The Annals of Applied Probability
2010, Vol. 20, No. 3, 907–950
DOI: 10.1214/09-AAP640

A FUNCTIONAL LIMIT THEOREM FOR THE PROFILE OF b-ARY TREES

BY EVA-MARIA SCHOPP

University of Freiburg

In this paper we prove a functional limit theorem for the weighted profile of a b-ary tree. For the proof we use classical martingales connected to branching Markov processes and a generalized version of the profile-polynomial martingale. By embedding, choosing weights and a branch factor in a right way, we finally rediscover the profiles of some well-known discrete time trees.

1. Introduction. The profile is the set or sequence of numbers of nodes at each level of a tree. It is a fine tree shaped parameter related to many other important shape characteristics such as the total path length (the sum of the distances of all nodes to the root), depth (the distance of a random node to the root), height (the maximal distance of a node to the root), saturation level (the minimal distance of an external node to the root) and width (number of nodes at the most abundant level).

In general, we distinguish between two main groups of trees. In the first group we collect all the trees by their height which increases through the square root of the number of nodes. The prime examples of these kind of trees are Galton–Watson trees conditioned on the total progeny or simply generated trees. It is shown in Aldous (1993) that the simply generated trees studied by Meir and Moon (1978) and conditioned Galton–Watson trees are the same. Moreover, the profile of conditioned Galton–Watson trees is further investigated in Drmota and Gittenberger (1997), Kersting (1998) and Pitman (1999) with further references.

In the second group we collect all the trees with logarithmically growing height. In contrast to conditioned Galton–Watson trees, the profiles of these trees have received less attention in the past. However, there has been much done on this topic in the last few years. The methods used to derive limit theorems for normalized profiles range from the method of moments, the contraction method to analytical tools, including saddlepoint methods, Mellin transforms, Poissonization, de-Poissonization, singularity analysis, application of generating functions and uniform asymptotic analysis. We refer to Aldous and Shields (1988), Drmota and Hwang (2005a, 2005b), Fuchs, Hwang and Neininger (2006), Hwang (2005),

Received July 2008; revised August 2009.

AMS 2000 subject classifications. Primary 60F17; secondary 68Q25, 68P10.

Key words and phrases. Functional limit theorem, b-ary trees, profile of trees, random trees, analysis of algorithms, martingales.

Devroye and Hwang (2006), Drmota, Janson and Neininger (2008), Park (2006) and Park et al. (2009).

Furthermore, Chauvin, Drmota and Jabbour-Hattab (2001) and Chauvin et al. (2005) used martingale methods to obtain limit theorems for the profile of the random binary search tree. It is a classical result of Jabbour-Hattab (2001) that the profile-polynom of the binary search tree, $M_n(z) := \frac{1}{C_n(z)} \sum_l U_l(n) z^l$, is a discrete time martingale where $U_l(n)$ denotes the number of nodes in generation l of a random binary search tree of size n [cf. (11)] and $C_n(z)$ is a normalizing deterministic constant for $z \in \mathbb{C}$ fixed [see (14) for a definition]. With this in mind, classical martingale convergence results may now be applied to investigate the (asymptotic) behavior of this martingale and the asymptotics of the profile of the binary search tree [see Chauvin, Drmota and Jabbour-Hattab (2001)]. Instead of analyzing $M_n(z)$ directly, Chauvin et al. (2005) showed that the discrete time martingale $M_n(z)$ is deeply related to the well-studied classical Yule-time martingale $M(t,z)$ of the corresponding continuous time tree, the Yule tree [cf. (9)]. Finally, they strengthened classical convergence results in order to relate the uniform convergence of $M(t,z)$ in compact sets to corresponding uniform asymptotic results of $M_n(z)$.

Apart from the profile of the binary search tree, the profile of the random recursive tree has been studied with martingale methods [cf. Drmota and Hwang (2005a)]. Drmota and Hwang (2005b) showed that the profile is not concentrated around the mean. Their proof is based on showing that the variance undergoes four phase transitions and exhibits a bimodal behavior in contrast to the unimodality of the expected value of the profiles. As a consequence, the profile is not concentrated around the mean. For example, around the most numerous level (where the width is attained) the variance is small [$O(\log n)$] and the profile is concentrated with a Gaussian limit, but $\log n$ away from this level the variance increases, and there is no concentration anymore [cf. Drmota and Hwang (2005a, 2005b), and further remarks in Example 7.4]. Recently, Sulzbach (2008) used the martingale method of Chauvin et al. (2005) to show a limit theorem for the profile of plane oriented recursive trees.

Obviously not all trees have a martingale structure, and, furthermore, the method rapidly becomes costly when we leave the well-known binary search tree and recursive tree cases. The last statement is based on the fact that corresponding statements about other trees, and especially their continuous time matches, are not available which differs from assertions about the well-studied Yule tree process.

However, there is a large class of trees for which the martingale method gives uniform convergence results for the profile in any suitable compact set. In this paper we are interested in the kind of trees that can be studied with martingale methods. We want to find out how the method of Chauvin et al. (2005) can be generalized to derive asymptotic results for other trees besides the ones that have been studied so far.

Broutin and Devroye (2006) constructed a class of continuous time edge-weighted trees and proved, using large deviation techniques, a general law of large numbers for the height of these trees. This class of edge-weighted trees leads, when the tree process is stopped correctly, to various well-known discrete time trees as, for example, to random binary search trees and to random recursive trees. The initial point of their model is the construction of an infinite, b-ary tree where each node u in that tree is assigned, independently, a random vector,

$$((Z_1^{(u)}, E_1^{(u)}), (Z_2^{(u)}, E_2^{(u)}), \ldots, (Z_b^{(u)}, E_b^{(u)})),$$

where, for instance, $E_j^{(u)}$ is the lifetime of u's jth child, and $Z_j^{(u)}$ is a specific weight assigned to the jth child of u [cf. (2)]. Let $\mathrm{Exp}(\mu)$ denote the exponential distribution with the first moment equal to $\frac{1}{\mu}$, $\mu > 0$, in other words, the distribution with Lebesgue density $f(x) = \mu e^{-\mu x}\mathbb{1}_{[0,\infty)}(x)$ for $x \in \mathbb{R}$. A continuous time tree process is Markovian with a classical martingale structure if and only if the lifetimes are independent of each other and exponentially distributed (see, e.g., Harris [(1963), Chapter V.2], Watanabe (1967), Joffe, Le Cam and Neveu (1973), Athreya and Ney [(1972), Chapter III], Kingman (1975), Biggins (1977), Wang (1980), Uchiyama (1982), Neveu (19887), Biggins (1991) and Biggins (1992)). Essentially, this follows from the memoryless property of the exponential distribution.

Our purpose is now to show that these trees have the right martingale properties in order to generalize the method of Chauvin et al. (2005). The main theorem of this paper, Theorem 6.6, states that the normalized profile of a b-ary weighted tree converges almost surely to a limit as the number of nodes in the tree tends to infinity. This limit can be identified as the almost sure limit of a discrete time martingale and, also, as the unique solution of expectation one of a fixed point equation.

The plan for the rest of the paper is the following: First, in Section 2 we will introduce the tree model of Broutin and Devroye (2006) and define a closely related branching Markov process. Second, in Section 3 we will define the corresponding continuous time martingale associated with the branching Markov process of Section 2 and its discrete time analog in Section 4. Next, in Section 5 we will elaborate the relationship between those two martingales and formulate our main result, Theorem 6.6 in Section 6. Finally, in Section 7 we will show the applicability of our main Theorem 6.6 based on some examples as, for instance, the well-known random binary search tree, the random recursive tree, random lopsided trees and random plane oriented trees.

2. The framework. In this section we describe the tree model of Broutin and Devroye (2006). Let $\hat{T}_\infty$ be an infinite, complete b-ary tree with $b \geq 2$. We assign to each node a label,

$$u \in \mathcal{U} := \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \{1, \ldots, b\}^n, \tag{1}$$

in the natural way; the root node, which will be denoted by ø, has b children which are called $1, \dots, b$. In the same manner, every node u has children denoted by $u1, \dots, ub$. Generally, if $u = u_1 \cdots u_l$ is a node, and $v := v_1 \cdots v_k$ is a sequence with $v_j \in \{1, \dots, b\}$, $j = 1, \dots, k$, we set

$$uv = u_1 \cdots u_l v_1 \cdots v_k$$

and call u an ancestor of uv.

For each node u we create independently a random b-vector,

$$\left((Z_1^{(u)}, E_1^{(u)}), (Z_2^{(u)}, E_2^{(u)}), \dots, (Z_b^{(u)}, E_b^{(u)})\right), \tag{2}$$

where for $j = 1, \dots, b$, $(Z_j^{(u)}, E_j^{(u)})$ is the vector assigned to the edge from node u to its jth child. Here, $Z_j^{(u)}$ represents the weight of the edge from node u to its jth child, and $E_j^{(u)}$ is the lifetime of node u's jth child.

Each couple $(Z_j^{(u)}, E_j^{(u)})$ is distributed as (Z, E) for independent Z and E. Note that the lifetime $E_j^{(u)}$ is independent of $E_r^{(v)}$ for different u and v or different j and r. All E's are independent of any Z's, but we allow dependence of $Z_1^{(u)}, \dots, Z_b^{(u)}$ for any node u in the tree $\hat{T}_\infty$.

We also assume that E is exponentially distributed with mean one to ensure Markov properties and that Z is a lattice distribution with values in $\mathbb{Z}^d$ for some $d \in \mathbb{N}$.

REMARK 2.1. In Broutin and Devroye (2006) it is assumed that E and Z are nonnegative independent random variables where E is not mono-atomic, has no atom at zero and the following property holds:

$$\inf\{x : P(E > x) > 0\} = 0.$$

As mentioned in their concluding remark, their model and their proof of the law of large numbers for the height can be extended to more general cases allowing dependence of E and Z as well as multi-dimensional versions. In Broutin, Devroye and McLeish (2008) the assumption that the components of the random vectors attained to each node are independent is skipped in order to obtain further height results of, for example, pebbled trees and others.

Let $\pi(u)$ be the set of edges from the root to node u. Further, let (Z_e, E_e) be the couple of random variables assigned to edge e, and let

$$\hat{T}_t := \left\{u \in \hat{T}_\infty : G_u := \sum_{e \in \pi(u)} E_e \le t\right\} \tag{3}$$

be the subtree of $\hat{T}_\infty$ consisting of the nodes that deceased before time t. We are interested in the external profile of $\hat{T}_t$,

$$\hat{\rho}_t(l) := \left|\left\{u \in \partial\hat{T}_t : D_u := \sum_{e \in \pi(u)} Z_e = l\right\}\right| \tag{4}$$

for $t \in \mathbb{R}_+$ and $l \in \mathbb{Z}^d$, where

$$\partial \hat{T}_t := \{u \in \hat{T}_\infty : \text{ if } u = u_1 \cdots u_n \text{ for some } n, \\ \text{then } u_1 \cdots u_{n-1} \in \hat{T}_t, u \notin \hat{T}_t\}$$

is the set of all external nodes or leaves in $\hat{T}_t$.

