)$,

$$m(\zeta, \theta) := \mathbb{E} \sum_{i=1}^{N} e^{-\zeta \eta_i - \theta \xi_i} \in (0, \infty]. \tag{13.1}$$

Note that by (5.7), $m(0, \theta) = \widehat{\mu}(\theta)$. In our case with $\eta_i \equiv 1$, we thus simply have

$$m(\zeta, \theta) = e^{-\zeta} \widehat{\mu}(\theta). \tag{13.2}$$

Furthermore, let

$$\gamma = \inf\Bigl\{ a : \inf_{\zeta < 0} \log m(\zeta, -a\zeta) < 0 \Bigr\}. \tag{13.3}$$

One of the main results of Biggins [13, 14] is the following (valid for general η_i under some conditions that are satisfied in our case $\eta_i \equiv 1$):

Theorem 13.1 (Biggins [13, 14])**.** *As $t \to \infty$,*

$$B_t/t \xrightarrow{\text{a.s.}} \gamma. \tag{13.4}$$

□

In our case $H(\mathcal{T}_t) = B_t$, so this yields the asymptotic height of $\mathcal{T}_t$; this translates to the height of $T_n = \mathcal{T}_{\tau(n)}$ as follows.

Theorem 13.2. *Under the assumptions (A1)–(A5) and (A6ψ), as $n \to \infty$,*

$$\frac{H(T_n)}{\log n} \xrightarrow{\text{a.s.}} \bar{\gamma} := \frac{\gamma}{\alpha}. \tag{13.5}$$

Proof. By Theorem 13.1, $H(T_n)/\tau(n) \xrightarrow{\text{a.s.}} \gamma$, and the result follows by (5.17). □

Remark 13.3. The fragmentation trees in Section 9 are of a slightly different type than the trees T_n that are our main object of study, since they appear as the family tree $\mathcal{T}_t$ stopped at a fixed time $t = \log(x_0/x_1)$ instead of a random time $\tau(n)$, see Remark 9.2. This means that asymptotics for the height of fragmentation trees follow directly from Theorem 13.1 rather than from Theorem 13.2. In this section we usually consider only trees of the type T_n, and leave corresponding results for fragmentation trees to the reader.

Remark 13.4. Also the split trees defined by Devroye [35], see Remark 9.7, are in general not exactly of the type of trees studied here, but for the purpose of studying the height, they can be approximated by fragmentation trees and similar results can be obtained, see Broutin and Devroye [20] and Broutin, Devroye and McLeish [21].

By (13.5), $\gamma \geqslant 0$, and thus the fundamental constant γ in (13.3) and (13.5) can also be evaluated as

$$\begin{aligned}\gamma &= \inf\bigl\{a>0 : \inf_{\zeta<0} \log m(\zeta,-a\zeta) < 0\bigr\}\\ &= \inf\bigl\{a>0 : \inf_{\theta>0} \log m(-\theta/a,\theta) < 0\bigr\}.\end{aligned} \tag{13.6}$$

In our case, when (13.2) holds, this simplifies to

$$\begin{aligned}\gamma &= \inf\bigl\{a>0 : \inf_{\theta>0}\bigl\{\theta/a + \log\widehat{\mu}(\theta)\bigr\} < 0\bigr\}\\ &= \inf\Bigl\{a>0 : \inf_{\theta>0}\Bigl\{\frac{1}{a} + \frac{\log\widehat{\mu}(\theta)}{\theta}\Bigr\} < 0\Bigr\}\\ &= \inf\Bigl\{a>0 : \frac{1}{a} < -\inf_{\theta>0}\frac{\log\widehat{\mu}(\theta)}{\theta}\Bigr\}\end{aligned} \tag{13.7}$$

and thus

$$\gamma^{-1} = -\inf_{\theta>0}\frac{\log\widehat{\mu}(\theta)}{\theta} = -\inf_{\theta\geqslant\alpha}\frac{\log\widehat{\mu}(\theta)}{\theta}. \tag{13.8}$$

Geometrically, (13.8) says that $-\gamma^{-1}$ is the slope of the tangent from the origin to the curve $\log\widehat{\mu}(\theta)$, $\theta>0$, provided such a tangent exists. (Otherwise, $-\gamma^{-1}$ is the slope of the asymptote, as follows from Lemma 13.5(ii) and Remark 13.6 below.) Analytically, γ can be found as follows.

Lemma 13.5. $0 < \beta^{-1} \leqslant \gamma < \infty$.

(i) *If $\theta>0$ is a solution of the equation*

$$\theta\frac{\widehat{\mu}'(\theta)}{\widehat{\mu}(\theta)} = \log\widehat{\mu}(\theta), \tag{13.9}$$

then

$$\gamma^{-1} = -\frac{\log\widehat{\mu}(\theta)}{\theta} = -\frac{\widehat{\mu}'(\theta)}{\widehat{\mu}(\theta)} = -\bigl(\log\widehat{\mu}\bigr)'(\theta). \tag{13.10}$$

Furthermore, (13.9) *has at most one positive solution.*

(ii) *If* (13.9) *has no positive solution, then*

$$\gamma^{-1} = -\lim_{\theta\to\infty} \frac{\log \widehat{\mu}(\theta)}{\theta} = \sup\{x : \mu[0,x) = 0\}. \tag{13.11}$$

Proof. Since $\log \widehat{\mu}(\theta) < 0$ for $\theta > \alpha$, (13.8) yields $\gamma^{-1} > 0$ and thus $\gamma < \infty$.

It is well-known, and follows from Hölder's inequality, that $\log \widehat{\mu}(\theta)$ is convex; furthermore, using (5.39),

$$\bigl(\log \widehat{\mu}\bigr)'(\alpha) = \frac{\widehat{\mu}'(\alpha)}{\widehat{\mu}(\alpha)} = \widehat{\mu}'(\alpha) = -\beta. \tag{13.12}$$

Hence, for any real θ,

$$\log \widehat{\mu}(\theta) \geqslant \log \widehat{\mu}(\alpha) + (\theta-\alpha)(\log \widehat{\mu})'(\alpha) = -\beta(\theta-\alpha) > -\beta\theta. \tag{13.13}$$

Hence, for $\theta > 0$, $\log \widehat{\mu}(\theta)/\theta > -\beta$, and (13.8) yields $\gamma^{-1} \leqslant \beta$. Thus $\gamma \geqslant \beta^{-1}$.

Next, for any differentiable convex function $f(\theta)$ defined on an interval $\mathcal{I} \subseteq (0,\infty)$, the function $g(\theta) := \theta f'(\theta) - f(\theta)$ is increasing since $g'(\theta) = \theta f''(\theta) \geqslant 0$. Hence, $g(\theta)$ has at most one zero in $\mathcal{I}$, and since $\bigl(f(\theta)/\theta\bigr)' = g(\theta)/\theta^2$, a zero of $g(\theta)$ is a global minimum point of $f(\theta)/\theta$ in $\mathcal{I}$. Taking $f(\theta) := \log \widehat{\mu}(\theta)$ on $\mathcal{I} := \{\theta > 0 : \widehat{\mu}(\theta) < \infty\}$, we see that $g(\theta) = 0$ is equivalent to (13.9). If (13.9) has a positive solution, it is thus unique and a minimum point of $\log \widehat{\mu}(\theta)/\theta$ which yields (13.10) by (13.8) and (13.9).

On the other hand, if (13.9) has no positive solution, then $g(\theta)$ has a fixed sign in $\mathcal{I}$. Since $g(\alpha) = \alpha(\log \widehat{\mu})'(\alpha) < 0$, $g(\theta) < 0$ for all $\theta \in \mathcal{I}$ and $f(\theta)/\theta$ is strictly decreasing. Thus, the infimum in (13.8) is the limit as $\theta \to \infty$, which yields the first equality in (13.11). The final equality is a straightforward property of Laplace transforms. □

Remark 13.6. The case (ii) in Lemma 13.5 is exceptional. We see from (13.11) that μ has no mass in $[0,\gamma^{-1})$, so no child is ever born to a parent of age less than γ^{-1}. Moreover, by (13.8), $\widehat{\mu}(\theta) \geqslant e^{-\gamma^{-1}\theta}$ for all $\theta > 0$, and it follows easily that $\mu\{\gamma^{-1}\} \geqslant 1$, so μ has a point mass at γ^{-1}. This case is thus exceptional, and does not appear in any of our examples.

Example 13.7. For the random recursive tree in Example 6.1, originally treated by Devroye [30] in a related but somewhat different way, see also [34], we have $\widehat{\mu}(\theta) = 1/\theta$, $\theta > 0$, see (6.1). Consequently, (13.9) is

$$-1 = -\log\theta \tag{13.14}$$

with the solution $\theta = e$, and then (13.10) yields

$$\gamma^{-1} = -\bigl(\log \widehat{\mu}\bigr)'(\theta) = \frac{1}{\theta} = e^{-1} \tag{13.15}$$

i.e., $\gamma = \theta = e$. Since $\alpha = 1$, the limit in (13.5) is $\bar{\gamma} = \gamma$.

Example 13.8. For the binary search tree in Example 6.2, originally treated by Devroye [29], see also [30], [14] and [34], we have $\widehat{\mu}(\theta) = 2/(1+\theta)$, $\theta > -1$, see (6.5). Consequently, (13.9) is

$$-\frac{\theta}{1+\theta} = \log\frac{2}{1+\theta} \tag{13.16}$$

and (13.10) is

$$\gamma^{-1} = -\bigl(\log\widehat{\mu}\bigr)'(\theta) = \frac{1}{1+\theta}, \tag{13.17}$$

i.e., $\gamma = 1+\theta$. Since $\alpha = 1$, $\bar{\gamma} = \gamma$. We may substitute (13.17) in (13.16), noting that $\theta > 0$ corresponds to $\gamma > 1$, and obtain

$$\gamma^{-1} - 1 = \log\bigl(2\gamma^{-1}\bigr) = -\log(\gamma/2) \tag{13.18}$$

or

$$\gamma\log(\gamma/2) + 1 - \gamma = 0, \tag{13.19}$$

which has the root $\gamma \doteq 4.311070$. (The theory above implies that (13.19) has a unique root $\gamma > 1$.)

Example 13.9. More generally, consider a linear preferential attachment tree as in Example 6.4, with weights $w_k = \chi k + \rho$ as in (6.17). (This was originally treated by Pittel [102], at least in the case $\chi \geqslant 0$; see also Biggins and Grey [15].) Since only the quotient χ/ρ matters, and $\rho > 0$, we may and shall for simplicity assume $\rho = 1$. (It follows that (13.24) below holds also in the general case provided we replace χ by χ/ρ.)