For this purpose we will study a closely related branching Markov process (a jump or step Markov process) $(T_t)_{t\geq 0}$ which will be described next. We start the tree T_0 with one particle, the root ø, which is alive at time $t = 0$ and set accordingly $T_0 := \{ø\}$. This initial ancestor dies at a random time τ_1 where τ_1 is exponentially distributed with mean one and bears b children. These children behave independently from and similarly to their ancestor. After the first birth the individuals $1, \dots, b$ are alive with independent lifetimes equal to $(E_1^{(ø)}, \dots, E_b^{(ø)})$, respectively. At time $\tau_2 := \min\{E_1^{(ø)}, \dots, E_b^{(ø)}\} + \tau_1$ the corresponding individual deceases and gives birth to b new individuals, namely its children. Because of the memoryless property of the exponential distribution, all individuals alive just after τ_2 (namely the $b-1$ remaining children of the root and the b children of the root's child that deceased at time τ_2) behave similarly to and independently from each other, having an exponentially distributed lifetime. We define T_t as the tree corresponding to the process described above when it is stopped at time $t > 0$. It is clear that we have $\hat{T}_t = T_{t+\tau_1}$. In general, let $(\tau_j)_{0\leq j\leq\infty}$ be a sequence of Markov times with

$$\tau_0 = 0, \qquad \tau_j := \min\{t : N_t := (b-1)j + 1\}, \qquad N_t = |\partial T_t|, \tag{5}$$

where N_t is the number of external nodes or leaves in the tree T_t. Consequently, τ_j is the time of the jth death. By convention, we consider all internal nodes at time t deceased and the remaining external nodes alive. There are k internal nodes if and only if there are exactly $(b-1)k + 1$ external ones.

Since $\tau_1 \stackrel{d}{=} \mathrm{Exp}(1)$, we obtain

$$\tau_k - \tau_{k-1} \stackrel{d}{=} \min\{E_1, \dots, E_{(b-1)(k-1)+1}\} \stackrel{d}{=} \mathrm{Exp}\big((b-1)(k-1)+1\big),$$

where (E_j) are identically distributed, independent random variables, distributed as E and representing the remaining lifetimes of the $(b-1)(k-1)+1$ nodes alive at time τ_{k-1} after the $(k-1)$th death. For the distribution of τ_k we obtain then

$$\tau_k \stackrel{d}{=} \sum_{j=1}^{k} \frac{E_j}{(b-1)(j-1)+1}. \tag{6}$$

General notation. We denote for functions $f, g : \mathbb{N} \mapsto \mathbb{R}$ where $g(n) \neq 0$ $\forall n \in \mathbb{N}$,

$$f \sim_{\text{a.s.}} g \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{f(n)}{g(n)} = 1 \quad \Leftrightarrow \quad f(n) = g(n)\big(1 + o(1)\big)$$

with $o(1) \to 0, n \to \infty$. In the same manner we define

$$f \le_{\text{a.s.}} g \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{f(n)}{g(n)} \le 1 \quad \Leftrightarrow \quad f(n) \le g(n)\bigl(1+o(1)\bigr)$$

with $o(1) \to 0, n \to \infty$.

We let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

We summarize the results in the following lemma which is the result of Chauvin et al. [(2005), Lemma 2.1] in the Yule tree case.

LEMMA 2.2. (1) *We have* $(\tau_k - \tau_{k-1})_{k\ge 1}$ *independent and*

$$\tau_k - \tau_{k-1} \stackrel{d}{=} \operatorname{Exp}\bigl((b-1)(k-1)+1\bigr).$$

(2) *Further,* $(\tau_k)_{k\ge 1}$ *and* $(T_{\tau_k})_{k\ge 1}$ *are independent.*

(3) *We have* $(b-1)\tau_n \sim_{a.s.} \log n$ *almost surely.*

PROOF. The first two statements can be proven with the same arguments given in Chauvin et al. (2005). The last assertion is Proposition 1 in Broutin and Devroye (2006). □

Exactly in the same manner as for the tree $\hat{T}_\infty$, we assign to each node in the tree T_t a label $u \in \mathcal{U}$ in the natural way [cf. (1)]. Additionally, we weighted the edges in the tree $(\hat{T}_t)_{t\ge 0}$ as the values induced by $\hat{T}_\infty$. We define for $u = u_1 \cdots u_n \in \mathcal{U}$,

$$D_u := Z_{u_1}^{(\emptyset)} + \sum_{k=1}^{n-1} Z_{u_{k+1}}^{(u_1\cdots u_k)}$$

as its weight or weighted position and accordingly for $l \in \mathbb{Z}^d$ and $t \ge 0$,

$$\rho_t(l) := |\{u \in \partial T_t : D_u = l\}|.$$

Note that we obtain for $t \ge 0$, $\partial T_{t+\tau_1}$ is equal to $\partial \hat{T}_t$.

3. Continuous time martingales. Our ambition is the study of the profile of a class of discrete time trees $(\tilde{\mathcal{T}}_k)_k$ that can be constructed from the class of continuous time trees considered in Section 2 [cf. Section 7]. We will follow the ideas used by Chauvin et al. (2005) in the case of the random binary search tree and the Yule tree process and construct trees $(T_t)_{t\ge 0}$ described in the last section with $T_{\tau_k} \stackrel{d}{=} \tilde{\mathcal{T}}_k$. Alternatively, we deal with trees $\mathcal{T}_k := T_{\tau_k}$ whose profiles are at least comparable to the profile of $\tilde{\mathcal{T}}_k$. Recall that if we stop the tree process T_t at time $t = \tau_k$, we have k internal nodes and $(b-1)k+1$ external nodes or leaves in the tree.

These general trees are not as well studied as the binary search tree and its continuous time analog, the Yule tree process. Chauvin et al. (2005) used classical results from the Yule tree process and a corresponding fragmentation process [cf. Chauvin et al. (2005), Section 2.2] to formulate their main profile convergence result (see their Theorem 4.1). In general, it seems difficult to use a connection with a fragmentation processes.

Note that in the following we consider the natural filtrations $(\mathcal{F}_t)_{t\geq 0}$ and $(\mathcal{F}_{(k)})_{k\in\mathbb{N}_0}$ where

$$\mathcal{F}_t := \sigma(T_s, s \leq t) \quad \text{and} \quad \mathcal{F}_{(k)} := \sigma(T_{\tau_j}, 1 \leq j \leq k).$$

For $v = (v_1, \dots, v_d)$ and $w = (w_1, \dots, w_d) \in \mathbb{C}^d$ we define

$$v + w = (v_1 + w_1, \dots, v_d + w_d) \quad \text{and} \quad v \cdot w := \sum_{j=1}^{d} v_j w_j.$$

Set

$$\tilde{Z}^{(t)}(dx) := \sum_{u\in\partial T_t} \bar{\delta}_{\{D_u\}}(dx), \tag{7}$$

where $\bar{\delta}$ denotes the Dirac measure.

Let $m(\lambda)^t := E\int e^{-\lambda\cdot x}\tilde{Z}^{(t)}(dx)$ for $t \geq 0$. Then, for $|m(\lambda)| < \infty$ we define the classical martingale $W^{(t)}(\lambda)$ for $\lambda \in \mathbb{C}^d$, $t \geq 0$, as

$$W^{(t)}(\lambda) := \frac{1}{m(\lambda)^t}\int e^{-\lambda\cdot x}\tilde{Z}^{(t)}(dx). \tag{8}$$

Using Biggins [(1992), page 148], it follows for $t \geq 0$ that

$$m(\lambda)^t = \exp[t(bEe^{-\lambda\cdot Z} - 1)].$$

For the special case $d = 1$ we define, setting $z := e^{\lambda} \in \mathbb{C}$,

$$M(t, z) := W^{(t)}(-\lambda) = W^{(t)}(-\log z) = \sum_{u\in\partial T_t} z^{D_u} e^{-t(bEz^Z - 1)}. \tag{9}$$

This continuous time martingale was studied for $Z = 1$ and $b = 2$ (the Yule tree case) in Chauvin et al. (2005).

4. Discrete time martingales. Let $\mathcal{T}_n := T_{\tau_n}$ be the discrete time edge-weighted tree with n internal nodes. We define for $n \in \mathbb{N}$ and $\lambda \in \mathbb{C}^d \setminus \{\lambda : C_n(\lambda) = 0\}$,

$$W_n(\lambda) := \frac{1}{C_n(\lambda)} \sum_{l\in\mathbb{Z}^d} U_l(n) e^{-\lambda\cdot l} = \frac{1}{C_n(\lambda)} \sum_{u\in\partial\mathcal{T}_n} e^{-\lambda\cdot D_u}, \tag{10}$$

where $C_n(\lambda)$ is a multiplicative factor which we will specify below, and $U_l(n)$ is the number of external nodes in $\mathcal{T}_n$ at (weighted) level $l \in \mathbb{Z}^d$;

$$U_l(n) := \rho_{\tau_n}(l) = |\{u \in \partial T_{\tau_n} : D_u = l\}|, \qquad l \in \mathbb{Z}^d,\ n \in \mathbb{N}. \tag{11}$$

We choose $C_n(\lambda)$ in order to make $W_n(\lambda)$ a martingale with respect to the filtration $\mathcal{F}_{(n)} = \sigma\{\mathcal{T}_j, 1 \le j \le n\}$. In the case $d = 1, b = 2$ and $Z = 1$, the random binary tree case, the martingale

$$M_n(z) := W_n(-\log(z)), \tag{12}$$

where $z \in \mathbb{C}$ was found and $C_n(-\log z)$ calculated by Jabbour-Hattab (2001).

For the general case, let D_n be the weighted depth of the nth inserted (internal) node in $(\mathcal{T}_m)_{m\in\mathbb{N}_0}$ respectively $(T_t)_{t\in\mathbb{R}+}$. Because of the exponential distribution of the lifetimes, each alive individual (external node) is equally likely to be the next one to die and to become an internal node. Therefore we obtain for $l \in \mathbb{Z}^d$

$$P(D_{n+1} = l | T_{\tau_n}) = \frac{U_l(n)}{(b-1)n+1}.$$

Assume now that node $u \in T_{\tau_n}$ is the next node to expire and to bear b new individuals in the tree $T_{\tau_{n+1}}$. Denote by $Z_j^{(n+1)}$, $1 \le j \le b$, the d-dimensional weight assigned to the edge from node u to its jth child. Then we have

$$U_l(n+1) - U_l(n) = -\mathbb{1}_{\{D_{n+1}=l\}} + \sum_{j=1}^{b} \mathbb{1}_{\{D_{n+1}+Z_j^{(n+1)}=l\}}. \tag{13}$$

With $\tilde{W}_n(\lambda) := \sum_{l\in\mathbb{Z}^d} U_l(n) \exp(-\lambda \cdot l)$ and $\tilde{W}_0(\lambda) := 1$, we obtain from (13)

$$\begin{aligned} E\big(\tilde{W}_{n+1}(\lambda) | \mathcal{F}_{(n)}\big) &= \sum_l e^{-\lambda\cdot l} E\big(U_l(n+1) | \mathcal{F}_{(n)}\big) \\ &= \tilde{W}_n(\lambda) - \sum_l e^{-\lambda\cdot l} \frac{U_l(n)}{(b-1)n+1} \\ &\quad + \sum_{j=1}^{b} \sum_l e^{-\lambda\cdot l} P\big(D_{n+1} = l - Z_j^{(n+1)} | \mathcal{F}_{(n)}\big) \\ &= \tilde{W}_n(\lambda) \frac{(b-1)n + bEe^{-\lambda\cdot Z}}{(b-1)n+1}. \end{aligned}$$

Iterating this we have that

$$(W_n(\lambda))_{n\in\mathbb{N}} = \left(\frac{\tilde{W}_n(\lambda)}{\prod_{j=0}^{n-1}((b-1)j + bEe^{-\lambda\cdot Z})/((b-1)j+1)} \right)_{n\in\mathbb{N}}$$

is a $(\mathcal{F}_{(n)})_{n\in\mathbb{N}}$ adapted martingale. Consequently, we set it according to our notation for $n \geq 1$

$$C_n(\lambda) := \prod_{j=0}^{n-1} \frac{(b-1)j + bE\exp(-\lambda\cdot Z)}{(b-1)j+1}, \tag{14}$$

and $C_0(\lambda) := 1$. We define for further calculations and, for reference, the following set:

$$\begin{aligned} N_C &:= \{\lambda \in \mathbb{C}^d : C_n(\lambda) = 0 \text{ for some } n\} \\ &= \left\{\lambda : -\frac{b}{b-1} E\exp(-\lambda\cdot Z) \in \mathbb{N}_0\right\}. \end{aligned} \tag{15}$$

5. Relationships between discrete and continuous time martingales. In this section we will study the relationship of the two martingales $(W_n(\lambda))_{n\in\mathbb{N}_0}$ and $(W^{(t)}(\lambda))_{t\geq 0}$ defined in Sections 3 and 4. With $C_n(\lambda)$, $n \in \mathbb{N}_0$, defined in (14), we set for $\lambda \in \mathbb{C}^d \setminus N_C$ [cf. (15)],

$$\mathcal{H}_n(\lambda) := C_n(\lambda) e^{\tau_n(1-bEe^{-\lambda\cdot Z})}, \qquad n \geq 0.$$

We claim that $(\mathcal{H}_n(\lambda))_{n\in\mathbb{N}_0}$ is a martingale with respect to the filtration $(\mathcal{F}_{\tau_n})_{n\in\mathbb{N}_0}$ with expectation 1.

LEMMA 5.1. *Let $\lambda \in \mathbb{C}^d \setminus N_C$. Then*:

(1) $W^{(\tau_n)}(\lambda) = \mathcal{H}_n(\lambda) W_n(\lambda)$, $n \in \mathbb{N}_0$.

(2) $(\mathcal{H}_n(\lambda))_{n\in\mathbb{N}_0}$ *is a martingale with respect to the filtration* $(\mathcal{F}_{\tau_n})_{n\in\mathbb{N}_0}$ *with expectation* 1.

(3) $(\mathcal{H}_n(\lambda))_{n\in\mathbb{N}_0}$ *and* $(W_n(\lambda))_{n\in\mathbb{N}_0}$ *are independent*.

PROOF. The first statement is a direct consequence of the definition. The other two statements follow from an application of Lemma 2.2. □

5.1. *Asymptotic behavior and further relationships*. In this subsection we are interested in the asymptotic behavior of $C_n(\theta)$ for $\theta \in \mathbb{R}^d$.

LEMMA 5.2. *For $\theta \in \mathbb{R}^d$, $b > 1$, and $n \to \infty$ we have*

$$C_n(\theta) \sim_{a.s.} n^{1/(b-1)(bEe^{-\theta\cdot Z}-1)} \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta\cdot Z})/(b-1))}.$$

PROOF. Let $\alpha \neq 0$ and $\beta \in \mathbb{R}$. From Stirling's formula for the Gamma function [see, e.g., Flajolet and Odlyzko (1990)], we obtain

$$\begin{aligned} \prod_{j=0}^{n-1} \frac{\alpha j + \beta}{\alpha j + 1} &= \binom{n-(-\beta/\alpha)-1}{n} \left[\binom{n-(-1/\alpha)-1}{n}\right]^{-1} \\ &= n^{(1/\alpha)(\beta-1)} \frac{\Gamma(1/\alpha)}{\Gamma(\beta/\alpha)} \left(1 + O\left(\frac{1}{n}\right)\right). \end{aligned}$$

With $\alpha = (b-1)$ and $\beta = bEe^{-\theta\cdot Z}$ we obtain the statement. $\square$

Since $\mathcal{H}_n(\lambda)$ is a positive martingale for $Ee^{-\lambda\cdot Z} > 0$, we obtain immediately from a well-known, classical martingale result, for every $\lambda \in \mathbb{R}^d$, $\mathcal{H}_n(\lambda)$ converges almost surely to a limit $\mathcal{H}(\lambda)$ as $n \to \infty$. More details are given in the next lemma.

LEMMA 5.3. *Let $\lambda \in \mathbb{R}^d \setminus N_C$ with N_C defined in* (15). *Then we have*, *almost surely*,

$$\mathcal{H}_n(\lambda) \to \left(\frac{Y}{b-1}\right)^{(bEe^{-\lambda\cdot Z}-1)/(b-1)} \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\lambda\cdot Z})/(b-1))}, \qquad n \to \infty,$$

where $Y \stackrel{d}{=} \Gamma(\frac{1}{b-1}, \frac{1}{b-1})$.

PROOF. First we show that $e^{-(b-1)t}N_t$, defined in (5), converges almost surely and that the limit law is gamma distributed. From Athreya and Ney [(1972), Remark 1, page 109] we detect that

$$Es^{N_t} = se^{-t}\bigl[1 - \bigl(1 - e^{-(b-1)t}\bigr)s^{b-1}\bigr]^{-1/(b-1)}.$$

As a consequence of setting $\alpha := b-1 > 0$ and $u = -is$ for $s \in \mathbb{R}$, we obtain by standard calculations

$$Ee^{isN_t e^{-\alpha t}} \to \left(\frac{1}{\alpha is + 1}\right)^{1/\alpha}, \qquad t \to \infty. \tag{16}$$

The right-hand side of (16) is the characteristic function of a $\Gamma(1/\alpha, 1/\alpha)$ distributed random variable with density

$$f_\Gamma(x) := \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} x^{(1/\alpha)-1} e^{-x/\alpha} \mathbb{1}_{(0,\infty)}(x).$$

With Doob's limit law it is immediately verified that the continuous parameter nonnegative martingale $(N_t e^{-(b-1)t})_{t\geq 0}$ converges with probability one to a finite limit Y which is gamma distributed with parameters $(\frac{1}{\alpha}, \frac{1}{\alpha})$:

$$N_t e^{-(b-1)t} \to Y, \qquad t \to \infty, \text{a.s.} \tag{17}$$

Consequently, we have for $n \to \infty$,

$$(b-1)ne^{-(b-1)\tau_n} \to Y \qquad \text{a.s.}, \tag{18}$$

since τ_n converges toward infinity almost surely. Therefore, we obtain almost surely with Lemma 5.2 that for $n \to \infty$ and $\lambda \in \mathbb{R}^d \setminus N_C$,

$$\begin{aligned}\mathcal{H}_n(\lambda) &\sim_{\text{a.s.}} \bigl(n^{1/(b-1)}e^{-\tau_n}\bigr)^{(bEe^{-\lambda\cdot Z}-1)} \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\lambda\cdot Z})/(b-1))}\\ &\to \left(\frac{Y}{b-1}\right)^{(bEe^{-\lambda\cdot Z}-1)/(b-1)} \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\lambda\cdot Z})/(b-1))}.\end{aligned}$$

$\square$

Note that for $\tilde{Z}^{(t)}$ defined in (7), we have

$$\tilde{Z}^{(t)}(dx) = \sum_{r=1}^{N_t} \mathbb{1}_{\{z_r^{(t)}\}}(dx),$$

where $z_r^{(t)}$ is the weighted position of the rth individual alive at time t. If the rth individual is equal to node u, then $z_r^{(t)}$ is equal to $D_u \in \mathbb{Z}^d$. Note that the positions of the individuals alive just before τ_n—the time of the nth death—are $\{z_r^{(\tau_{n-1})} : 1 \le r \le (b-1)(n-1)+1\}$ as the particles do not move during their lifetimes.

Before we state the next theorem, we formulate the following preliminary proposition. The following set will play an essential role in the sequel (cf. Theorem 5.5). For $1 < \gamma \le 2$, $\lambda = \theta + i\eta$, define $\Omega_\gamma^1, \Omega_\gamma^2 \subset \mathbb{C}^d$ by

$$\Omega_\gamma^1 := \operatorname{int}\{\lambda : bEe^{-\gamma\theta\cdot Z} < \infty\},$$

$$\Omega_\gamma^2 := \operatorname{int}\left\{\lambda : \frac{m(\gamma\theta)}{|m(\lambda)|^\gamma} < 1\right\}.$$

We set

$$\Lambda := \bigcup_{1<\gamma\le 2} \Omega_\gamma^1 \cap \Omega_\gamma^2. \tag{19}$$

PROPOSITION 5.4. *For $\theta \in \Lambda$, $\delta > 0$, $n \ge 1$, $\alpha \in \mathbb{R}$ we have*:

(1) $EW^{(\tau_n)}(\theta) = (Em(\theta)^{-\tau_n})^{-1}$;
(2) $Ee^{-\theta z_1^{(\tau_n)}} = \frac{1}{Em(\theta)^{-\tau_n}((b-1)n+1)}$;
(3) $E(\delta^\alpha)^{\tau_n} \sim_{a.s.} \frac{\Gamma((\log(1/\delta)^\alpha+1)/(b-1))}{\Gamma(1/(b-1))} n^{-\log(1/\delta)^\alpha/(b-1)}$.

PROOF. For the proof of the first statement we let $n \ge 1$, setting $a_n = (b-1)n+1$,

$$\xi_n := \xi_n(\theta) := EW^{(\tau_n)}(\theta) = E\int e^{-\theta\cdot x}\tilde{Z}^{(\tau_n)}(dx) = a_n \underbrace{Ee^{-\theta\cdot z_1^{(\tau_n)}}}_{=:c_n}.$$

In particular, we designate $\xi_1 = \log m(\theta) + 1$. Using induction over n, we find

$$\begin{aligned}\xi_n &= E\left[\sum_{r=2}^{a_{n-1}} e^{-\theta\cdot z_r^{(\tau_{n-1})}} + \sum_{j=1}^{b} e^{-\theta\cdot(z_r^{(\tau_{n-1})}+Z_j)}\right] \\ &= \prod_{j=1}^{n} \frac{(b-1)(j-1)+1+\log m(\theta)}{(b-1)(j-1)+1}.\end{aligned}$$

Using (6) we obtain

$$
\begin{aligned}
E(\delta^\alpha)^{\tau_n} &= E\exp\left(\sum_{j=1}^{n}\frac{E_j}{(b-1)(j-1)+1}\log\delta^\alpha\right)\\
&= \prod_{j=1}^{n}\frac{(b-1)(j-1)+1}{(b-1)(j-1)+1-\log\delta^\alpha}.
\end{aligned}
\tag{20}
$$

From this we can finally conclude that $\xi_n = \xi_n(\theta) = (Em(\theta)^{-\tau_n})^{-1}$.

For the second statement we obtain immediately from (1) that

$$Ee^{-\theta\cdot z_1^{(\tau_n)}} = c_n = \frac{1}{a_n}\xi_n = \frac{1}{a_n}(Em(\theta)^{-\tau_n})^{-1}.$$

Finally for the last assertion, using the same asymptotic result as in Lemma 5.2, we obtain

$$\prod_{j=1}^{n}\frac{aj+1}{aj+1+x} \sim_{\text{a.s.}} (x+1)\frac{\Gamma((x+1)/a)}{\Gamma(1/a)}(n+1)^{-x/a},$$

and further, using (20),

$$E(\delta^\alpha)^{\tau_n} \sim_{\text{a.s.}} \frac{\Gamma((\log(1/\delta)^\alpha+1)/(b-1))}{\Gamma(1/(b-1))}n^{-\log(1/\delta)^\alpha/(b-1)}. \qquad \square$$

In the following theorem we describe the convergence behavior of the discrete time martingale of Section 4 [see (10)] and of the continuous time martingale of Section 3 [cf. (8)]. The first part is based on Biggins (1992), Theorem 6. The proof of the second part is based on the application of the first part of Theorem 5.5 and the relationship between the two martingales (cf. Lemma 5.1).