We then have $\widehat{\mu}(\theta) = 1/(\theta - \chi)$ by (6.19), and thus (13.9) is, with $\theta > \chi$,

$$-\frac{\theta}{\theta-\chi} = -\log(\theta-\chi) \tag{13.20}$$

and (13.10) is

$$\gamma^{-1} = \frac{1}{\theta-\chi}, \tag{13.21}$$

i.e., $\gamma = \theta - \chi$. We may substitute this in (13.20) and obtain

$$\log\gamma = \frac{\theta}{\theta-\chi} = \frac{\gamma+\chi}{\gamma} \tag{13.22}$$

or

$$\gamma\log\gamma - \gamma = \chi. \tag{13.23}$$

The limit in (13.5) is $\bar{\gamma} := \gamma/\alpha$; we have $\alpha = \chi + 1$ by (6.20), and thus (13.23) yields

$$\bar{\gamma}\log\bigl((\chi+1)\bar{\gamma}\bigr) - \bar{\gamma} = \frac{\chi}{\chi+1}. \tag{13.24}$$

Example 13.7 is the case $\chi = 0$. For the binary search tree in Example 13.8 we have $\chi = -1$ and $\rho = 2$, which is equivalent to $\chi = -\frac{1}{2}$ and $\rho = 1$; thus we

take $\chi = -\frac{1}{2}$ above. Indeed, (13.24) with $\chi = -\frac{1}{2}$ is the same as (13.19). More generally, the m-ary increasing tree in Example 6.7 is obtained by $\chi = -1/m$. For this m-ary tree, (13.24) yields, cf. the binary case (13.19),

$$\bar\gamma \log\Bigl(\frac{m-1}{m}\bar\gamma\Bigr) - \bar\gamma + \frac{1}{m-1} = 0. \tag{13.25}$$

The height of the m-ary increasing tree was found by Broutin, Devroye, McLeish and de la Salle [22], see also Drmota [40, Theorem 6.47].

Example 13.10. A somewhat more complicated example is the m-ary search tree in Section 7.1 or 7.2. (For this example, it does not matter whether we include external nodes or not, since this only changes the height by 1. Furthermore, μ is the same for both versions, so the calculations are the same.) This was originally treated by Devroye [31], see also Mahmoud [85], Pittel [102], Biggins [14] and Devroye [34].

Recall that $\alpha = 1$, so $\bar\gamma = \gamma$. By (7.2),

$$\log \widehat\mu(\theta) = \log m! - \sum_{i=1}^{m-1} \log(i+\theta). \tag{13.26}$$

Hence (13.9) is

$$\theta \sum_{i=1}^{m-1} \frac{1}{\theta+i} - \sum_{i=1}^{m-1} \log(\theta+i) + \log m! = 0 \tag{13.27}$$

and (13.10) is

$$\gamma^{-1} = \sum_{i=1}^{m-1} \frac{1}{\theta+i}, \tag{13.28}$$

which yields γ after (numerically) finding the unique positive root of (13.27).

Example 13.11. We can extend Example 13.10 for m-ary search trees to the generalization in Example 8.4. (The case $m = 2$, the median-of-$(2\ell+1)$ binary search tree, was treated by Devroye [33]; the general case was studied, by different methods, by Chauvin and Drmota [24], see also Drmota [40, Section 6.5.2].) We still have $\alpha = 1$, so $\bar\gamma = \gamma$. It follows from (8.7) that (13.9) is

$$\sum_{i=1}^{(m-1)(\ell+1)} \frac{\theta}{\theta+i+\ell} - \sum_{i=1}^{(m-1)(\ell+1)} \log(\theta+i+\ell) + \log\frac{(m\ell+m)!}{(\ell+1)!} = 0 \tag{13.29}$$

and (13.10) is

$$\gamma^{-1} = \sum_{i=1}^{(m-1)(\ell+1)} \frac{1}{\theta+i+\ell}. \tag{13.30}$$

Example 13.12. Another example is the binary pyramids in Example 6.8, where the height was found by Mahmoud [86] by similar methods, see also Biggins and Grey [15] and Broutin, Devroye and McLeish [21]. The equations, now using (6.44), become a bit involved; we refer to [86], [15] and [21] for a numerical solution.

Example 13.13. For the fragmentation tree in Example 9.4, we have a branching process that differs from the one for the binary search tree in Example 13.8, but the intensity μ is the same, so all calculations in Example 13.8 are valid for this tree too. Thus, see Theorem 13.1 and Remark 13.3, $H(\mathcal{T}_t)/t \xrightarrow{\text{a.s.}} \gamma$ with γ given by (13.19). Furthermore, if we stop at n nodes as in Remark 9.2, $H(T_n)/\log n \xrightarrow{\text{a.s.}} \gamma$, just as for the binary search tree. More precise results for the height of this fragmentation tree, and m-ary generalizations of it, are given by Chauvin and Drmota [24].

13.2. Moment convergence

We can also obtain moment convergence in Theorem 13.2, in particular convergence of the expectation $\mathbb{E}\,H(T_n)/\log n$ to $\bar{\gamma}$, at least if we assume the following additional condition on the birth times for an individual in the Crump–Mode–Jagers process.

(A*) There exists $\delta > 0$ such that $\mathbb{E}\,e^{\delta\xi_2} < \infty$.

In other words, each individual gets at least two children ($N \geqslant 2$), and the age when the second child is born has an exponential moment. (Equivalently, it has exponentially decreasing tails.)

The condition (A*) is satisfied in all examples in Sections 6–8, since ξ_2 is the sum of one or several exponential waiting times.

Remark 13.14. We use (A*) in the proof of Lemma 13.16 below. Some extra condition is clearly needed for Lemma 13.16 (at least $\mathbb{E}\,e^{\delta\xi_1} < \infty$ for some $\delta > 0$, since $\tau(n) \geqslant \xi_1$ if $n > 1$ and, say, $\psi(t) = 1$.) However, we do not know whether (A*) really is needed for Lemmas 13.17 and 13.18 and for Theorem 13.19. In fact, we conjecture that Theorem 13.19 holds assuming only (A1)–(A5) and (A6ψ).

We begin with some lemmas. The first two are stated somewhat more generally than actually needed here.

Lemma 13.15. *For every $r > 0$, there exists c_r such that, for large t,*

$$\mathbb{P}\bigl(H(\mathcal{T}_t) \geqslant c_r t\bigr) \leqslant e^{-rt}. \tag{13.31}$$

Proof. Let $N_t[x,\infty)$ be the number of nodes in $\mathcal{T}_t$ that have depth $\geqslant x$, and let $\eta_t[x,\infty) := \mathbb{E}\bigl(N_t[x,\infty)\bigr)$. We also introduce, as in [13, 14], the function

$$\alpha(\zeta) := \inf\{\theta : m(\zeta,\theta) \leqslant 1\} = \inf\bigl\{\theta : \widehat{\mu}(\theta) \leqslant e^{\zeta}\bigr\} = \inf\bigl\{\theta : \log\widehat{\mu}(\theta) \leqslant \zeta\bigr\}, \tag{13.32}$$

and its one-sided Legendre transform

$$\alpha^*(x) := \inf_{\zeta<0}\bigl\{x\zeta + \alpha(\zeta)\bigr\}. \tag{13.33}$$

Then, by [13, Theorem 4 and its proof], for any real x,

$$\limsup_{t\to\infty} \frac{\log \eta_t[tx,\infty)}{t} \leqslant \alpha^*(x). \tag{13.34}$$

In fact, [13, Theorem 4] shows that the limit exists and equals $\alpha^*(x)$ except possibly for one exceptional x. The upper bound (13.34) (for every x) is the simple part of the proof and follows from [13, Corollary 1].

Using (13.32), we can rewrite (13.33) as, for $x \geqslant 0$,

$$\begin{aligned}\alpha^*(x) &= \inf_{\substack{\zeta<0\\ \log\widehat{\mu}(\theta)\leqslant\zeta}} \bigl\{x\zeta+\theta\bigr\} = \inf_{\widehat{\mu}(\theta)<1}\bigl\{x\log\widehat{\mu}(\theta)+\theta\bigr\}\\ &= \inf_{\theta>\alpha}\bigl\{x\log\widehat{\mu}(\theta)+\theta\bigr\}.\end{aligned} \tag{13.35}$$

Since $\log\widehat{\mu}(\theta) < 0$ for any $\theta > \alpha$, (13.35) implies

$$\lim_{x\to\infty} \alpha^*(x) = -\infty. \tag{13.36}$$

Consequently, we can choose c_r such that $\alpha^*(c_r) < -r$, and then (13.34) yields $\log\eta_t[c_rt,\infty) \leqslant -rt$ for large t. This yields the result (13.31), since

$$\mathbb{P}\bigl(H(\mathcal{T}_t) \geqslant c_rt\bigr) = \mathbb{P}\bigl(N_t[c_rt,\infty) \geqslant 1\bigr) \leqslant \eta_t[c_rt,\infty) \leqslant e^{-rt}. \tag{13.37}$$

□

Lemma 13.16. *Assume (A1)–(A5) and (A*). Then, for every $r \geqslant 0$ there exists c_r' such that $\mathbb{P}\bigl(\tau(n) \geqslant c_r'\log n\bigr) = O\bigl(n^{-r}\bigr)$.*

Proof. First, if (A6ψ) holds, then $\tau(n)/\log n \xrightarrow{\text{a.s.}} \alpha^{-1}$ by Theorem 5.12. In general, we may as in the proof of Theorem 5.12 consider the truncation $\psi_1(t) := \psi(t)\wedge 1$ and the corresponding stopping time $\tau_1(n)$. Then $\tau_1(n)/\log n \xrightarrow{\text{a.s.}} \alpha^{-1}$ and $\tau(n) \leqslant \tau_1(n)$. Hence, in any case, if $b := 2\alpha^{-1}$, then $\mathbb{P}(\tau(n) < b\log n) \to 1$; in particular, if n is large enough (which we assume in the rest of the proof; the result is trivial for small n),

$$\mathbb{P}(\tau(n) \geqslant b\log n) < e^{-1}. \tag{13.38}$$

Let $L := \lceil r\log n\rceil$. Let $v_0 = o$, the root, and let v_i be the second child of v_{i-1}, $i \geqslant 1$. (Thus, denoting the individuals by strings in $V_\infty := \bigcup_{n=0}^\infty \mathbb{N}^n$, see Section 5, the nodes v_i are $\emptyset, 2, 22, 222, \dots$.) Let $\tau_L := \sigma_{v_L}$, the time v_L is born. Let $A := \mathbb{E}\, e^{\delta\xi_2}$, which is finite by (A*), and let $B := r(\log A+1)/\delta$. Then,

$$\begin{aligned}\mathbb{P}(\tau_L \geqslant B\log n) &\leqslant e^{-\delta B\log n}\,\mathbb{E}\, e^{\delta\tau_L} = e^{-\delta B\log n}\bigl(\mathbb{E}\, e^{\delta\xi_2}\bigr)^L = e^{-\delta B\log n}A^L\\ &\leqslant e^{-\delta B\log n+(1+r\log n)\log A} = An^{-\delta B+r\log A} = An^{-r}.\end{aligned} \tag{13.39}$$

Let w_i be the first child of v_{i-1}, $i \geqslant 1$. Then w_i is born before (or at the same time as) v_i, so $w_1, \dots, w_L$ are all born before (or at) τ_L. Each w_i starts an independent copy $\mathcal{T}_t^{(i)}$ of the branching process and its family tree (we do not distinguish between these two), with local time $t = 0$ when w_i is born, i.e., at global time σ_{w_i}. Let $Z_t^{\psi,(i)}$ be the total weight of $\mathcal{T}_t^{(i)}$ and let $\tau^{(i)}(n)$ be the corresponding stopping time $\tau^{(i)}(n) := \inf\{t : Z_t^{\psi,(i)} \geqslant n\}$.