For a set $A \subset \mathbb{C}^d$ let $\operatorname{int}(A)$ denote the set of interior points of A.

THEOREM 5.5. *With Λ defined in (19) we have:*

(1) *as $t\to\infty$, $\{W^{(t)}(\lambda)\}$ converges, a.s. and in L^1, uniformly on every compact subset C of Λ;*

(2) *as $n\to\infty$, $\{W_n(\lambda)\}$ converges, a.s. and in L^1, uniformly on every compact subset C of Λ.*

The limits are denoted as $W^{(\infty)}(\lambda)$, respectively, $W_\infty(\lambda)$.

PROOF. The first part was proven by Biggins [(1992), Theorem 6], so we only need to prove the second part.

Let $C \subset \Lambda$ be a compact subset of Λ. Therefore,

$$\lim_N \sup_{n\ge N} E \sup_{\lambda\in C}|W_n(\lambda) - W_N(\lambda)| = 0$$

which implies the uniform L^1 convergence. Additionally, due to the fact that $(\sup_{\lambda\in C}|W_n(\lambda)-W_N(\lambda)|)_{n\geq N}$ is a submartingale, this implies also the a.s. uniform convergence. From Lemma 5.1 we have

$$W_n(\lambda)-W_N(\lambda)=E\big(W^{(\tau_n)}(\lambda)-W^{(\tau_N)}(\lambda)|\mathcal{F}_{(n)}\big).$$

Now taking the supremum and expectations we further deduce

$$E\sup_{\lambda\in C}|W_n(\lambda)-W_N(\lambda)|\leq E\Big(\sup_{\lambda\in C}\big|W^{(\tau_n)}(\lambda)-W^{(\tau_N)}(\lambda)\big|\Big).$$

Taking the supremum over $n\geq N$ we get

$$\begin{aligned}\sup_{n\geq N}E\sup_{\lambda\in C}|W_n(\lambda)-W_N(\lambda)|&\leq E\sup_{n\geq N}\Big(\sup_{\lambda\in C}\big|W^{(\tau_n)}(\lambda)-W^{(\tau_N)}(\lambda)\big|\Big)\\&\leq E\Delta_N,\end{aligned}$$

where we set

$$\Delta_n:=\sup_{T\geq\tau_n}\Big(\sup_{\lambda\in C}\big|W^{(T)}(\lambda)-W^{(\tau_n)}(\lambda)\big|\Big).$$

Since $W^{(t)}(\lambda)$ converges a.s. uniformly, we have a.s. that $\lim_{n\to\infty}\Delta_n=0$. With the triangle inequality we obtain that $\Delta_n\leq 2\Delta_0$. If we show that Δ_0 is integrable, we deduce $\lim_n E\Delta_n=0$ and the statement of the theorem using dominated convergence.

Let $M_\lambda(s):=\{s>0: W^{(s)}(\lambda)\neq W^{(s-)}(\lambda)\}$ for $\lambda\in\Lambda$ and $s>0$. As it is shown in Bertoin and Rouault [(2003), proof of Proposition 3] it is sufficient for proving the integrability of Δ_0 to show that for all $x\in\Lambda$ a disk,

$$D_x(\bar\rho):=\{\lambda\in\mathbb{C}^d:\|\lambda-x\|<\bar\rho\},$$

exists with

$$\sup_{\lambda\in D_x(\bar\rho)}E\Big(\sum_{s\in M_\lambda(s)}\big|W^{(s)}(\lambda)-W^{(s-)}(\lambda)\big|^q\Big)<\infty$$

for some $q\in(1,2]$. Note that the set $M_\lambda(x)$ is a.s. countable. We observe that $W^{(t)}(\lambda)\neq W^{(t-)}(\lambda)$ if and only if $t=\tau_n$ for some $n\in\mathbb{N}$, and for that reason we can conclude that

$$\begin{aligned}&E\Big(\sum_{s\in M_\lambda(s)}\big|W^{(s)}(\lambda)-W^{(s-)}(\lambda)\big|^q\Big)\\&\qquad=E\sum_{n=1}^{\infty}\big|W^{(\tau_n)}(\lambda)-W^{(\tau_n-)}(\lambda)\big|^q\\&\qquad=\sum_{n=1}^{\infty}E\bigg|m(\lambda)^{-\tau_n}\bigg[\int e^{-\lambda\cdot x}\,d\tilde Z^{(\tau_n)}(x)-\int e^{-\lambda\cdot x}\,d\tilde Z^{(\tau_{n-1})}(x)\bigg]\bigg|^q.\end{aligned}$$

Now set $a_{n-1} := (b-1)(n-1)+1$ for $n \in \mathbb{N}$. Then we derive the following formula for $\tilde{Z}^{(\tau_n)}$:

$$\tilde{Z}^{(\tau_n)} \stackrel{d}{=} \sum_{r=2}^{a_{n-1}} \mathbb{1}_{\{z_r^{(\tau_{n-1})}\}} + \sum_{j=1}^{b} \mathbb{1}_{\{z_1^{(\tau_{n-1})}+Z_j\}}, \tag{21}$$

where $(Z_1, \dots, Z_b)$ are independent of $\{z_r^{(\tau_{n-1})}, 1 \le r \le a_{n-1}\}$ and distributed as $(Z_1^{(\emptyset)}, \dots, Z_b^{(\emptyset)})$ [cf. (2)].

Then, with Lemma 2.2, Jensen's inequality, independence and Proposition 5.4, we have for $q \in (1,2]$, $F_{\tau,n} := \sigma(\tau_j, 1 \le j \le n)$ and $n \in \mathbb{N}_0$, that

$$\begin{aligned} &E\left|W^{(\tau_n)}(\lambda) - W^{(\tau_n-)}(\lambda)\right|^q \\ &\quad = E\left(|m(\lambda)|^{-q\tau_n} E\left[\left|e^{-\lambda\cdot z_1^{(\tau_{n-1})}}\left(\sum_{j=1}^{b} e^{-\lambda\cdot Z_j} - 1\right)\right|^q \;\Big|\; F_{\tau,n}\right]\right) \\ &\quad \le E|m(\lambda)|^{-q\tau_n} E\left|e^{-\lambda z_1^{(\tau_{n-1})}}\right|^q 2^{q-1}\left(E\left(\sum_{j=1}^{b} |e^{-\lambda\cdot Z_j}|\right)^q + 1\right) \\ &\quad = \frac{2^{q-1}}{a_{n-1}} \frac{E|m(\lambda)|^{-q\tau_n}}{Em(\theta q)^{-\tau_{n-1}}}\left(E\left(\sum_{j=1}^{b} e^{-\theta\cdot Z_j}\right)^q + 1\right). \end{aligned} \tag{22}$$

Recall the definition of Λ as

$$\Lambda = \bigcup_{1<\gamma\le 2} (\Omega_\gamma^1 \cap \Omega_\gamma^2).$$

For $x \in \Omega_\gamma^1 \cap \Omega_\gamma^2$ for some $\gamma \in (1,2]$, we can choose $\bar{\rho}$ sufficiently small so that $D_{\bar{\rho}}(x) \subset \Omega_\gamma^1 \cap \Omega_\gamma^2$. Then there exist some $\delta < 1$ so that for all $\lambda \in D_{\bar{\rho}}(x)$ we have $\frac{m(\theta\gamma)^{1/\gamma}}{|m(\lambda)|} \le \delta$ for some $\gamma \in (1,2]$. Using this we obtain with Proposition 5.4,

$$E|m(\lambda)|^{-\gamma\tau_n}(Em(\theta\gamma)^{-\tau_{n-1}})^{-1} \le_{\text{a.s.}} C(\gamma,\lambda,b) n^{\log\delta^\gamma/(b-1)}.$$

With (22) by setting $q = \gamma \in (1,2]$ we obtain

$$\begin{aligned} &E\left|W^{(\tau_n)}(\lambda) - W^{(\tau_n-)}(\lambda)\right|^\gamma \\ &\quad \le_{\text{a.s.}} 2^{\gamma-1}\frac{1}{a_{n-1}}\left(E\left(\sum_{j=1}^{b} e^{-\theta\cdot Z_j}\right)^\gamma + 1\right) C(\gamma,\lambda,b) n^{\log\delta^\gamma/(b-1)} \\ &\quad \sim_{\text{a.s.}} \tilde{C}(\gamma,\lambda,b) n^{-(1+\varepsilon)} \end{aligned}$$

for a suitable constant $\tilde{C}(\gamma,\lambda,b)$ independent of n, and $\varepsilon := \frac{\log(1/\delta)^\gamma}{b-1}$. Since $\sum_n n^{-1-\varepsilon} < \infty$ and $\sup_{\lambda\in D_{\bar{\rho}}(x)} \tilde{C}(\gamma,\lambda,b) < \infty$ for $\bar{\rho}$ sufficiently small, we obtain the statement. $\square$

REMARK 5.6. The proof of the second part of Theorem 5.5 is roughly the same as the corresponding proof of Chauvin et al. [(2005), Theorem 3.1] up to the point where the integrability of Δ_0 is proven. Their arguments can also be used in our more general case except of the one concerning the integrability of $\Delta_0 = \sup_{T\geq 0}\sup_{\lambda\in C}|W^{(t)}(\lambda) - 1|$. Since the Yule tree process is a special kind of a fragmentation process, the integrability of Δ_0 in Chauvin et al. (2005) can be obtained from a result of Bertoin and Rouault (2005). To be more precise for fragmentation processes, the integrability can be verified by an application of the compensation formula for Poisson point processes applied to the Poissonian construction of the fragmentation (see Bertoin and Rouault [(2003), Proposition 3] and Bertoin [(2003), Theorem 2] for more details). Is seems difficult to use a similar argument in the general case.

REMARK 5.7. Recall that we have already defined

$$\Lambda = \bigcup_{1<\gamma\leq 2} (\Omega_\gamma^1 \cap \Omega_\gamma^2)$$

with

$$\Omega_\gamma^1 = \mathrm{int}\{\lambda : bEe^{-\gamma\theta\cdot Z} < \infty\} \quad \text{and}$$

$$\Omega_\gamma^2 = \mathrm{int}\left\{\lambda : \frac{m(\gamma\theta)}{|m(\lambda)|^\gamma} < 1\right\}$$

in (19). Using Biggins [(1992), Theorem 6], we have that $\{W^{(t)}(\lambda)\}$ converges uniformly on any compact subset of Λ, almost surely and in mean, as $t\to\infty$ [cf. Theorem 5.5].