For every $t \geqslant 0$, the individuals (nodes) in $\mathcal{T}_t^{(i)}$ form a subset of $\mathcal{T}_{\sigma_{w_i}+t}$, and they have the same ages in both trees; hence the definition (5.1) implies that $Z^{\psi}_{\sigma_{w_i}+t} \geqslant Z_t^{\psi,(i)}$. Consequently, $Z^{\psi}_{\sigma_{w_i}+\tau^{(i)}(n)} \geqslant n$ and thus, for every $i \leqslant L$,

$$\tau(n) \leqslant \sigma_{w_i} + \tau^{(i)}(n) \leqslant \tau_L + \tau^{(i)}(n). \tag{13.40}$$

In other words,

$$\tau(n) \leqslant \tau_L + \min_{i \leqslant L} \tau^{(i)}(n). \tag{13.41}$$

Moreover, the stopping times $\tau^{(i)}(n)$ are independent, and have the same distribution as $\tau(n)$. Hence, (13.38) implies that

$$\mathbb{P}\bigl(\min_{i \leqslant L} \tau^{(i)}(n) \geqslant b \log n\bigr) = \mathbb{P}\bigl(\tau(n) \geqslant b \log n\bigr)^L < e^{-L} \leqslant n^{-r}. \tag{13.42}$$

Combining (13.41), (13.39) and (13.42), we obtain

$$\mathbb{P}\bigl(\tau(n) \geqslant (b+B)\log n\bigr) \leqslant (A+1)n^{-r}. \tag{13.43}$$

□

The next lemma will immediately be improved in Lemma 13.18. Lemma 13.18 is trivially true for $n = 1$ too; however, we assume $n \geqslant 2$ since as said in Section 5, in principle we do not require n to be an integer; any real positive n is possible. (We use this in the proof of Lemma 13.18 below, for convenience, when we do not round m to an integer.)

Lemma 13.17. *Assume (A1)–(A5) and (A*). Then there exists c such that, for all n*

$$\mathbb{P}\bigl(H(T_n) \geqslant c \log n\bigr) = O\bigl(n^{-1}\bigr). \tag{13.44}$$

Proof. Let c_1 and c_1' be as in Lemmas 13.15 and 13.16 with $r = 1$; we may assume $c_1' \geqslant 1$. Then by Lemmas 13.16 and 13.15, recalling that $T_n := \mathcal{T}_{\tau(n)}$, for large n,

$$\begin{aligned}\mathbb{P}\bigl(H(T_n) \geqslant c_1 c_1' \log n\bigr) &\leqslant \mathbb{P}\bigl(\tau(n) > c_1' \log n\bigr) + \mathbb{P}\bigl(H(\mathcal{T}_{c_1' \log n}) \geqslant c_1 c_1' \log n\bigr) \\ &\leqslant O\bigl(n^{-1}\bigr) + e^{-c_1' \log n} = O\bigl(n^{-1}\bigr).\end{aligned} \tag{13.45}$$

This completes the proof, since (13.44) is trivial for small n. □

Lemma 13.18. *Assume (A1)–(A5) and (A*). Then there exists C such that, for all $n \geqslant 2$ and $t \geqslant 1$,*

$$\mathbb{P}\bigl(H(T_n) \geqslant Ct \log n\bigr) \leqslant n^{-t}. \tag{13.46}$$

Proof. By Lemma 13.17, there exists c and A such that, for all n,

$$\mathbb{P}\bigl(H(T_n) \geqslant c\log n\bigr) \leqslant A/n. \tag{13.47}$$

We may assume $A \geqslant 1$.

Given $n \geqslant 2$ and $t \geqslant 1$, let $m := An^t$. Then $m \geqslant n$ and thus $\tau(m) \geqslant \tau(n)$ and $H(T_m) \geqslant H(T_n)$. Consequently, (13.47) applied to m yields

$$\mathbb{P}\bigl(H(T_n) \geqslant c\log m\bigr) \leqslant \mathbb{P}\bigl(H(T_m) \geqslant c\log m\bigr) \leqslant A/m = n^{-t}. \tag{13.48}$$

Moreover, $\log m = \log A + t\log n \leqslant (1 + \log A/\log 2)t\log n$. Hence, with $C := c(1 + \log A/\log 2)$, (13.48) implies

$$\mathbb{P}\bigl(H(T_n) \geqslant Ct\log n\bigr) \leqslant \mathbb{P}\bigl(H(T_n) \geqslant c\log m\bigr) \leqslant n^{-t}. \tag{13.49}$$

□

Theorem 13.19. *Assume (A1)–(A5), (A6ψ) and (A*). Then the convergence in (13.5) holds also in L^r for every $r > 0$, i.e.,*

$$\mathbb{E}\,|H(T_n)/\log n - \bar\gamma|^r \to 0. \tag{13.50}$$

In particular,

$$\frac{\mathbb{E}\bigl(H(T_n)^r\bigr)}{\log^r n} \to \bar\gamma^r, \qquad r > 0. \tag{13.51}$$

Proof. Let $X_n := H(T_n)/\log n$. By Lemma 13.18, for $n \geqslant 2$ and $t \geqslant 1$, $\mathbb{P}(X_n \geqslant Ct) \leqslant n^{-t} \leqslant 2^{-t} \leqslant 2^{1-t}$, which obviously holds also for $t < 1$. Hence, for $r > 0$,

$$\mathbb{E}\,X_n^r = r\int_0^\infty t^{r-1}\,\mathbb{P}(X_n \geqslant t)\,\mathrm{d}t \leqslant r\int_0^\infty t^{r-1}2^{1-t/C}\,\mathrm{d}t < \infty. \tag{13.52}$$

This shows that each moment $\mathbb{E}\,X_n^r$ is uniformly bounded for $n \geqslant 2$. As is well-known, this implies uniform integrability of X_n^r for each r, and thus also of $|X_n - \bar\gamma|^r$; since $X_n \xrightarrow{\text{a.s.}} \bar\gamma$ by Theorem 13.2, this implies (13.50) and (13.51). (See e.g. [54, Theorems 5.4.2 and 5.5.2] for uniform integrability.) □

In particular, Theorem 13.19 shows that $\mathbb{E}\,H(T_n) \sim \bar\gamma\log n$, and similarly for higher moments, in Examples 13.7–13.12. We obtain also corresponding results for the fragmentation tree in Example 13.13, using Lemma 13.15.

Remark 13.20. It follows from (13.50), with $r = 2$, that the variance $\mathbb{E}\,|H(T_n) - \mathbb{E}\,H(T_n)|^2 = o(\log^2 n)$. In the case of a binary search tree, Reed [103] showed the much sharper result that $\mathbb{E}\,|H(T_n) - \mathbb{E}\,H(T_n)|^2 = O(1)$; this was extended to higher central moments and to m-ary search trees by Drmota [41].

14. Asymptotic normality?

As said in the introduction, it is natural to try to show asymptotic normality of the number of fringe trees of a given type. There are several previous results of

this type for special cases. Central limit laws for fringe trees have been shown, by several different methods, for binary search trees in e.g. [32], [36], [49], [50], [23] and [59], and for random recursive trees in e.g. [50] and [59]. For m-ary search trees, the situation is more complicated: no results for general fringe trees have been published (this is work in progress [62]), but some special cases (such as the degree distribution and the number of fringe trees of a given size) and related quantities (the number of internal nodes) have been treated, and it turns out that central limit theorems hold for $m \leqslant 26$ but not for $m \geqslant 27$, see e.g. [87], [85], [83], [26], [63], [25], [48], [67] and [60]. Further examples of asymptotic normality include the degree distribution of plane oriented recursive trees (preferential attachment trees, see Example 6.5) [8], [88], [67], [68], and the number of internal nodes in median-of-$(2\ell+1)$ binary search tree for $\ell \leqslant 58$, but not for $\ell \geqslant 59$, see [26], [27].

The examples of m-ary search tree and median-of-$(2\ell+1)$ binary search trees thus show that central limit theorems do not always hold for fringe trees of the random trees generated by Crump–Mode–Jagers branching processes as in the present paper.

Problem 14.1. Find a characterization of the Crump–Mode–Jagers processes that yield asymptotic normality for the number of fringe trees of a given type.

Using the methods of Section 5, Problem 14.1 can be seen as a special case of the following problem for branching processes:

Problem 14.2. Find a characterization of the Crump–Mode–Jagers branching processes such that for suitable characteristics ϕ and ψ, and with $\tau(n)$ as in Section 5, $Z^{\phi}_{\tau(n)}$ is asymptotically normal as $n \to \infty$.

Problem 14.2 considers a stopped branching process. It is closely related to the following problem for fixed times:

Problem 14.3. Find a characterization of the Crump–Mode–Jagers branching processes such that for suitable characteristics ϕ, $(Z^{\phi}_t - m_\phi Z_t)/\sqrt{Z_t}$ is asymptotically normal as $t \to \infty$.

This problem has been studied, at least for some branching processes. Asmussen and Hering [2, Theorems VIII.3.1 and VIII.12.1] give a central limit theorem of this type for a somewhat different class of branching processes, viz. multi-type Markov branching processes. In principle, as pointed out in [2], this class includes the Crump–Mode–Jagers branching processes studied here (with the "type" taken as the entire previous history of the individual), but the resulting type space is typically so large that the technical conditions in [2] are not satisfied. (In particular, "Condition (M)".) However, for the Crump–Mode–Jagers branching processes used in the examples above, with life histories that are composed of one or several independent waiting times, the process can be described using a finite dimensional type space. It seems that the results in [2] then apply and can be translated to conditions for these Crump–Mode–Jagers branching processes. Presumably, the same conditions then apply to Problems 14.1 and 14.2 too, but that remains an open problem.

Moreover, it seems likely that the same type of conditions apply to much more general Crump–Mode–Jagers branching processes. The conditions in [2] are stated in terms of eigenvalues of a certain operator A defined by the process, and the result says (under some technical assumptions) that if λ_1 is the largest eigenvalue of A (this eigenvalue is real), then we have asymptotic normality if every other eigenvalue λ has $\operatorname{Re}\lambda \leqslant \frac{1}{2}\lambda_1$, but (typically, at least) not otherwise. The same condition also appears in the different but closely related context of generalized Pólya urns, see [67]. We conjecture that this condition (in a suitable form) applies to rather general Crump–Mode–Jagers branching processes. This has been proved in the discrete-time case [72], but the continuous-time case relevant here is more challenging.

Remark 14.4. In contrast, for conditioned Galton–Watson trees (see Remark 1.1), asymptotic normality for fringe trees holds in general, see [70]. (Such trees are not treated in the present paper.)

Appendix A: Birth processes

Recall that a *pure birth process* with birth rates $\lambda_k \geqslant 0$ is a continuous-time stochastic process $(\mathcal{X}_t)_{t\geqslant 0}$, taking values in $\mathbb{Z}_{\geqslant 0}$ and with some given initial value $\mathcal{X}_0 = x_0$ (usually 0 or 1), which is Markov and such that when $\mathcal{X}_t = k$, the process jumps to $k+1$ with rate λ_k; equivalently, the process jumps from k to $k+1$ after an exponential waiting time $Y_k \sim \operatorname{Exp}(\lambda_k)$, and all these waiting times are independent. (We allow $\lambda_k = 0$; we then let $Y_k = \infty$.) Thus, the time the process jumps to k is

$$S_k := \sum_{i=x_0}^{k-1} Y_i, \qquad k \geqslant x_0 + 1, \tag{A.1}$$

and, for $k \geqslant x_0$, with $S_{x_0} := 0$,

$$\mathcal{X}(T) = k \iff S_k \leqslant t < S_{k+1}. \tag{A.2}$$

Example A.1. In the branching process corresponding to a general preferential attachment tree in Example 6.3, the children of an individual are born at the jumps of a pure birth process (with birth rates $\lambda_k = w_k$); the birth process $\mathcal{X}_t$ is thus the number of children at time t.

Example A.2. As a special case (see Example 6.1), the counting process $\Xi[0,t]$ corresponding to a Poisson process Ξ with intensity 1 is a pure birth process with constant intensity $\lambda_k = 1$, started at 0. More generally, a pure birth process with constant birth rate $\lambda_k = \lambda$, started at 0, is a Poisson process with intensity λ. (We have earlier defined a Poisson process as a point process; the corresponding pure birth process considered here is also called a Poisson process. There is an obvious equivalence between the two points of view, and hardly any risk of confusion.)

Example A.3. The *Yule process* in e.g. Example 6.1 is a pure birth process with birth rates $\lambda_k = k$, started at 1. More generally, for a Crump–Mode–Jagers process where each individual gets children according to a Poisson process with intensity $\alpha > 0$, the total size (number of individuals Z_t) is a pure birth process with birth rates $\lambda_k = \alpha k$, started at 1; we call this a *Yule process with rate* $\alpha > 0$. (It evidently differs from the standard case $\alpha = 1$ only by a simple change of time.)

If $(\mathcal{Y}_t)$ is a Yule process with rate α, it thus follows from (5.9) that

$$e^{-\alpha t}\mathcal{Y}_t \to W \tag{A.3}$$

for some random variable W. (Note that the intensity measure μ is $\alpha \mathrm{d}t$, so (5.4) holds and the Malthusian parameter α equals the rate α.) It is easy to verify (5.11) and thus $W > 0$ a.s.; in fact, it follows from (A.23) below that $W \sim \mathrm{Exp}(1)$. (This is one of the few cases with a simple explicit distribution for the limit W.)

We state a general result on stopping a pure birth process by an exponential clock τ.

Theorem A.4. *Let $(\mathcal{X}_t)$ be a pure birth process with birth rates $\lambda_k \geqslant 0$, started at $\mathcal{X}_0 = 0$. Furthermore, let $\tau \sim \mathrm{Exp}(\alpha)$ be independent of the birth process. Then $X := \mathcal{X}_\tau$ has the distribution*

$$\mathbb{P}(X = k) = \frac{\alpha}{\lambda_k + \alpha} \prod_{i=0}^{k-1} \frac{\lambda_i}{\lambda_i + \alpha} = \frac{\alpha \prod_{i=0}^{k-1} \lambda_i}{\prod_{i=0}^{k} (\lambda_i + \alpha)}, \qquad k \geqslant 0. \tag{A.4}$$

We give two different proofs (both simple) to illustrate different ways of arguing with exponential random variables; the first proof is more direct probabilistic and the second more analytic. (The second proof is essentially the same as (6.13)–(6.14) given in Example 6.3; it was there given for a special case but the argument is general, as is shown below.)

First proof of Theorem A.4. Regard τ as an exponential random clock that strikes and stops the process. When $\mathcal{X}_t = k$ and $\tau > t$, so the process has not yet stopped, the next event that happens is either that the clock strikes (rate α), and then $X = k$, or that $\mathcal{X}_t$ jumps to $k+1$ (rate λ_k), and then $X > k$. Consequently,

$$\mathbb{P}(X = k \mid X \geqslant k) = \frac{\alpha}{\alpha + \lambda_k}, \tag{A.5}$$

$$\mathbb{P}(X \geqslant k+1 \mid X \geqslant k) = \frac{\lambda_k}{\alpha + \lambda_k}. \tag{A.6}$$

By (A.6) and induction,

$$\mathbb{P}(X \geqslant k) = \prod_{i=0}^{k-1} \frac{\lambda_i}{\lambda_i + \alpha}, \tag{A.7}$$

and the result follows by (A.5). □

Second proof of Theorem A.4. With the notation above, we have by (A.2)

$$\mathbb{P}(X \geqslant k) = \mathbb{P}(\mathcal{X}_\tau \geqslant k) = \mathbb{P}(S_k \leqslant \tau). \tag{A.8}$$

Conditioning on S_k, we have $\mathbb{P}(S_k \leqslant \tau \mid S_k) = e^{-\alpha S_k}$, and taking the expectation we find, using (A.1) and independence of $Y_1, \dots, Y_{k-1}$,

$$\mathbb{P}(X \geqslant k) = \mathbb{E}\, e^{-\alpha S_k} = \prod_{i=0}^{k-1} \mathbb{E}\, e^{-\alpha Y_i} = \prod_{i=0}^{k-1} \frac{\lambda_i}{\lambda_i + \alpha}. \tag{A.9}$$

The result follows by taking the difference $\mathbb{P}(X \geqslant k) - \mathbb{P}(X \geqslant k+1)$. □

We consider in particular the linear case, when

$$\lambda_k = \chi k + \rho, \tag{A.10}$$

for some constants χ and ρ. (As in Example 6.4, only the ratio χ/ρ matters, up to a change of time scale, so we might assume $\chi \in \{1, 0, -1\}$, but we shall not require this.) Note that Examples A.2 and A.3 both are of this type, with $(\chi, \rho) = (0, \rho)$ and $(1, 0)$ (or $(\alpha, 0)$), respectively.

Note that $\rho = \lambda_0 > 0$, while χ can be any real number. As in Example 6.4, if $\chi < 0$, we have to assume that $\rho = m|\chi|$ for a non-negative integer m (and $\mathcal{X}_0 \leqslant m$); then $\lambda_m = 0$ and the process stops when it reaches m, so the values λ_k, $k > m$, can be ignored.

Theorem A.5. *Let $(\mathcal{X}_t)$ be a pure birth process with birth rates $\lambda_k = \chi k + \rho$ as in (A.10), for some χ and ρ, started at $\mathcal{X}_0 = 0$. Furthermore, let $\tau \sim \mathrm{Exp}(\alpha)$ be independent of the birth process. Then $X := \mathcal{X}_\tau$ has the distribution, when $\chi \neq 0$,*

$$\begin{aligned} \mathbb{P}(X = k) &= \frac{\alpha \prod_{i=0}^{k-1} (\chi i + \rho)}{\prod_{i=0}^{k} (\chi i + \rho + \alpha)} \\ &= \frac{\alpha}{\rho + \alpha} \cdot \frac{\langle \rho/\chi \rangle_k}{\langle (\rho+\alpha)/\chi + 1 \rangle_k}, \qquad k \geqslant 0. \end{aligned} \tag{A.11}$$

Thus, using the notation in Definition B.1, X has the hypergeometric distribution $\mathsf{HG}\big(\rho/\chi, 1; (\rho+\alpha)/\chi + 1\big)$.

In the special case $\chi = 0$ (so $\lambda_k = \rho$ is constant), we have instead

$$\mathbb{P}(X = k) = \frac{\alpha}{\rho + \alpha} \left(\frac{\rho}{\rho + \alpha} \right)^k, \qquad k \geqslant 0. \tag{A.12}$$

Thus, in this case X has the geometric distribution $\mathrm{Ge}_0\big(\alpha/(\alpha+\rho)\big)$.

Proof. An immediate corollary of Theorem A.4. □

In the linear case, it is also easy to find the distribution of $\mathcal{X}_t$ for a fixed t. We begin with the expectation.

Theorem A.6. *Let $(\mathcal{X}_t)$ be a pure birth process with birth rates $\lambda_k = \chi k + \rho$ as in (A.10), for some χ and ρ, started at $\mathcal{X}_0 = x_0$. Then, for every $t \geqslant 0$,*

$$\mathbb{E}\,\mathcal{X}_t = \begin{cases} \left(\frac{\rho}{\chi} + x_0\right) e^{\chi t} - \frac{\rho}{\chi} & \chi \neq 0, \\ \rho t + x_0, & \chi = 0. \end{cases} \tag{A.13}$$

Consequently, for every χ and every $t \geqslant 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\,\mathcal{X}_t = (\rho + \chi x_0) e^{\chi t}. \tag{A.14}$$

Proof. Since $\mathcal{X}_t$ grows with a rate that is a linear function $\chi k + \rho$ of the current state k, its expectation $\mathbb{E}\,\mathcal{X}_t$ grows at rate $\chi\,\mathbb{E}\,\mathcal{X}_t + \rho$, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\,\mathcal{X}_t = \chi\,\mathbb{E}\,\mathcal{X}_t + \rho. \tag{A.15}$$

This differential equation, with the initial value $\mathbb{E}\,\mathcal{X}_0 = x_0$, has the solution (A.13). (The reader that finds this argument too informal may note that $\mathcal{X}_t - \int_0^t (\chi \mathcal{X}_s + \rho)\,\mathrm{d}s$ is a martingale, and take the expectation to obtain (A.15).) □

We give the distribution of $\mathcal{X}_t$ only for the case $x_0 = 0$, leaving the general case to the reader.

Theorem A.7. *Let $(\mathcal{X}_t)$ be a pure birth process with birth rates $\lambda_k = \chi k + \rho$ as in (A.10), for some $\chi \neq 0$ and ρ, started at $\mathcal{X}_0 = 0$. Then, for all $t \geqslant 0$ and $k \geqslant 0$,*

$$\mathbb{P}(\mathcal{X}_t = k) = \frac{\langle \rho/\chi \rangle_k}{k!} (1 - e^{-\chi t})^k e^{-\rho t} = (-1)^k \binom{-\rho/\chi}{k} (1 - e^{-\chi t})^k e^{-\rho t}. \tag{A.16}$$

Equivalently, the probability generating function is given by

$$\mathbb{E}\, z^{\mathcal{X}_t} = e^{-\rho t} \big(1 - z(1 - e^{-\chi t})\big)^{-\rho/\chi} = \left(\frac{e^{-\chi t}}{1 - z(1 - e^{-\chi t})}\right)^{\rho/\chi}. \tag{A.17}$$

In the case $\chi > 0$, this says $\mathcal{X}_t \sim \mathrm{NBin}(\rho/\chi, e^{-\chi t})$, a negative binomial distribution.

In the case $\chi < 0$, when necessarily $\rho = m|\chi|$ for some positive integer m, (A.16) can be written

$$\mathbb{P}(\mathcal{X}_t = k) = \binom{m}{k} (1 - e^{-|\chi| t})^k e^{-(m-k)|\chi| t}, \tag{A.18}$$

and thus $\mathcal{X}_t \sim \mathrm{Bin}(\rho/|\chi|, 1 - e^{-|\chi| t})$, a binomial distribution.