Now define

$$\tilde{\Lambda} := \Lambda \cap \mathbb{R}^d$$

as the restriction of Λ on $\mathbb{R}^d$. Then we can rewrite $\tilde{\Lambda}$ and show that

$$\tilde{\Lambda} = \bigcup_{1<\gamma\leq 2} \Omega_\gamma^1 \cap \tilde{\Omega}^3 \tag{23}$$

with

$$\tilde{\Omega}^3 := \left\{\theta \in \mathbb{R}^d : \theta \in \Omega_0 : -\log m(\theta) < \frac{-\theta\cdot m'(\theta)}{m(\theta)}\right\} \quad \text{and}$$

$$\Omega_0 := \mathrm{int}\{\lambda \in \mathbb{C}^d : m(\mathrm{Re}(\lambda)) < \infty\}$$

(cf. Biggins [(1992), page 141]). Since

$$m(\theta) = \exp(bEe^{-\theta\cdot Z} - 1) \quad \text{and}$$

$$m'(\theta) = m(\theta)(-bEZe^{-\theta\cdot Z}),$$

we have for any $\theta \in \Omega_0 \cap \mathbb{R}^d$,

$$\theta \in \tilde{\Omega}^3 \quad \Leftrightarrow \quad 1 - bEe^{-\theta\cdot Z} < b\theta\cdot EZe^{-\theta\cdot Z}.$$

5.1.1. *Subtree sizes*. The rest of Section 5 is devoted to a further characterization of $W^{(\infty)}(\cdot)$, resp. $W_\infty(\cdot)$, defined in Theorem 5.5 as solutions of fixed point equations. For deriving these fixed point equations we will split the original tree in the b subtrees which are growing from the children of the root. For that purpose we will investigate the sizes of the subtrees which are growing from a node in the tree. For every $u \in \mathcal{U}$ let

$$\tau^{(u)} := \inf\{t : u \in T_t\}$$

be the time of the first appearance (birth) of node u in the tree. For $t > 0$ set

$$T_t^{(u)} := \{v \in \mathcal{U} : uv \in T_{t+\tau^{(u)}}\},$$

the tree process growing from u. Further set

$$N_t^{(u)} = |\partial T_t^{(u)}| \quad \text{and} \quad n_t^{(u)} := N_{t-\tau^{(u)}}^{(u)},$$

the number of leaves at time $t \geq \tau^{(u)}$ in the subtree growing from node u. Then, using the same arguments as in (17), we obtain

$$\lim_{t\to\infty} e^{-(b-1)t} N_t^{(u)} = Y_u \quad \text{and} \tag{24}$$

$$\lim_{t\to\infty} e^{-t(b-1)} n_t^{(u)} = Y_u e^{-\tau^{(u)}(b-1)}, \tag{25}$$

where Y_u is distributed as Y; that is, it is $\Gamma(\frac{1}{b-1}, \frac{1}{b-1})$ distributed. If u, v are not in the same line of descent, we can conclude that by the branching property Y_u and Y_v are independent. Since for $t \geq \tau^{(u)}$ we have

$$n_t^{(u)} = n_t^{(u1)} + n_t^{(u2)} + \cdots + n_t^{(ub)}, \qquad \tau^{(u1)} = \tau^{(u2)} = \cdots = \tau^{(ub)},$$

we obtain using (25)

$$e^{-\tau^{(u1)}(b-1)} \sum_{j=1}^{b} Y_{uj} = e^{-\tau^{(u)}(b-1)} Y_u,$$

and for that reason we further get

$$\frac{n_t^{(uj)}}{n_t^{(u)}} \to \frac{Y_{uj}}{\sum_{j=1}^{b} Y_{uj}} =: U^{(uj)}, \qquad 1 \leq j \leq b. \tag{26}$$

Finally, with $\tau_1 = \tau^{(1)}$, we have

$$Y := Y_\varnothing = e^{-\tau_1(b-1)}(Y_1 + \cdots + Y_b), \tag{27}$$

$$Y_j = U^{(j)} Y e^{\tau_1(b-1)}, \qquad 1 \leq j \leq b \quad \text{and} \tag{28}$$

$$1 = \sum_{j=1}^{b} U^{(j)}. \tag{29}$$

The distribution of the subtree sizes and their limit distributions can now be further calculated using a generalized Pólya–Eggenberger urn model.

5.1.2. *Limit martingale equation and splitting formulas*. In the following statement we derive representations of $W^{(\infty)}(\cdot)$ and $W_\infty(\cdot)$ as solutions of fixed point equations. Furthermore, the first part of Theorem 5.8 emphasizes the close relationship of $W^{(\infty)}(\cdot)$ and $W_\infty(\cdot)$.

THEOREM 5.8. *Let us assume $\lambda \in \tilde{\Lambda}$ [cf. (23)]. Then the following formulas hold*:

(1) *limit martingale connection*,

$$W^{(\infty)}(\lambda) = \left(\frac{Y}{b-1}\right)^{(bEe^{-\lambda\cdot Z}-1)/(b-1)} \times \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\lambda\cdot Z})/(b-1))} W_\infty(\lambda) \qquad a.s., \tag{30}$$

with $Y \stackrel{d}{=} \Gamma(\frac{1}{b-1}, \frac{1}{b-1})$;

(2) *splitting formula*,

(a) *for the continuous time process*

$$W^{(\infty)}(\lambda) = \sum_{j=1}^{b} e^{-\lambda\cdot Z_j} e^{-\tau_1(bEe^{-\lambda\cdot Z}-1)} W_j^{(\infty)}(\lambda), \tag{31}$$

where $W_1^{(\infty)}(\lambda), \dots, W_b^{(\infty)}(\lambda)$ are independent, distributed as $W^{(\infty)}(\lambda)$ and independent of τ_1 and $(Z_1, \dots, Z_b)$ where $(Z_1, \dots, Z_b)$ are the weights assigned to the edge from node $\varnothing$ to its children $1, \dots, b \in \mathcal{U}$,

(b) *for the discrete time process*,

$$W_\infty(\lambda) = \sum_{j=1}^{b} e^{-\lambda\cdot Z_j} \big(U^{(j)}\big)^{(bEe^{-\lambda\cdot Z}-1)/(b-1)} W_{\infty,(j)}(\lambda), \tag{32}$$

where $W_{\infty,(1)}(\lambda), \dots, W_{\infty,(b)}(\lambda)$ are independent, distributed as $W_\infty(\lambda)$ and independent of $(U^{(j)})$ where $(U^{(j)})$ are defined in (26).

PROOF. (1) This is a consequence of Lemmas 5.1 and 5.3.

(2) (a) For $t > \tau_1$ we have

$$W^{(t)}(\lambda) = \sum_{j=1}^{b} e^{-\lambda\cdot Z_j} e^{-\tau_1(bEe^{-\lambda\cdot Z}-1)} W_j^{(t-\tau_1)}(\lambda),$$

where for $j = 1, \dots, b$ we set

$$W_j^{(t)}(\lambda) := \sum_{u \in \partial T_t^{(j)}} e^{-\lambda\cdot D_u} e^{-t(bEe^{-\lambda\cdot Z}-1)},$$

and we let $(Z_1, \dots, Z_b)$ be the weights assigned to the edges from $\varnothing$ to its children $1, \dots, b$. Now let $t \to \infty$. Then the assertion follows.

(b) Follows now from (1), (2)(a) and (28). □

6. Profiles and large deviation results. To prove our main theorem, Theorem 6.6, we will need some preliminary lemmas that will be stated at the beginning of this section. The first one, Lemma 6.2, is based on Biggins [(1992), Lemma 5] in the discrete time case with nonlattice weights Z. The proof for the continuous time nonlattice version can be managed with some additional arguments [see Biggins (1992), page 150]. For the lattice case note that the critical points in the proof of Lemma 6.2 are the values $\eta \in \mathbb{R}^d$ with

$$\left|\frac{m(\lambda)}{m(\theta)}\right| = \exp\bigl(bEe^{-\theta\cdot Z}(\cos(\eta\cdot Z)-1)\bigr) = 1, \qquad \theta \in \Omega_0.$$

This can only occur if Z is a lattice distribution or if $\eta = 0$. Let

$$N := \{l \in \mathbb{Z}^d : P(\{Z = l\}) > 0\} \tag{33}$$

denote the support of Z. For $\eta = (\eta_1, \dots, \eta_d) \in \mathbb{R}^d$ we denote by

$$|\eta| := \max\{|\eta_j|, j = 1, \dots, d\}$$

the maximum-norm. Denote for $v_1, \dots, v_r \in \mathbb{R}^d$ by

$$\mathcal{E}\{v_1, \dots, v_r\} := \left\{v : \exists \lambda = (\lambda_1, \dots, \lambda_r) \in \mathbb{R}^r : v = \sum_{j=1}^r \lambda_j v_j\right\} \subset \mathbb{R}^d$$

the subspace generated from the vectors $v_1, \dots, v_r$ and, analogously, let for some subset $S \subset \mathbb{R}^d$, $\mathcal{E}(S)$ be the subspace in $\mathbb{R}^d$ generated from the vectors in the set S. The next proposition is obvious and will be stated for further reference in the paper.

PROPOSITION 6.1. *Let Z be a lattice distribution with support N; define*

$$f_\theta(\eta) := bEe^{-\theta\cdot Z}(\cos(\eta\cdot Z)-1), \qquad \theta \in \Omega_0, \eta \in \mathbb{R}^d$$

and let $a \in \mathbb{R}^1_+$.

(1) *If the dimension of $\mathcal{E}(N)$ is equal to d, then there are only finitely many roots $\eta_0 = 0, \eta_1, \dots, \eta_m$, of $f_\theta(\cdot)$ in $[-a, a]^d$ independently for all $\theta \in \Omega_0$ where $m = m(a) \in \mathbb{N}_0$.*

(2) *If the dimension of $\mathcal{E}(N)$ is $r < d$, then there are $\eta_{0,1}, \dots, \eta_{0,d-r} \in \mathbb{R}^d$ (linearly) independent vectors, $\eta_0 = 0, \eta_1, \dots, \eta_m \in [-a, a]^d$ where $m := m(a) \in \mathbb{N}_0$ so that*

$$L := \{\eta \in [-a, a]^d : f_\theta(\eta) = 0\} \subset \bigcup_{j=0}^m C_j,$$

where

$$C_j := \eta_j + \mathcal{E}\{\eta_{0,1}, \dots, \eta_{0,d-r}\}, \qquad 0 \le j \le m.$$

LEMMA 6.2. *For every lattice-distribution Z we have almost surely*

$$\lim_{t\to\infty} \sup_{\theta\in\tilde{K}} \int_{|\eta|\le\pi} \sqrt{t}^d \big|W^{(t)}(\theta+i\eta) - W^{(\infty)}(\theta)\big| \left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta = 0 \tag{34}$$

for every compact subset $\tilde{K} \subset \tilde{\Lambda}$.

PROOF. Let C and K represent here and in the rest of the proof some arbitrary generic constants in $\mathbb{R}_+$ whose values may differ. Let $f_\theta(\cdot)$ be defined as in Proposition 6.1.

Using the notation of Proposition 6.1, we will first assume that the dimension of $\mathcal{E}(N)$ is equal to d. As in Biggins (1992) we divide the integral into two parts for $|\eta| < \varepsilon$ and $\varepsilon \le |\eta| \le \pi$.

We will consider the case $|\eta| < \varepsilon$ first. With the standard Taylor series estimation we have for small ε and $|\eta| < \varepsilon$ with $f_\theta(0) = f_\theta'(0) = 0$,

$$\sup_{\theta\in\tilde{K}} \left|\frac{m(\lambda)}{m(\theta)}\right| \le \exp(-C|\eta|^2) \tag{35}$$

for some constant $C > 0$. Now let

$$\tilde{K}_\varepsilon := \{\lambda : \mathrm{Re}(\lambda) \in \tilde{K}, |\mathrm{Im}(\lambda)| \le \varepsilon\}.$$

Then, with Theorem 5.5, we have

$$\begin{aligned}
&\sup_{\theta\in\tilde{K}} \int_{|\eta|<\varepsilon} \sqrt{t}^d \big|W^{(t)}(\lambda) - W^{(\infty)}(\theta)\big| \left|\frac{m(\lambda)}{m(\theta)}\right|^t d\eta \\
&\quad \le \sup_{\lambda\in\tilde{K}_\varepsilon} \big|W^{(t)}(\lambda) - W^{(\infty)}(\theta)\big| \int_{|\eta|<\varepsilon\sqrt{t}} \exp(-C|\eta|^2)\, d\eta \\
&\quad \to K \sup_{\lambda\in\tilde{K}_\varepsilon} \big|W^{(\infty)}(\lambda) - W^{(\infty)}(\theta)\big|, \qquad t\to\infty.
\end{aligned}$$

The last expression can be made arbitrarily small by choosing ε sufficiently small.