In the case $\chi = 0$, we have instead

$$\mathbb{P}(\mathcal{X}_t = k) = \frac{(\rho t)^k}{k!} e^{-\rho t} \tag{A.19}$$

and

$$\mathbb{E}\, z^{\mathcal{X}_t} = e^{(z-1)\rho t}. \tag{A.20}$$

Thus $\mathcal{X}_t \sim \mathrm{Po}(\rho t)$, a Poisson distribution, when $\chi = 0$.

Note the well-known fact for the Poisson process in Example A.2, we thus have (A.19), yielding a Poisson distribution $\mathrm{Po}(\rho t)$.

Proof. With $p_k(t) := \mathbb{P}(\mathcal{X}_t = k)$, we have the initial values $p_0(0) = 1$, $p_k(0) = 0$, $k \geqslant 1$, and the Kolmogorov forward equation (with $p_{-1} := 0$)

$$p_k'(t) = \lambda_{k-1}p_{k-1}(t) - \lambda_k p_k(t) = (\chi(k-1)+\rho)p_{k-1}(t) - (\chi k+\rho)p_k(t). \quad \text{(A.21)}$$

This system of differential equations determines $p_k(t)$ uniquely for $k = 0, 1, \dots$, and it is easily verified that the functions in (A.16) and (A.19) give a solution. The probability generating functions (A.17) and (A.20) follow by summation. (Alternatively, one may derive and solve a partial differential for $g(z,t) := \mathbb{E}\, z^{\mathcal{X}_t}$.)

In the case $\chi < 0$ and $\rho = m|\chi| = -m\chi$, (A.16) can be written as

$$\mathbb{P}(\mathcal{X}_t = k) = \binom{m}{k}(e^{|\chi|t}-1)^k e^{-m|\chi|t} \quad \text{(A.22)}$$

and (A.18) follows. □

Example A.8. Let $\mathcal{Y}_t$ be a Yule process, see Example A.3. This birth process starts at $\mathcal{Y}_0 = 1$, so we consider instead $\mathcal{Y}_t - 1$ which is a pure birth process started at 0, with birth rates $\lambda_k = k+1$, $k \geqslant 0$. Theorem A.7 applies with $\chi = \rho = 1$ and yields the well-known result $\mathcal{Y}_t - 1 \sim \mathrm{NBin}(1, e^{-t}) = \mathrm{Ge}_0(e^{-t})$, and thus

$$\mathcal{Y}_t \sim \mathrm{Ge}_1(e^{-t}), \qquad t \geqslant 0. \quad \text{(A.23)}$$

For the expectation we get from (A.23), or directly from Theorem A.6 with $\chi = 1$, $\rho = 0$ and $x_0 = 1$, $\mathbb{E}\,\mathcal{Y}_t = e^t$.

Similarly, Theorem A.5 applied to $\mathcal{Y}_t - 1$ shows that if $\tau \sim \mathrm{Exp}(\alpha)$ is independent of the Yule process, then

$$\mathcal{Y}_\tau - 1 \sim \mathsf{HG}(1, 1; \alpha+2). \quad \text{(A.24)}$$

Appendix B: Hypergeometric functions and distributions

Recall that the hypergeometric function $F(a,b;c;z)$ (also denoted by ${}_2F_1(a,b;c;z)$) is defined by the sum

$$F(a,b;c;z) := \sum_{n=0}^{\infty} \frac{\langle a\rangle_n \langle b\rangle_n}{\langle c\rangle_n} \cdot \frac{z^n}{n!}, \quad \text{(B.1)}$$

see e.g. [97, §15.2] or [53, §5.5]. In general, the parameters a, b, c can be arbitrary complex numbers (except that $c = 0, -1, -2, \dots$ is allowed only in special cases), and z may be a complex number, but we are here only interested in real a, b, c and z. If $a \in \mathbb{Z}_{\leqslant 0}$ or $b \in \mathbb{Z}_{\leqslant 0}$, then the hypergeometric terms in (B.1) vanish for $n > |a|$ or $n > |b|$, respectively, so $F(a,b;c;z)$ is a polynomial; otherwise the series (B.1) converges for $|z| < 1$ and diverges for $|z| > 1$. (The hypergeometric

function $F(a,b;c;z)$ extends by analytic continuation to $z \in \mathbb{C} \setminus [1,\infty)$, but we have no use for this extension here.)

The hypergeometric series (B.1) converges for $z=1$ if and only if $a \in \mathbb{Z}_{\leqslant 0}$, $b \in \mathbb{Z}_{\leqslant 0}$ (in these cases the sum is finite, as said above) or $\operatorname{Re}(c-a-b)>0$, and then its sum is, as shown by Gauss [52], see also [97, (15.4.20), (15.4.24)],

$$F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \tag{B.2}$$

We say that a random variable has a *(general) hypergeometric distibution* if its probability generating function is, up to a normalization constant, a hypergeometric function $F(a,b;c;z)$, for some a,b,c. We denote such a distribution by $\mathsf{HG}(a,b;c)$. (There seems to be no standard notation.) Some such distributions appear above in the study of random trees, and we give here some general properties and examples of such distributions, as a background and for easy reference. See further e.g. Johnson, Kemp and Kotz [74, Chapter 6] and the references given there.

We repeat the definition somewhat more formally:

Definition B.1. The *general hypergeometric distribution* $\mathsf{HG}(a,b;c)$ is the distribution of a non-negative integer-valued random variable X such that

$$\mathbb{P}(X=k) = C\frac{\langle a\rangle_k \langle b\rangle_k}{\langle c\rangle_k \, k!}, \qquad k \geqslant 0, \tag{B.3}$$

for some constant C. Equivalently, the probability generating function is

$$\mathbb{E}\, z^X = CF(a,b;c;z). \tag{B.4}$$

By (B.4), the normalizing constant C is necessarily $1/F(a,b;c;1)$ and thus, by (B.2),

$$C = \frac{1}{F(a,b;c;1)} = \frac{\Gamma(c-a)\Gamma(c-b)}{\Gamma(c)\Gamma(c-a-b)}. \tag{B.5}$$

Furthermore, the probability generating function (B.4) is

$$\mathbb{E}\, z^X = \frac{F(a,b;c;z)}{F(a,b;c;1)}. \tag{B.6}$$

Note the symmetry $\mathsf{HG}(a,b;c) = \mathsf{HG}(b,a;c)$.

Remark B.2. The definition (B.3) is equivalent to

$$\frac{\mathbb{P}(X=k+1)}{\mathbb{P}(X=k)} = \frac{(k+a)(k+b)}{(k+c)(k+1)}, \qquad k \geqslant 0, \tag{B.7}$$

with a suitable interpretation if $\mathbb{P}(X=k)=0$.

A hypergeometric distribution $\mathsf{HG}(a,b;c)$ does not exist for all real parameters a,b,c. We see from (B.3) and (B.5) that a necessary and sufficient condition

for the existence of $\mathsf{HG}(a,b;c)$ is that $\langle a\rangle_k\langle b\rangle_k/\langle c\rangle_k \geqslant 0$ and that $F(a,b;c;1) < \infty$. We do not give precise necessary and sufficient conditions for this here, see e.g. [74], but we note the following cases where $\mathsf{HG}(a,b;c)$ exists; these comprise all cases of interest to us (and to others as far as we know), if we recall that a and b can be interchanged.

(i) $a,b>0$, $c>a+b$. The distribution has support $\mathbb{Z}_{\geqslant 0}$. (Examples are (6.33) and Theorem A.5 for $\chi>0$.)

(ii) $a\in\mathbb{Z}_{\leqslant 0}$, $b>0$, $c<a+1$. The distribution has support $\{0,\dots,|a|\}$. (Examples are (6.43) and Theorem A.5 for $\chi<0$.)

(iii) $a\in\mathbb{Z}_{\leqslant 0}$, $b<a+1$, $c>0$. The distribution has support $\{0,\dots,|a|\}$. (One example is Example B.4 below. Typically, as there, we have both $a,b\in\mathbb{Z}_{\leqslant 0}$ and $c>0$; then, if we do not assume $b\leqslant a$, the support is $\{0,\dots,\min(|a|,|b|)\}$.)

Remark B.3. As said above, the hypergeometric function (B.1) in general does not exist when $c\in\mathbb{Z}_{\leqslant 0}$. However, it is still possible to define $\mathsf{HG}(a,b;c)$ is some cases. We assume that also $a\in\mathbb{Z}_{\leqslant 0}$. (Of course, the case $b\in\mathbb{Z}_{\leqslant 0}$ is similar, by symmetry.) If $c<a$ (and $b>0$, included in (ii)), there is no problem with the definition (B.3), letting X have support $\{0,\dots,|a|\}$. Also the case $c=a$ is interpreted in this way; in particular, note that $\mathsf{HG}(-n,1;-n)$ is the uniform distribution on $\{0,\dots,|a|\}$ [74, §6.10.1].

In the case $a,c\in\mathbb{Z}_{\leqslant 0}$ and $a<c$, (B.3) does not make sense, since it would yield infinite values. We can extend the definition to this case by rewriting (B.3) as

$$\mathbb{P}(X=k)=C'\frac{\langle a\rangle_k\langle b\rangle_k}{\Gamma(c+k)\,k!},\qquad k\geqslant 0, \tag{B.8}$$

cf. (2.1), where now $C'=\Gamma(c)C=\Gamma(c)/F(a,b;c;1)$, which exists by continuity also for $c\in\mathbb{Z}_{\leqslant 0}$. (The function $F(a,b;c;z)/\Gamma(c)$ is denoted $\mathbf{F}(a,b;c;z)$ in [97].) With this interpretation, we can define $\mathsf{HG}(a,b;c)$ also in this case, for suitable b. (See Example B.4 for an example.) Note that for this case, $\mathbb{P}(X=k)=0$ for $0\leqslant k\leqslant |c|$. It is easily verified that if $X\sim\mathsf{HG}(a,b;c)$ with $c=-n\in\mathbb{Z}_{\leqslant 0}$, then $X=X'+n+1$ with $X'\sim\mathsf{HG}(a+n+1,b+n+1;n+2)$. (This makes it possible to reduce to the case $c\notin\mathbb{Z}_{\leqslant 0}$.)

Example B.4. The classical hypergeometric distribution describes the number of red balls in a sample of m balls drawn (without replacement) from a population of N balls, of which n are red; see e.g. [47, Section II.6]. (Here n,m,N are integers with $0\leqslant n\leqslant N$ and $0\leqslant m\leqslant N$.) The probability function is

$$\begin{aligned}\mathbb{P}(X=k)&=\frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}\\&=\frac{n!\,(N-n)!\,(N-m)!\,m!}{k!\,(n-k)!\,(m-k)!\,(N-m-n+k)!\,N!}\\&=\frac{(N-m)!\,(N-n)!}{N!\,(N-m-n)!}\cdot\frac{\langle -m\rangle_k\langle -n\rangle_k}{\langle N-m-n+1\rangle_k}.\end{aligned} \tag{B.9}$$

(In the case $m+n>N$, the final line in (B.9) is not valid, but it can be interpreted as in Remark B.3.) This is thus the distribution $\mathsf{HG}(-m,-n;N-m-n+1)$.