Next, we consider

$$\sup_{\theta\in\tilde{K}} \int_{\varepsilon\le|\eta|\le\pi} \sqrt{t}^d \big|W^{(t)}(\theta+i\eta) - W^{(\infty)}(\theta)\big| \left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta.$$

Let $\{\eta_1,\dots,\eta_{m(\pi)}\}$ be the roots of $f_\theta(\eta)$ or equivalently the values for which $|\frac{m(\theta+i\eta)}{m(\theta)}| = 1, \theta \in \Omega_0$ and $0 < |\eta| \le \pi$. This is the case if and only if $\eta \cdot l \in 2\pi\mathbb{Z}$ for all $l \in N$ and $0 < |\eta| \le \pi$. Consequently, we have $\cos(\eta_j \cdot Z) = 1$, $\sin(\eta_j \cdot Z) = 0$, $e^{-i\eta_j\cdot Z} = 1$ and $m(\theta) = m(\theta + i\eta_j)$ a.s. for $j = 1,\dots,m = m(\pi)$. With the triangle inequality we obtain

$$\begin{aligned}
&\sup_{\theta\in\tilde K}\int_{|\eta-\eta_j|<\varepsilon}\sqrt{t}^d\big|W^{(t)}(\theta+i\eta)-W^{(\infty)}(\theta)\big|\left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta\\
&\quad\le \sup_{\theta\in\tilde K}\int_{|\eta-\eta_j|<\varepsilon}\sqrt{t}^d\big|W^{(t)}(\theta+i\eta)-W^{(t)}(\theta)\big|\left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta\\
&\qquad+\sup_{\theta\in\tilde K}\int_{|\eta-\eta_j|<\varepsilon}\sqrt{t}^d\big|W^{(t)}(\theta)-W^{(\infty)}(\theta)\big|\left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta\\
&\quad=: I_1^{(j)}+I_2^{(j)}.
\end{aligned}$$

For the second integral, $I_2^{(j)}$, let $\tilde K_\varepsilon^{(j)} := \{\lambda=\theta+i\eta\in\mathbb{C}^d : \theta\in\tilde K, |\eta-\eta_j|<\varepsilon\}$. Again with a Taylor series estimation we have for ε small and $|\eta-\eta_j|<\varepsilon$,

$$\sup_{\theta\in\tilde K}\left|\frac{m(\lambda)}{m(\theta)}\right|\le\exp(-C|\eta-\eta_j|^2).$$

Then, for every $\varepsilon>0$ we have

$$\begin{aligned}
&\sup_{\theta\in\tilde K}\int_{|\eta-\eta_j|<\varepsilon}\sqrt{t}^d\big|W^{(t)}(\theta)-W^{(\infty)}(\theta)\big|\left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta\\
&\quad\le \sup_{\theta\in\tilde K_\varepsilon^{(j)}}\big|W^{(t)}(\theta)-W^{(\infty)}(\theta)\big|\\
&\qquad\times\int_{|\eta-\eta_j|<\varepsilon} t^{d/2}\exp(-Ct|\eta-\eta_j|^2)\,d\eta\\
&\quad=\sup_{\theta\in\tilde K}\big|W^{(t)}(\theta)-W^{(\infty)}(\theta)\big|\underbrace{\int_{|\eta|<\varepsilon\sqrt{t}}\exp(-C|\eta|^2)\,d\eta}_{\to(\pi/C)^{d/2}<\infty, t\to\infty}\to 0,\qquad t\to\infty,
\end{aligned}$$

by Theorem 5.5.

Next, we claim that $W^{(t)}(\theta+i\eta_j)=W^{(t)}(\theta)$ almost surely for $j=1,\dots,m$. For the proof note that $W^{(t)}(\theta+i\eta_j)=W^{(t)}(\theta)$ because of $\eta_j\cdot Z_e\in 2\pi\mathbb{Z}$ a.s. for $e\in\pi(u), u\in\partial T_t$. In particular we have also $W^{(\infty)}(\theta+i\eta_j)=W^{(\infty)}(\theta)$ for $j=1,\dots,m$. Then, calculating the first integral, $I_1^{(j)}$, using (35),

$$\begin{aligned}
&\sup_{\theta\in\tilde K}\int_{|\eta-\eta_j|<\varepsilon}\sqrt{t}^d\big|W^{(t)}(\theta+i\eta)-W^{(t)}(\theta)\big|\left|\frac{m(\theta+i\eta)}{m(\theta)}\right|^t d\eta\\
&\quad\le \sup_{\lambda\in\tilde K_\varepsilon^{(j)}}\big|W^{(t)}(\lambda)-W^{(t)}(\theta)\big|\int_{|\eta|<\varepsilon\sqrt{t}}\exp(-C|\eta|^2)\,d\eta\\
&\quad\to(\pi/C)^{d/2}\sup_{\lambda\in\tilde K_\varepsilon^{(j)}}\big|W^{(\infty)}(\lambda)-W^{(\infty)}(\theta)\big|
\end{aligned}$$

with Theorem 5.5 letting $t \to \infty$. This can be made arbitrarily small by letting $\varepsilon \searrow 0$ and, therefore, $\eta \to \eta_j$.

Next we show that for $B := \{\eta : \varepsilon \le |\eta| \le \pi, |\eta - \eta_j| \ge \varepsilon, j = 1, \dots, m\}$, we obtain

$$\int_B \sqrt{t}^d |W^{(t)}(\lambda)| \left|\frac{m(\lambda)}{m(\theta)}\right|^t d\eta \to 0, \qquad t \to \infty,$$

uniformly in a neighbourhood of any $\theta_0 \in \tilde{\Lambda}$ and hence uniformly on $\tilde{K}$. The convergence of

$$\int_B \sqrt{t}^d |W^{(\infty)}(\theta)| \left|\frac{m(\lambda)}{m(\theta)}\right|^t d\eta$$

toward zero will immediately follow from these considerations, completing the proof.

Now we have $B = \bigcup_{j \in I} \{\eta : a_j \le |\eta| \le b_j\} \subset \{\varepsilon \le |\eta| \le \pi\}$ for some finite index set I and suitable $a_j < b_j$. Additionally, $\mu > 1$ so that for all $j \in I$ we have $f_\theta(\eta) \neq 0$ if $\theta \in \Omega_0$ and $\frac{1}{\mu} a_j \le |\eta| \le \mu b_j$. For $\theta_0 \in \tilde{\Lambda}$ let $B_{\bar{\rho}} := \{\theta \in \mathbb{R}^d : |\theta - \theta_0| \le \bar{\rho}\}$ and let $G^{(j)}_{c\bar{\rho}} := \{\lambda : \theta \in B_{c\bar{\rho}}, c^{-1} a_j \le |\eta| \le c b_j\}$, $j \in I$. It follows that $G^{(j)}_{c_1\bar{\rho}} \subset G^{(j)}_{c_2\bar{\rho}}$ for $0 < c_1 \le c_2$, $j \in I$. As $\theta_0 \in \tilde{\Lambda}$, there is $\gamma \in (1, 2]$ so that $\theta_0 \in \Omega^2_\gamma$. Then as a conclusion from the definition of Ω^2_γ we have

$$\frac{m(\gamma\theta_0)^{1/\gamma}}{m(\theta_0)} < 1.$$

Therefore, we can choose $\bar{\rho}$ sufficiently small, so that for some $\delta < 1$ we have $B_{\mu\bar{\rho}} \subset \Omega^2_\gamma$, and also

$$\frac{\sup\{m(\gamma\theta)^{1/\gamma} : \theta \in B_{\mu\bar{\rho}}\}}{\inf\{m(\theta) : \theta \in B_{\mu\bar{\rho}}\}} \le \delta.$$

Further for $\theta \in \Omega_0$ we have $|\frac{m(\lambda)}{m(\theta)}| < 1$ for all $\frac{1}{\mu} a_j < |\eta| \le \mu b_j$ and $j \in I$. We obtain that for r sufficiently small

$$\frac{\sup\{|m(\lambda)| : \lambda \in G^{(j)}_{\mu r}\}}{\inf\{m(\theta) : \theta \in B_{\mu r}\}} \le \delta, \qquad j \in I.$$

Let $B^{(t)}(\lambda) := \int e^{-\lambda \cdot x} \tilde{Z}^{(t)}(dx) - m(\lambda)^t$; then we obtain

$$W^{(t)}(\lambda) \left(\frac{m(\lambda)}{m(\theta)}\right)^t = m(\theta)^{-t} B^{(t)}(\lambda) + \left(\frac{m(\lambda)}{m(\theta)}\right)^t.$$

Let $I_n := \{t : n < t \le n+1\}$. Then for $j \in I$ we find

$$\begin{aligned} &\sup_{\theta \in B_{\bar{\rho}}, t \in I_n} \sqrt{t}^d \int_{a_j \le |\eta| \le b_j} |W^{(t)}(\lambda)| \left|\frac{m(\lambda)}{m(\theta)}\right|^t d\eta \\ &\qquad \le K \sqrt{n+1}^d \left(\frac{\sup\{|B^{(t)}(\lambda)| : \lambda \in G^{(j)}_{\bar{\rho}}, t \in I_n\}}{\inf\{m(\theta)^{n+1} : \theta \in B_{\bar{\rho}}\}} + \delta^n \right). \end{aligned} \tag{36}$$

Now since $B^{(t)}(\lambda)$ is an analytic function, we can use Cauchy's integral formula, the triangle inequality [see Biggins (1992), Lemma 3], and a compactness argument to show that

$$\sup_{\lambda \in G_{\tilde{\rho}}^{(j)}} \left|B^{(t)}(\lambda)\right| \le \frac{1}{\pi} \sum_{\tilde{r}} \int_{C_{\tilde{r}}^{(j)}} \left|B^{(t)}(z(s))\right| ds, \qquad j \in I, \tag{37}$$

where $\{C_{\tilde{r}}^{(j)}\}$ parameterize the distinguished boundaries of a finite number of disks covering $G_{\tilde{\rho}}^{(j)}$ and lying within $G_{\mu\tilde{\rho}}^{(j)}$ for $j \in I$. As in Biggins (1992) we take expected values of (37) and use Jensen's inequality for some $\alpha > 1$ to obtain that for $j \in I$,

$$E \sup_{\lambda \in G_{\tilde{\rho}}^{(j)}, t \in I_n} \left|B^{(t)}(\lambda)\right| \le K \sup_{\lambda \in G_{\mu\tilde{\rho}}^{(j)}} \Big(E \sup_{t \in I_n} \left|B^{(t)}(\lambda)\right|^{\alpha}\Big)^{1/\alpha} \tag{38}$$

for some constant K. Note that $m(\lambda)$ is never zero and that $|B^{(t)}(\lambda)/m(\lambda)^t|$ is a regular submartingale. Hence with a standard martingale inequality [Williams (1979), Lemma 43.3], we obtain

$$E \sup_{t \in I_n} \left|\frac{B^{(t)}(\lambda)}{m(\lambda)^t}\right|^{\alpha} \le \left(\frac{\alpha}{\alpha - 1}\right)^{\alpha} E \left|\frac{B^{(n+1)}(\lambda)}{m(\lambda)^{n+1}}\right|^{\alpha},$$

and so the right-hand side of (38) is less than

$$K \sup_{\lambda \in G_{\mu\tilde{\rho}}^{(j)}} \big(E\left|B^{(n+1)}(\lambda)\right|^{\alpha}\big)^{1/\alpha}, \qquad j \in I. \tag{39}$$

Now the proof continues exactly as in Biggins [(1992), Lemma 5], bounding (39) [cf. Biggins (1992), (4.6)] with Biggins [(1992), Lemma 6] and finally shows that the expected value of (36) converges toward zero, and thus (36) converges toward zero almost surely, if $n \to \infty$.