Remark B.5. In the case $a=-n\in\mathbb{Z}_{\leqslant 0}$, when $X\sim\mathsf{HG}(a,b;c)$ satisfies $0\leqslant X\leqslant n$, it is easy to see that $n-X\sim\mathsf{HG}(a,a-c+1;a-b+1)=\mathsf{HG}(-n,1-n-c;1-n-b)$. (Natural examples can be seen from Example B.4.)

Remark B.6. Most hypergeometric distributions in this paper are of the special form $\mathsf{HG}(a,1;c)$, i.e., with $b=1$. In this case, (B.3) and (B.5) simplify and yield

$$\mathbb{P}(X=k)=\frac{c-a-1}{c-1}\cdot\frac{\langle a\rangle_k}{\langle c\rangle_k}. \tag{B.10}$$

Such distributions are also called *Waring distributions* [74, §6.10.4].

As said above a hypergeometric variable $X\sim\mathsf{HG}(a,b;c)$ with $a\in\mathbb{Z}_{\leqslant 0}$ or $b\in\mathbb{Z}_{\leqslant 0}$ is bounded, and thus has moments of all orders. If $a,b\notin\mathbb{Z}_{\leqslant 0}$, the distribution has a power-law tail, and thus only a finite number of moments. We give a precise asymptotic formula for $\mathbb{P}(X=k)$ and then formulas for (factorial) moments.

Theorem B.7. *Suppose that $X\sim\mathsf{HG}(a,b;c)$, and that $a,b\notin\mathbb{Z}_{\leqslant 0}$. Then, as $k\to\infty$,*

$$\mathbb{P}(X=k)\sim\frac{\Gamma(c-a)\Gamma(c-b)}{\Gamma(a)\Gamma(b)\Gamma(c-a-b)}k^{a+b-c-1}. \tag{B.11}$$

Proof. By (B.3) (or (B.8)), (B.5) and (2.1),

$$\mathbb{P}(X=k)=\frac{\Gamma(c-a)\Gamma(c-b)}{\Gamma(a)\Gamma(b)\Gamma(c-a-b)}\cdot\frac{\Gamma(k+a)\Gamma(k+b)}{\Gamma(k+c)\Gamma(k+1)}. \tag{B.12}$$

The result follows since $\Gamma(x+s)/\Gamma(x)\sim x^s$ as $x\to\infty$ for every fixed s. □

Theorem B.8. *Suppose $X\sim\mathsf{HG}(a,b;c)$ and $m\in\mathbb{Z}_{\geqslant 0}$.*

(i) *If $a,b\notin\mathbb{Z}_{\leqslant 0}$ and $m\geqslant c-a-b$, then the moment $\mathbb{E}\,X^m$ and the factorial moment $\mathbb{E}(X)_m$ are infinite.*

(ii) *If $a\in\mathbb{Z}_{\leqslant 0}$, $b\in\mathbb{Z}_{\leqslant 0}$ or $m<c-a-b$, then*

$$\mathbb{E}(X)_m=\frac{\langle a\rangle_m\langle b\rangle_m}{\langle c-a-b-m\rangle_m}. \tag{B.13}$$

In particular, for $a\in\mathbb{Z}_{\leqslant 0}$, $b\in\mathbb{Z}_{\leqslant 0}$ or $c>a+b+1$,

$$\mathbb{E}\,X=\frac{ab}{c-a-b-1} \tag{B.14}$$

and, for $a\in\mathbb{Z}_{\leqslant 0}$, $b\in\mathbb{Z}_{\leqslant 0}$ or $c>a+b+2$,

$$\operatorname{Var}X=\frac{ab(c-a-1)(c-b-1)}{(c-a-b-1)^2(c-a-b-2)}. \tag{B.15}$$

Proof. (i): Follows from Theorem B.7.

(ii): It follows from (B.1), using (B.2) and (2.1), that

$$E(X)_m = \frac{\frac{\mathrm{d}^m}{\mathrm{d}z^m}F(a,b;c;z)|_{z=1}}{F(a,b;c;1)} = \frac{\langle a\rangle_m\langle b\rangle_m}{\langle c\rangle_m}\cdot\frac{F(a+m,b+m;c+m;1)}{F(a,b;c;1)} = \frac{\langle a\rangle_m\langle b\rangle_m}{\langle c-a-b-m\rangle_m}.$$

This proves (B.13). Taking $m=1$ we find (B.14), and the cases $m=1$ and $m=2$ yield (B.15) after a small calculation, using $\operatorname{Var} X = \mathbb{E}(X)_2 + \mathbb{E}\,X - (\mathbb{E}\,X)^2$. □

We see from Theorems A.5 and A.7 that a hypergeometric distribution can arise as a mixture $\mathrm{NBin}(a, e^{-\chi\tau})$ of negative binomial distributions, or a mixture $\mathrm{Bin}(m, 1-e^{-|\chi|\tau})$ of binomial distributions. In this case, $\tau \sim \mathrm{Exp}(\alpha)$, so the random parameter $e^{-|\chi|\tau}$ in the negative binomial or binomial distribution has the beta distribution $B(\alpha/|\chi|, 1)$, and $1-e^{-|\chi|\tau}$ has the beta distribution $B(1, \alpha/|\chi|)$. This extends to mixing using arbitrary beta distributions.

Theorem B.9. (i) *Let $a, b, r > 0$ be real numbers. Let X have a distribution that is a mixture of the negative binomial distribution $\mathrm{NBin}(r,p)$ with $p \sim B(a,b)$. Then $X \sim \mathsf{HG}(r, b; r+a+b)$.*

(ii) *Let $a, b, m > 0$, with m an integer. Let X have a distribution that is a mixture of the binomial distribution $\mathrm{Bin}(m,p)$ with $p \sim B(a,b)$. Then $X \sim \mathsf{HG}(-m, a; 1-b-m)$.*

Proof. The proofs of both parts are similar: we use the definitions of the negative binomial, binomial and beta distributions, evaluate a beta integral and make some manipulations using (2.1) and (2.2). It is not difficult to keep track of the constant factors during the calculations (and, indeed, this is a useful check, which we leave to the reader), but it is simpler to ignore them, since the final factor is determined by (B.5) and thus does not need to be computed; we thus just write $C_1, \dots$ for various constants (depending on the parameters but not on k).

(i):

$$\begin{aligned}\mathbb{P}(X=k) &= \int_0^1 \frac{\langle r\rangle_k}{k!}p^r(1-p)^k\cdot\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,\mathrm{d}p \\ &= \frac{\langle r\rangle_k}{k!}\frac{B(r+a,k+b)}{B(a,b)} = C_1\frac{\langle r\rangle_k\Gamma(b+k)}{k!\,\Gamma(r+a+b+k)} \\ &= C_2\frac{\langle r\rangle_k\langle b\rangle_k}{\langle r+a+b\rangle_k k!}.\end{aligned} \tag{B.16}$$

(ii):

$$\mathbb{P}(X=k) = \int_0^1 \frac{(m)_k}{k!}p^k(1-p)^{m-k}\cdot\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,\mathrm{d}p$$

$$= \frac{(m)_k}{k!}\frac{B(k+a,m-k+b)}{B(a,b)} = C_3\frac{(m)_k\Gamma(k+a)\Gamma(b+m-k)}{k!} \tag{B.17}$$

$$= C_4\frac{(m)_k\Gamma(k+a)}{(b+m-1)_k k!} = C_5\frac{\langle -m\rangle_k\langle a\rangle_k}{\langle 1-b-m\rangle_k k!}. \qquad \square$$

Remark B.10. In this context the resulting hypergeometric distributions are known as *beta-negative binomial distributions* and *beta-binomial distributions*. (Note that we can obtain any distribution of type (i) or (ii) above.) There are also several other names used for various cases of the general hypergeometric distribution, see [74]; in particular, case (ii) is sometimes called *negative hypergeometric*.

Example B.11. Yule [108] considered a simple model of evolution, where each existing species creates new species in the same genus with a constant rate λ_s, and also (independently) new species in new genera with another rate λ_g. What is the limiting distribution of the number of species in a random genus?

Note that the evolution of all species, ignoring their genus, is a Crump–Mode–Jagers branching process, where each individual gets children according to a Poisson process with intensity $\lambda_s+\lambda_g$. Hence, assuming that we start with a single species, the total number of species forms a Yule process with rate $\lambda_s+\lambda_g$, see Example A.3. Similarly, the number of species in the same genus as the root (the original species) forms a Yule process with rate λ_s.

One way to treat this problem is to consider each genus as an individual in a Crump–Mode–Jagers process, where each individual has an internal Yule process $\mathcal{Y}_t$ with rate λ_s (the number of species in the genus), and new births occur with rate $\lambda_g\mathcal{Y}_t$. Since $\mathbb{E}\,\mathcal{Y}_t = e^{\lambda_s t}$, see Theorem A.6, the offspring process has intensity $\mu(\mathrm{d}t) = \lambda_g e^{\lambda_s t}\,\mathrm{d}t$, from which it follows that (5.4) holds with the Malthusian parameter $\alpha = \lambda_s+\lambda_g$. The assumptions (A1)–(A5) hold, and Theorem 5.14 shows that the number of species in a random genus converges in distribution to $\mathcal{Y}_{\overline{\tau}}$, the number of species in the root at time $\overline{\tau}\sim\mathrm{Exp}(\lambda_s+\lambda_g)$. The Yule process $\mathcal{Y}_t$ starts as 1, but we may as in Example A.8 (which is the case $\lambda_s=1$) apply Theorem A.5 to $\mathcal{Y}_t-1$, with $\chi=\rho=\lambda_s$, and it follows that the asymptotic distribution of the number of species in a given genus is $1+\mathsf{HG}(1,1;3+\lambda_g/\lambda_s)$. (One can, as said above, also use Theorem B.9(i), since Theorem A.7 implies that $\mathcal{Y}_t-1$ has the geometric distribution $\mathrm{NBin}(1,e^{-\lambda_s t}) = \mathrm{Ge}_0(e^{-\lambda_s t})$, and $e^{-\lambda_s\overline{\tau}}\sim B((\lambda_s+\lambda_g)/\lambda_s,1)$.)

This result was found by Yule [108] (by a different method), and a distribution of the form $1+\mathsf{HG}(1,1;c)$ is therefore called a *Yule distribution*, see further [106] and [74, §6.10.3]. (Here $c>2$. Often one writes $c=2+\rho$, with $\rho>0$; in the present example, thus $\rho=(\lambda_s+\lambda_g)/\lambda_s$.) Note that the case $c=3$ appears in (6.2) and (shifted to $\mathsf{HG}(1,1;3)$) in (7.6).

An alternative method to treat this example is to consider the Yule process (with rate $\lambda_s+\lambda_g$) of all species. Call the first species in each genus the *progenitor* of the genus, and give each progenitor a mark; then each species (except the original one) is marked with probability $p=\lambda_g/(\lambda_s+\lambda_g)$, and these marks are

independent of each other and of everything else. Hence, we obtain the same asymptotic distribution of fringe trees (except for the mark at the root) if we sample a random progenitor as if we sample a random species, cf. Section 12. Using Theorem 5.14 it follows that the number of species in a random genus converges in distribution to the number of descendants in the same genus of a fixed individual stopped at a random age $\tau \sim \mathrm{Exp}(\lambda_s + \lambda_g)$. This yields the same result as above.