For the second case when the dimension of $\mathcal{E}(N)$ is $r < d$ we choose $a = \pi$ in Proposition 6.1 and define

$$D_j := D_j(\varepsilon, \pi) := \{\eta : |\eta| \le \pi, d(\eta, C_j) < \varepsilon\},$$

where

$$d(\eta, C_j) := \min\{|\eta - w|, w \in C_j\}$$

for $0 \le j \le m = m(\pi)$ (C_j was defined in the second part of Proposition 6.1). Note that we can choose $\delta > 0$ so small that we have $m(\pi) = m(\pi + \delta)$. Set

$$g_\theta(\eta)^{(t)} := t^{d/2} \left|W^{(t)}(\theta + i\eta)\right| \left|\frac{m(\theta + i\eta)}{m(\theta)}\right|^{t}.$$

We have $|\frac{m(\theta+i\eta)}{m(\theta)}| < 1$ for $\theta \in B = B(\varepsilon, \pi) := [-\pi, \pi]^d \setminus \bigcup_{j=0}^m D_j$, and

$$B(\varepsilon, \pi) \subset [-\pi - \delta, \pi + \delta]^d \Big\backslash \bigcup_{j=0}^m D_j(\varepsilon/2, \pi + \delta) =: B(\varepsilon/2, \pi + \delta).$$

We can now use similar arguments as in the case $r = d$ to obtain that uniformly in a neighborhood of any $\theta_0 \in \tilde{\Lambda}$, and hence uniformly on $\tilde{K}$ we have

$$\int_B g_\theta(\eta)^{(t)}\, d\eta \to 0, \qquad t \to \infty.$$

Let $\tilde{D}_j := \{\lambda \in \mathbb{R}^{d-r} : \eta_j + \sum_{k=1}^{d-r} \lambda_k \eta_{0,k} \in [-\pi, \pi]^d\}, 0 \le j \le m$, with $(\eta_{0,k})_{1 \le k \le d-r}$ as in Proposition 6.1. Then we conclude that

$$\sum_{j=0}^m \int_{D_j} g_\theta(\eta)^{(t)}\, d\eta \le \sum_{j=0}^m \int_{\lambda \in \tilde{D}_j} \int_{|\eta - (\eta_j + \sum_{k=1}^{d-r} \lambda_k \eta_{0,k})| < \varepsilon} g_\theta(\eta)^{(t)}\, d\eta\, d\lambda.$$

With the same calculations as in the case $r = d$, we get for every $\rho > 0$,

$$\int_{|\eta - (\eta_j + \sum_{k=1}^{d-r} \lambda_k \eta_{0,k})| < \varepsilon} g_\theta(\eta)^{(t)}\, d\eta < \rho$$

if $\varepsilon \le \varepsilon(\rho)$ and $t \ge T(\rho)$ independently of the choice of η_j, λ and θ. Therefore, for some $C \in (0, \infty)$ we have

$$\sum_{j=0}^m \int_{D_j} g_\theta(\eta)^{(t)}\, d\eta \le \sum_{j=0}^m \int_{\lambda \in \tilde{D}_j} \rho\, d\lambda = C\rho,$$

and this can be made arbitrarily small by choosing ρ sufficiently small. This completes the proof. $\square$

LEMMA 6.3. *For any compact set $C \subset \tilde{\Lambda}$ we have almost surely*

$$\lim_{t\to\infty} \sup_{l \in \mathbb{Z}^d, \theta \in C} e^{-\theta \cdot l} e^{t(1 - bEe^{-\theta\cdot Z})} t^{d/2} [\rho_t(l) - W^{(\infty)}(\theta) G_{(l,t)}(\theta)] = 0,$$

with

$$G_{(l,t)}(\theta) := \frac{1}{e^{-\theta \cdot l}} e^{t(bEe^{-\theta\cdot Z} - 1)} \frac{1}{(2\pi)^d} \int_{|\eta| \le \pi} e^{-btEe^{-\theta\cdot Z}(1 - e^{i\eta\cdot Z})} e^{-i\eta\cdot l}\, d\eta.$$

PROOF. Generally we have $W^{(t)}(\theta) = e^{t(1 - bEe^{-\theta\cdot Z})} \sum_l \rho_t(l) e^{-\theta\cdot l}$ [cf. (8)] and

$$\left(\frac{1}{2\pi}\right)^d \int_{|\eta| \le \pi} e^{i\eta\cdot l} e^{-i\eta\cdot \tilde{l}}\, d\eta = \begin{cases} 0, & \text{if } \tilde{l} \ne l, \\ 1, & \text{if } \tilde{l} = l. \end{cases}$$

Using this and Lemma 6.2, we obtain

$$\begin{aligned}&\rho_t(l)e^{-\theta\cdot l}e^{t(1-bEe^{-\theta\cdot Z})}\sqrt{t}^d\\&\quad= W^{(\infty)}(\theta)\sqrt{t}^d\frac{1}{(2\pi)^d}\int_{|\eta|\le\pi}e^{-btEe^{-\theta\cdot Z}(1-e^{i\eta\cdot Z})}e^{-i\eta\cdot l}\,d\eta+o(1),\end{aligned}$$

where the error term $o(1)$ is uniform in l and in θ in any compact subset of $\tilde{\Lambda}$. □

REMARK 6.4. The next corollary deals with the special case when $d=1$ and when Z takes only finitely many values in $\mathbb{N}_0$. In particular, Corollary 6.5 gives detailed information about the term $G_{l,t}(\theta)$ defined in Lemma 6.3. Note that for the proof of Corollary 6.5 we use the Cauchy formula to obtain

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{btEe^{-\theta Z}e^{i\eta Z}}e^{-i\eta l}\,d\eta=\frac{1}{l!}\frac{\partial^l}{\partial x^l}\big(e^{btE(e^{-\theta}x)^Z}\big)_{|x=0}.$$

Finally, standard calculations such as those used in Lemma 6.3 lead then to the assertion of Corollary 6.5.

COROLLARY 6.5. *Let $d=1$ and $N\subset\{0,1,\dots,L\}$ for some $L\in\mathbb{N}_0$. Then, almost surely, for any compact $C\subset\tilde{\Lambda}$, we have*

$$\lim_{t\to\infty}\sup_{l\ge 0,\theta\in C}e^{-\theta l}e^{t(1-bEe^{-\theta Z})}\sqrt{t}\bigg[\rho_t(l)-W^{(\infty)}(\theta)e^{-t}\frac{A_l}{l!}\bigg]=0$$

with

$$A_l=e^{btP(Z=0)}\sum_{D_l}\frac{l!}{a_1!a_2!\cdots a_L!}\prod_{j=1}^{L}\big(btP(Z=j)\big)^{a_j},$$

where $D_l:=\{(a_1,\dots,a_L)\in\mathbb{N}_0^L:\sum_{j=1}^L ja_j=l\}$.

The terms in Corollary 6.5 can be further calculated. Let $p_j:=P(Z=j)$ for $j\in\mathbb{N}_0$ and $\mathcal{P}^{(\xi)}(l):=\frac{e^{-\xi}}{l!}\xi^l$ the Poisson measure with parameter $\xi>0$. Then

$$\begin{aligned}e^{-btEz^Z}\frac{z^lA_l}{l!}&=\sum_{\sum ja_j=l}e^{-bt(\sum_{j=1}^L z^jp_j)}\prod_{j=1}^{L}\frac{(btp_jz^j)^{a_j}}{a_j!}\\&=\sum_{\sum ja_j=l}\prod_{j=1}^{L}\mathcal{P}^{(tbz^jp_j)}(a_j).\end{aligned}$$

With the local limit theorem in Petrov [(1975), Theorem 7 of Chapter VII, Section 2],

$$\lim_{\lambda\to\infty}\sup_l\bigg|\sqrt{2\pi\lambda}\mathcal{P}^{(\lambda)}(l)-\exp\bigg(-\frac{(l-\lambda)^2}{2\lambda}\bigg)\bigg|=0,$$

we have

$$
\begin{aligned}
\rho_t(l) &= z^{-l} e^{-t(1-bEz^Z)} t^{-1/2} \\
&\quad \times \Bigg[\sqrt{t} W(\infty, -\log(z)) \\
&\quad \times \Bigg(\sum_{\sum j a_j = l} \prod_{j=1}^{L} \Bigg\{ \frac{1}{\sqrt{2\pi t b z^j p_j}} \\
&\qquad \times \Bigg(\exp\Bigg\{ -\frac{(a_j - t b z^j p_j)^2}{2 t b z^j p_j} \Bigg\} + o(1) \Bigg) \Bigg\} \Bigg) + o(1) \Bigg].
\end{aligned}
\tag{40}
$$

For this reason, using Corollary 6.5, we obtain an expression which can be easily calculated for $A_l, l \geq 1$ in the binary search tree case [$Z = 1$ a.s. and $A_l = (2t)^l$] [cf. Chauvin et al. (2005)], or also for the random recursive tree case [$P(Z = 0) = P(Z = 1) = \frac{1}{2}$ and $A_l = e^t t^l$].

Chauvin et al. (2005) calculated and estimated the expression in (40) to prove their convergence result. To prove our main theorem, Theorem 6.6, we chose a calculation inspired by a proof of Uchiyama (1982).

Note that for $g : \mathbb{R}^d \mapsto \mathbb{R}$ we let $Dg(x) = g'(x)$ denote its gradient and $D^2 g(x)$ its Hessian matrix at point $x \in \mathbb{R}^d$, if such exists. Further, for $M \in \mathbb{R}^d \times \mathbb{R}^d$ we denote by det M its determinant. Let

$$
A(-\theta) := bEe^{-\theta \cdot Z} - 1 = E \int e^{-\theta \cdot x} X(dx) - 1 \qquad \text{for } \theta \in \Omega_0 \tag{41}
$$

with

$$
X := \sum_{j=1}^{b} \mathbb{1}_{Z_j}, \tag{42}
$$

where $(Z_1, \ldots, Z_b) \stackrel{d}{=} (Z_1^{(\emptyset)}, \ldots, Z_b^{(\emptyset)})$ are distributed as the random weights attached to the edges that connect the root with its children $1, \ldots, b$. To prove our main Theorem 6.6 we assume that X is nondegenerate in the sense that the support of the intensity measure of X is not contained in any $d-1$ dimensional hyperplane of $\mathbb{R}^d$. Then $D^2 A(-\theta)$ is a positive definite for all $\theta \in \Omega_0$, by the inverse mapping theorem the set $\Omega_0^* := \{DA(-\theta) : \theta \in \Omega_0\}$ is open, and the mapping

$$
\theta \mapsto c = DA(-\theta)
$$

is a homeomorphism of Ω_0 onto Ω_0^*.