Note also that if we erase the edges between different genera and only keep the edges between species in the same genus in Yule's model, we obtain a growing forest. If we let F_n be this forest when it has reached n nodes, we obtain a growing forest process which is the same as UGROW defined by Devroye, McDiarmid and Reed [38]. Results for the size of the subtree rooted at a given node in UGROW are given by Devroye, McDiarmid and Reed [38] and Pakes [99].

Example B.12. Let n indistinguishable balls be distributed in $m \geqslant 2$ labelled boxes, so that each of the $\binom{n+m-1}{m-1}$ possibilities has the same probability. (This is called *Bose–Einstein statistics*, see e.g. [47, page 39] or [69, Example 12.2].) Equivalently, we consider a uniformly random composition of n into m (possibly empty) parts.

The number X_1 of balls in a given box, say box 1, has the distribution

$$\mathbb{P}(X_1 = k) = \frac{\binom{n-k+m-2}{m-2}}{\binom{n+m-1}{m-1}} = \frac{(m-1)(n)_k}{(n+m-1)(n+m-2)_k}, \qquad k = 0, \dots, n, \tag{B.18}$$

which shows that $X_1 \sim \mathsf{HG}(-n, 1; 2-n-m)$. See [74, Chapter 10.4.2].

Example B.13. Pólya's urn contains balls of different colours. We draw a ball uniformly at random and replace it together with c new balls of the same colour. This is repeated n times. Let W be the number of white balls drawn, assuming that the urn initially contains w white and b black (or non-white) balls.

We assume $c \neq 0$, to avoid the trivial case $c = 0$ when W has a binomial distribution, but $c < 0$ is allowed, meaning that balls are removed. In particular, $c = -1$ gives drawing without replacement, when W has the classical hypergeometric distribution in Example B.4. (It is natural to let c, b, w be integers, but the model has a natural interpretation also for real values of these parameters, see e.g. [67, Remark 4.2].) In the case $c < 0$, we assume that b, w and n are such that we never can be required to remove a ball of a colour that is no longer present, or draw a ball when the urn is empty.

It is easy to see that $\mathbb{P}(X = k)$ is proportional to

$$\begin{aligned}\binom{n}{k} \prod_{i=1}^{k} (w + (i-1)c) \prod_{j=1}^{n-k} (b + (j-1)c) &= \binom{n}{k} c^n \langle w/c \rangle_k \langle b/c \rangle_{n-k} \\ &= c^n \langle b/c \rangle_n \frac{(n)_k \langle w/c \rangle_k}{k!\,(b/c + n - 1)_k}\end{aligned} \tag{B.19}$$

and thus $W \sim \mathsf{HG}(-n, w/c; 1 - n - b/c)$, see [74, §6.2.4].

For a connection with Example B.12, suppose instead that the urn starts with one ball each of m colours (including white), and that $c = 1$. The number W of white balls drawn is the same as if we start with 1 white and $m-1$ black balls, and thus $W \sim \mathsf{HG}(-n, 1; 2-n-m)$. On the other hand, it is easy to see by induction, that for each n, the composition of the urn is uniform over all possible colour combinations. Thus W has the same distribution as X_1 in Example B.12 (with colours corresponding to boxes).

Appendix C: Order statistics of exponential variables

Let, for $1 \leqslant k \leqslant m$, $V_{m,k}$ be the k:th smallest of m i.i.d. Exp(1) random variables $E_1, \dots, E_m$; further, let $V_{m,0} = 0$.

Theorem C.1. *Let $1 \leqslant k \leqslant m$. Then the following holds.*

(i) *If $Y_i \sim \mathrm{Exp}(i)$ are independent, then*

$$V_{m,k} \stackrel{\mathrm{d}}{=} \sum_{i=m-k+1}^{m} Y_i. \tag{C.1}$$

(ii) *$V_{m,k}$ has the density function*

$$\frac{m!}{(m-k)!\,(k-1)!} e^{-(m-k+1)x} (1-e^{-x})^{k-1}, \qquad x > 0. \tag{C.2}$$

(iii) *$e^{-V_{m,k}}$ has the beta distribution $B(m-k+1, k)$.*

(iv) *For any $\theta > -(m-k+1)$,*

$$\mathbb{E}\, e^{-\theta V_{m,k}} = \prod_{i=m-k+1}^{m} \frac{i}{i+\theta} = \frac{(m)_k}{(m+\theta)_k}. \tag{C.3}$$

Proof. (i): Consider m independent exponential clocks that strike at $E_1, \dots, E_m$. As is well-known, by the lack of memory for exponential distributions, the waiting times $W_i := V_{m,i} - V_{m,i-1}$ between the strikes are independent and exponentially distributed, with $W_i \sim \mathrm{Exp}(m-i+1)$ since there are $m-i+1$ clocks left. Let $Y_i := W_{m-i+1} \sim \mathrm{Exp}(i)$ and note that $V_{m,k} = \sum_{i=1}^{k} W_i$.

(iii): Let $U_i := e^{-E_i}$. Then $U_1, \dots, U_m \sim U(0,1)$ are m i.i.d. uniform random variables, and $e^{-V_{m,k}}$ is the k:th largest of them; it is well-known that this has the beta distribution $B(m-1+k, k)$.

(ii): By (iii), $e^{-V_{m,k}}$ has the density

$$\frac{\Gamma(m+1)}{\Gamma(m-k+1)\Gamma(k)} x^{m-k}(1-x)^{k-1} = \frac{m!}{(m-k)!\,(k-1)!} x^{m-k}(1-x)^{k-1}, \tag{C.4}$$

for $0 < x < 1$, and (C.2) follows.

(iv): A simple consequence of (i), or alternatively of (iii). □

Acknowledgements

We thank Peter Jagers and Olle Nerman for helpful comments on branching processes and Shankar Bhamidi for suggesting the use of the limiting sin-tree to compute the constant in Theorem 11.3.

References

[1] David Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266. MR1102319

[2] Søren Asmussen and Heinrich Hering, *Branching Processes*. Birkhäuser, Boston, MA, 1983. MR0701538

[3] K. B. Athreya, Preferential attachment random graphs with general weight function. *Internet Math.* **4** (2007), no. 4, 401–418. MR2522950

[4] Krishna B. Athreya, Arka P. Ghosh and Sunder Sethuraman, Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci.* **118** (2008), no. 3, 473–494. MR2450248

[5] Ricardo A. Baeza-Yates. Some average measures in m-ary search trees. *Inform. Process. Lett.* **25** (1987), no. 6, 375–381. MR0905782

[6] Albert-László Barabási and Réka Albert, Emergence of scaling in random networks. *Science* **286** (1999), no. 5439, 509–512. MR2091634

[7] Jürgen Bennies and Götz Kersting, A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** (2000), no. 3, 777–803. MR1785529

[8] François Bergeron, Philippe Flajolet and Bruno Salvy, Varieties of increasing trees. *CAAP'92 (Rennes, 1992)*, 24–48, Lecture Notes in Comput. Sci. 581, Springer, Berlin, 1992 MR1251994

[9] Jean Bertoin, *Random Fragmentation and Coagulation Processes*. Cambridge University Press, Cambridge, 2006.

[10] Shankar Bhamidi, Universal techniques to analyze preferential attachment tree and networks: Global and local analysis. Preprint, 2007. `http://www.unc.edu/~bhamidi/preferent.pdf`

[11] J. D. Biggins, The first- and last-birth problems for a multitype age-dependent branching process. *Advances in Appl. Probability* **8** (1976), no. 3, 446–459.

[12] J. D. Biggins, Chernoff's theorem in the branching random walk. *J. Appl. Probability* **14** (1977), no. 3, 630–636.

[13] J. D. Biggins, The growth and spread of the general branching random walk. *Ann. Appl. Probab.* **5** (1995), no. 4, 1008–1024. MR1384364

[14] J. D. Biggins, How fast does a general branching random walk spread? *Classical and Modern Branching Processes (Minneapolis, MN, 1994)*, 19–39, Springer, New York, 1997.

[15] J. D. Biggins and D. R. Grey, A note on the growth of random trees. *Statist. Probab. Lett.* **32** (1997), no. 4, 339–342.

[16] Michael G. B. Blum and Olivier François, Minimal clade size and external branch length under the neutral coalescent. *Adv. in Appl. Probab.* **37** (2005), no. 3, 647–662.

[17] Béla Bollobás, Oliver Riordan, Joel Spencer and Gábor Tusnády, The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** (2001), no. 3, 279–290.

[18] Miklós Bóna, k-protected vertices in binary search trees. *Adv. in Appl. Math.* **53** (2014), 1–11.

[19] Miklós Bóna and Boris Pittel, On a random search tree: asymptotic enumeration of vertices by distance from leaves. Preprint, 2014. `arXiv:1412.2796`

[20] Nicolas Broutin and Luc Devroye, Large deviations for the weighted height of an extended class of trees. *Algorithmica* **46** (2006), nos 3–4, 271–297.

[21] Nicolas Broutin, Luc Devroye and Erin McLeish, Weighted height of random trees. *Acta Inform.* **45** (2008), no. 4, 237–277. MR2403296

[22] Nicolas Broutin, Luc Devroye, Erin McLeish, and Mikael de la Salle, The height of increasing trees. *Random Structures Algorithms* **32** (2008), no. 4, 494–518.

[23] Huilan Chang and Michael Fuchs, Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.* **60** (2010), no. 4, 481–512.

[24] Brigitte Chauvin and Michael Drmota, The random multisection problem, travelling waves and the distribution of the height of m-ary search trees. *Algorithmica* **46** (2006), nos 3–4, 299–327.

[25] Birgitte Chauvin and Nicolas Pouyanne, m-ary search trees when $m \geq 27$: a strong asymptotics for the space requirement. *Random Structures Algorithms* **24** (2004), 133–154. MR2035872

[26] Hua-Huai Chern and Hsien-Kuei Hwang, Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms* **19** (2001), nos 3–4, 316–358. MR1871558

[27] Hua-Huai Chern, Hsien-Kuei Hwang and Tsung-Hsi Tsai, An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms. *J. Algorithms* **44** (2002), no. 1, 177–225.

[28] Maria Deijfen. Random networks with preferential growth and vertex death. *J. Appl. Probab.* **47** (2010), no. 4, 1150–1163.

[29] Luc Devroye, A note on the height of binary search trees. *J. Assoc. Comput. Mach.* **33** (1986), 489–498.

[30] Luc Devroye, Branching processes in the analysis of the heights of trees. *Acta Inform.* **24** (1987), 277–298. MR0894557

[31] Luc Devroye, On the height of random m-ary search trees. *Random Structures Algorithms* **1** (1990), no. 2, 191–203.

[32] Luc Devroye, Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2** (1991), no. 3, 303–315.

[33] Luc Devroye, On the expected height of fringe-balanced trees. *Acta Inform.* **30** (1993), 459–466.

[34] Luc Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic*

Discrete Mathematics, 249–314, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer, Berlin, 1998.

[35] Luc Devroye, Universal limit laws for depth in random trees. *SIAM J. Comput.* **28** (1998), 409–432.

[36] Luc Devroye, Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32** (2002/03), no. 1, 152–171.

[37] Luc Devroye and Svante Janson, Protected nodes and fringe subtrees in some random trees. *Electronic Communications Probability* **19** (2014), no. 6, 1–10.

[38] Luc Devroye, Colin McDiarmid and Bruce Reed, Giant components for two expanding graph processes. *Mathematics and Computer Science, II (Versailles, 2002)*, 161–173, Birkhäuser, Basel, 2002.

[39] R. A. Doney, A limit theorem for a class of supercritical branching processes. *Journal of Applied Probability* **9** (1972), no. 4, 707–724.

[40] Michael Drmota, *Random Trees*. Springer, Vienna, 2009.

[41] Michael Drmota, An analytic approach to the height of binary search trees II. *J. ACM* **50** (2003), no. 3, 333–374. MR2146358

[42] Michael Drmota, Michael Fuchs and Yi-Wen Lee, Limit laws for the number of groups formed by social animals under the extra clustering model. (Extended abstract.) *Proceedings, 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA'14 (Paris, 2014)*, 73–84, *DMTCS Proceedings*, 2014.

[43] Michael Drmota, Bernhard Gittenberger, Alois Panholzer, Helmut Prodinger and Mark Daniel Ward, On the shape of the fringe of various types of random trees. *Math. Methods Appl. Sci.* **32** (2009), no. 10, 1207–1245.

[44] Michael Drmota, Svante Janson and Ralph Neininger, A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** (2008), no. 1, 288–333.

[45] Eric Durand, Michael G. B. Blum and Olivier François, Prediction of group patterns in social mammals based on a coalescent model. *J. Theoret. Biol.* **249** (2007), no. 2, 262–270.

[46] Eric Durand and Olivier François, Probabilistic analysis of a genealogical model of animal group patterns. *J. Math. Biol.* **60** (2010), no. 3, 451–468.

[47] William Feller, *An Introduction to Probability Theory and Its Application*, volume I, third edition, Wiley, New York, 1968.

[48] James Allan Fill and Nevin Kapur, Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Structures Algorithms* **26** (2005), no. 4, 359–391.

[49] Philippe Flajolet, Xavier Gourdon and Conrado Martínez, Patterns in random binary search trees. *Random Structures Algorithms* **11** (1997), no. 3, 223–244.

[50] Michael Fuchs, Subtree sizes in recursive trees and binary search trees: Berry–Esseen bounds and Poisson approximations. *Combin. Probab. Comput.* **17** (2008), no. 5, 661–680.

[51] J. L. Gastwirth and P. K. Bhattacharya, Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.* **32** (1984), no. 3, 527–536.

[52] Carlo Friderico [Carl Friedrich] Gauss, Disquisitiones generales circa seriem infinitam $1 + \frac{\alpha\beta}{1.\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1.2.\gamma(\gamma+1)}xx + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1.2.3.\gamma(\gamma+1)(\gamma+2)}x^3 +$ etc., pars prior. *Commentationes societatis regiae scientiarum Gottingensis recentiores* II (1813). Reprinted in *Werke*, Vol. 3, 123–162, Göttingen, 1863. http://gdz.sub.uni-goettingen.de/dms/load/toc/?PPN=PPN235999628

[53] Ronald L. Graham, Donald E. Knuth and Oren Patashnik, *Concrete Mathematics*. 2nd ed., Addison-Wesley, Reading, MA, 1994.

[54] Allan Gut, *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[55] Theodore E. Harris, *The Theory of Branching Processes*. Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[56] Axel Heimbürger, Asymptotic distribution of two-protected nodes in m-ary search trees. Master thesis, Stockholm University and KTH, 2014. diva-portal.org/smash/get/diva2:748258/FULLTEXT01.pdf

[57] Pascal Hennequin, Analyse en moyenne d'algorithmes, tri rapide et arbres de recherche. Ph.D. dissertation, École Polytechnique, Palaiseau, 1991. http://algo.inria.fr/AofA/Research/src/Hennequin.PhD.html

[58] Remco van der Hofstad, *Random Graphs and Complex Networks. Vol. I*. Lecture notes. October 20, 2014 version. http://www.win.tue.nl/~rhofstad/NotesRGCN.html

[59] Cecilia Holmgren and Svante Janson, Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.* **20** (2015), no. 4, 1–51.

[60] Cecilia Holmgren and Svante Janson, Asymptotic distribution of two-protected nodes in ternary search trees. *Electron. J. Probab.* **20** (2015), no. 9, 1–20.

[61] Cecilia Holmgren and Svante Janson, Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees. arXiv:1601.03691

[62] Cecilia Holmgren, Svante Janson and Matas Šileikis, Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. arXiv:1603.08125

[63] Hsien-Kuei Hwang. Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31** (2003), no. 2, 609–629.

[64] Peter Jagers, *Branching Processes with Biological Applications*. John Wiley & Sons, London, 1975.

[65] Peter Jagers and Olle Nerman, The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** (1984), no. 2, 221–259.

[66] Peter Jagers and Olle Nerman, The asymptotic composition of supercritical multi-type branching populations. *Séminaire de Probabilités, XXX*, 40–54, Lecture Notes in Math. 1626, Springer, Berlin, 1996.

[67] Svante Janson, Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.
[68] Svante Janson, Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms* **26** (2005), nos 1–2, 69–83.
[69] Svante Janson, Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probability Surveys* **9** (2012), 103–252.
[70] Svante Janson, Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Structures Algorithms* **48** (2016), no. 1, 57–101.
[71] Svante Janson, Maximal clades in random binary search trees. *Electron. J. Combin.* **22** (2015), no. 1, Paper P1.31.
[72] Svante Janson. Asymptotic normality in Crump–Mode–Jagers processes: the discrete time case. In preparation.
[73] Svante Janson and Ralph Neininger, The size of random fragmentation trees. *Probab. Theory Related Fields* **142** (2008), nos 3–4, 399–442.
[74] Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz, *Univariate Discrete Distributions*. 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005. MR2163227
[75] Olav Kallenberg, *Foundations of Modern Probability*. 2nd ed., Springer, New York, 2002.
[76] Ravi Kalpathy and Hosam Mahmoud, Degree profile of m-ary search trees: A vehicle for data structure compression. *Probab. Engrg. Inform. Sci.* **30** (2016), no. 1, 113–123. MR3436783
[77] J. F. C. Kingman, The first birth problem for an age-dependent branching process. *Ann. Probability* **3** (1975), no. 5, 790–801. MR0400438
[78] Donald E. Knuth, *The Art of Computer Programming. Vol. 3: Sorting and Searching*. 2nd ed., Addison-Wesley, Reading, MA, 1998. MR3077154
[79] A. N. Kolmogoroff, Über das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung. *C. R. (Doklady) Acad. Sci. URSS (N. S.)* **31** (1941), 99–101.
[80] P. L. Krapivsky and S. Redner, Organization of growing random networks. *Phys. Rev. E* **63** (2001), 066123.
[81] P. L. Krapivsky, S. Redner, and F. Leyvraz, Connectivity of Growing Random Networks. *Phys. Rev. Lett.* **85** (2000), 4629–4632.
[82] Markus Kuba and Alois Panholzer, Isolating a leaf in rooted trees via random cuttings. *Ann. Comb.* **12** (2008), no. 1, 81–99.
[83] William Lew and Hosam M. Mahmoud, The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* **23** (1994), no. 5, 1050–1074.
[84] Jiang Lu and Qilin Feng, Strong consistency of the number of vertices of given degrees in nonuniform random recursive trees. *Yokohama Math. J.* **45** (1998), no. 1, 61–69.
[85] Hosam M. Mahmoud, *Evolution of Random Search Trees*. John Wiley & Sons, New York, 1992.

[86] Hosam M. Mahmoud, A strong law for the height of random binary pyramids. *Ann. Appl. Probab.* **4** (1994), no. 3, 923–932.
[87] Hosam M. Mahmoud and Boris Pittel, Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10** (1989), no. 1, 52–75.
[88] Hosam M. Mahmoud and R. T. Smythe, Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* **3** (1992), no. 3, 255–266.
[89] Hosam M. Mahmoud, R. T. Smythe and Jerzy Szymański, On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms* **4** (1993), no. 2, 151–176.
[90] Hosam M. Mahmoud and Mark Daniel Ward, Asymptotic distribution of two-protected nodes in random binary search trees. *Appl. Math. Lett.* **25** (2012), no. 12, 2218–2222.
[91] Hosam M. Mahmoud and Mark D. Ward, Asymptotic properties of protected nodes in random recursive trees. *J. Appl. Probab.* **52** (2015), no. 1, 290–297.
[92] T. F. Móri, On random trees. *Studia Sci. Math. Hungar.* **39** (2002), nos 1–2, 143–155.
[93] Richard Muntz and Robert Uzgalis, Dynamic storage allocation for binary search trees in a two-level memory. *Proceedings of the Princeton Conference on Information Sciences and Systems* **4** (1971), 345–349.
[94] Olle Nerman, On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 3, 365–395.
[95] Olle Nerman and Peter Jagers, The stable double infinite pedigree process of supercritical branching populations. *Z. Wahrsch. Verw. Gebiete* **65** (1984), no. 3, 445–460.
[96] Jacques Neveu, Arbres et processus de Galton–Watson. *Ann. Inst. H. Poincaré Probab. Statist.* **22** (1986), no. 2, 199–207.
[97] *NIST Handbook of Mathematical Functions*. Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010. Also available as *NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/
[98] Roberto Oliveira and Joel Spencer, Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.* **2** (2005), no. 2, 121–163.
[99] Anthony G. Pakes, Limit laws for UGROW random graphs. *Statist. Probab. Lett.* **83** (2013), no. 12, 2607–2614. MR3118202
[100] Alois Panholzer and Helmut Prodinger, Level of nodes in increasing trees revisited. *Random Structures Algorithms* **31** (2007), no. 2, 203–226.
[101] Boris Pittel, On growing random binary trees. *J. Math. Anal. Appl.* **103** (1984), no. 2, 461–480.
[102] Boris Pittel, Note on the heights of random recursive trees and random m-ary search trees, *Random Structures Algorithms* **5** (1994), 337–347.
[103] Bruce Reed, The height of a random binary search tree. *J. ACM* **50** (2003), no. 3, 306–332.

[104] Anna Rudas and Bálint Tóth, Random tree growth with branching processes – a survey. *Handbook of Large-Scale Random Networks*, 171–202, Bolyai Soc. Math. Stud. 18, Springer, Berlin, 2009.

[105] Anna Rudas, Bálint Tóth and Benedek Valkó, Random trees and general branching processes. *Random Structures Algorithms* **31** (2007), no. 2, 186–202. MR2343718

[106] Herbert A. Simon, On a class of skew distribution functions. *Biometrika* **42** (1955), 425–440.

[107] Jerzy Szymański, On a nonuniform random recursive tree. *Annals of Discrete Math.* **33** (1987), 297–306.

[108] G. Udny Yule, A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. Roy. Soc. B* **213** (1925), 21–87.