THEOREM 6.6. *Let K be a compact subset of $\Lambda^* := \{DA(-\theta) : \theta \in \tilde{\Lambda}\} \subset \mathbb{R}^d$ and assume that X defined in (42) is nondegenerate. Then, almost surely*

$$
\lim_{n\to\infty} \sup_{c \in K} \left| \frac{U_{l_n(c)}(n)}{A_c(n)} - W_\infty(\theta(c)) \right| = 0,
$$

where $\theta(c) \in \tilde{\Lambda}$ is chosen so that for $c =: (c_1, \dots, c_d) \in \mathbb{R}^d$ we have

$$DA(-\theta(c)) = bEZe^{-\theta(c)\cdot Z} = c,$$
$$l_n(c) := \left[\frac{c\log n}{b-1}\right] := \left(\left[\frac{c_1\log n}{b-1}\right], \dots, \left[\frac{c_d\log n}{b-1}\right]\right)$$

and

$$A_c(n) := \frac{n^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}}{e^{-\theta(c)\cdot l_n(c)}\sqrt{(2\pi\log n/(b-1))^d \det D^2A(-\theta(c))}} \times \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta(c)\cdot Z})/(b-1))}.$$

For the limit $W_\infty(\cdot)$ we have

$$W_\infty(\theta) = \sum_{j=1}^{b} e^{-\theta\cdot Z_j}\big(U^{(j)}\big)^{(bEe^{-\theta\cdot Z}-1)/(b-1)} W_{\infty,(j)}(\theta), \tag{43}$$

where $W_{\infty,(1)}(\theta), \dots, W_{\infty,(b)}(\theta)$ are independent, distributed as $W_\infty(\theta)$ and independent of $(U^{(j)})$ that are defined in (26).

REMARK 6.7. From Theorem 5.5 it follows immediately from uniform convergence on compact subsets that $(W_\infty(\lambda), \lambda \in \Lambda)$ is a random analytic function. Furthermore, we have for $\theta \in \Lambda^*$, $W_\infty(\theta)$ is the unique solution of the fixed point equation (43) with expectation one. For this result, note that for $\theta \in \Lambda^*$, $(W_n(\theta))_n$ is a nonnegative martingale with an (absolute) first moment equal to one. From this we can conclude that $EW_\infty(\theta) = 1$ for all $\theta \in \Lambda^*$ (e.g., with Doob's limit law). Finally, using the result of Caliebe and Rösler (2004) we ascertain that the solution of the fixed point equation (43) with a finite nonzero expectation is unique up to a multiplicative constant.

We can also reformulate Theorem 6.6 in terms of l instead of in terms of c:

COROLLARY 6.8. *Let K be a compact subset of $\Lambda^* := \{DA(-\theta) : \theta \in \tilde{\Lambda}\}$. Assume that X defined in (42) is nondegenerate. Then, almost surely*

$$\lim_{n\to\infty} \sup_{l:(b-1)/\log n\, l\in K} \left|\frac{U_l(n)}{\bar{A}_{(b-1)l/\log n}(n)} - W_\infty(\theta_{l,n})\right| = 0,$$

where $\theta_{l,n} \in \tilde{\Lambda}$ is chosen so that

$$bEZe^{-\theta_{l,n}\cdot Z} = \frac{(b-1)}{\log n} l \in \mathbb{R}^d$$

and

$$\bar{A}_{(b-1)l/\log n}(n) := \frac{n^{(bEe^{-\theta_{l,n}\cdot Z}-1)/(b-1)}}{e^{-\theta_{l,n}\cdot l}\sqrt{(2\pi\log n/(b-1))^d \det D^2A(-\theta_{l,n})}} \times \frac{\Gamma(1/(b-1))}{\Gamma(bEe^{-\theta_{l,n}\cdot Z}/(b-1))}.$$

Further, for $W_\infty(\cdot)$ we have

$$W_\infty(\theta) = \sum_{j=1}^{b} e^{-\theta\cdot Z_j}\big(U^{(j)}\big)^{(bEe^{-\theta\cdot Z}-1)/(b-1)} W_{\infty,(j)}(\theta),$$

where $W_{\infty,(1)}(\theta),\dots,W_{\infty,(b)}(\theta)$ are independent, distributed as $W_\infty(\theta)$ and independent of $(U^{(j)})$ where $(U^{(j)})$ are defined in (26).

REMARK 6.9. Note that the following two procedures are equivalent:

(1) take the supremum over $c \in K \subset \Lambda^*$ with K a compact subset and then choose $\theta(c)$, or

(2) take the supremum over $\theta \in C \subset \tilde{\Lambda}$ with C a compact subset and then choose $c(\theta)$: $bEZe^{-\theta\cdot Z} = c(\theta)$.

PROOF OF THEOREM 6.6. For the proof we will use Lemma 6.3 and obtain

$$\begin{aligned} &e^{-\theta\cdot l}e^{t(1-bEe^{-\theta\cdot Z})}t^{d/2}\rho_t(l) \\ &\quad = W^{(\infty)}(\theta)e^{t(1-bEe^{-\theta\cdot Z})}\left(\frac{\sqrt{t}}{2\pi}\right)^d \\ &\qquad \times \int_{|\eta|\le\pi} e^{t(bEe^{-\theta\cdot Z}e^{i\eta Z}-1)}e^{-i\eta l}\,d\eta + o(1), \end{aligned} \tag{44}$$

where the error term is uniform for $\theta \in C \subset \tilde{\Lambda}$, a compact subset, and is uniform in l. We claim that

(a)

$$\begin{aligned} &e^{\tau_n(1-bEe^{-\theta(c)\cdot Z})}\left(\frac{\sqrt{\tau_n}}{2\pi}\right)^d \int_{|\eta|\le\pi} e^{\tau_n(bEe^{-\theta(c)\cdot Z}e^{i\eta\cdot Z}-1)}e^{-i\eta\cdot l_n(c)}\,d\eta \\ &\quad = \frac{1}{\sqrt{(2\pi)^d\det D^2A(-\theta(c))}} + o(1), \qquad n\to\infty, \end{aligned}$$

with $o(1)$ uniform for c in any compact subset $K \subset \Lambda^*$, and

(b)

$$\sup_{c\in K}\left|\left(\frac{Y}{b-1}\right)^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)} \times \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta(c)\cdot Z})/(b-1))} W_\infty(\theta(c))\right| < \infty \tag{45}$$

for every $K \subset \Lambda^*$ compact.

Since the functions in (45) are continuous in θ and then also in c, (b) follows immediately.

The left-hand side of (a) is equal to

$$\tau_n^{d/2} e^{-\tau_n A(-\theta(c))} \frac{1}{(2\pi)^d} \int_{|\eta|\le\pi} e^{\tau_n A(-\theta(c)+i\eta)} e^{-i\eta\cdot l_n(c)}\, d\eta$$
$$= \tau_n^{d/2} e^{-\tau_n A(-\theta(c))}\big(I_1(\tau_n) + I_2(\tau_n)\big),$$

where for $t \ge 1$ we set

$$I_1(t) = \frac{1}{(2\pi)^d} \int_{\{|\eta|<\pi t^{-1/3}\}} e^{tA(-\theta(c)+i\eta)} e^{-i\eta\cdot l_n(c)}\, d\eta \quad \text{and}$$
$$I_2(t) = \frac{1}{(2\pi)^d} \int_{\{\pi\ge|\eta|\ge\pi t^{-1/3}\}} e^{tA(-\theta(c)+i\eta)} e^{-i\eta\cdot l_n(c)}\, d\eta.$$

For $DA(-\theta(c)) = bEZe^{-\theta(c)\cdot Z} = c$ we have with $\lambda = -\theta(c) + i\eta$

$$A(\lambda) = A(-\theta(c)) + ic\cdot\eta - \tfrac{1}{2}(D^2A(-\theta(c))\eta)\cdot\eta + iB|\eta|^3 \tag{46}$$

for $\eta \to 0$ where $B = B(\lambda)$ is uniformly bounded for $\eta \to 0$ and $c \in K$. With (46), substituting $\eta = \mu/\sqrt{\tau_n}$, and using

$$\int e^{-(1/2)(D^2A(-\theta(c))\mu)\cdot\mu}\, d\mu = (2\pi)^d(\det D^2A(-\theta(c)))^{-1/2},$$

we obtain using Lemma 2.2 and (46), that

$$\begin{aligned}
&\tau_n^{d/2} e^{-\tau_n A(-\theta(c))} I_1(\tau_n)\\
&\quad= \frac{1}{(2\pi)^d} \int_{\{|\mu|<\pi\tau_n^{1/6}\}} e^{i(\mu/\sqrt{\tau_n})\cdot(c\tau_n - [c\tau_n + o(1)])}\\
&\qquad \times e^{-(1/2)(D^2A(-\theta(c))\mu)\cdot\mu} e^{iB|\mu|^3\tau_n^{-3/2}}\, d\mu\\
&\quad= \frac{1}{\sqrt{(2\pi)^d \det D^2A(-\theta(c))}} + o(1),
\end{aligned}$$

with $o(1)$ uniformly going to zero for $c \in K$.

Next we show $|t^{d/2}e^{-tA(-\theta(c))}I_2(t)| = o(1)$ uniformly for all c in a compact subset of Λ^*. We have

$$\left|t^{d/2}e^{-tA(-\theta(c))}I_2(t)\right| \le t^{d/2}\int_{\{\pi\ge|\eta|\ge\pi t^{-1/3}\}} e^{-q(\eta)t}\,d\eta$$

with

$$\begin{aligned} q(\eta) &:= A(-\theta(c)) - \operatorname{Re}\bigl(A\bigl(-\theta(c)+i\eta\bigr)\bigr)\\ &= \tfrac{1}{2}(D^2A(-\theta(c))\eta)\cdot\eta\bigl(1+o(1)\bigr), \qquad \eta\to 0, \end{aligned}$$

and $o(1)$ uniformly for all c in a compact subset of Λ^*. Additionally, if $\eta_1,\dots,\eta_m$ are those values with $0<|\eta|\le\pi$ and $\eta\cdot Z\in 2\pi\mathbb{Z}$ almost surely, we have

$$\begin{aligned} q(\eta) &= A(-\theta(c)) - \operatorname{Re}\bigl(A\bigl(-\theta(c)+i\eta\bigr)\bigr)\\ &= bEe^{-\theta(c)\cdot Z}\bigl(1-\cos(\eta\cdot Z)\bigr)\\ &= bEe^{-\theta(c)\cdot Z}\tfrac{1}{2}|\eta-\eta_j|^2\bigl(1+o(1)\bigr), \qquad \eta\to\eta_j,\ j=1,\dots,m. \end{aligned}$$

Now note that $\operatorname{Re}(A(\lambda)) < A(\theta(c))$ for $\pi\ge|\eta|>\delta, \eta\ne\eta_j$, $j=1,\dots,m$, for all $\delta>0$. We can therefore choose $\varepsilon>0$ so small and independently of $c\in K$ so that for all $\pi\ge|\eta|\ge\pi t^{-1/3}$ it follows that

$$q(\eta)t \ge \min\left\{\frac{\det D^2A(\theta(c))}{2}\pi^2t^{1/3}, t\frac{1}{2}bEe^{-\theta(c)\cdot Z}|\pi t^{-1/3}-\eta_j|^2, \tilde{C}t\right\} \ge \varepsilon t^{1/3}$$

for some suitable constant $\tilde{C}>0$ and t sufficiently large (resp. n if $t=\tau_n$). It follows $|\sqrt{t}e^{-tA(-\theta(c))}I_2(t)| = o(1)$ and also $|\sqrt{\tau_n}e^{-\tau_nA(-\theta(c))}I_2(\tau_n)| = o(1)$ with the error term as claimed. We obtain from Theorem 5.8 and (44), choosing $t=\tau_n$, $l=l_n(c)$,

$$\begin{aligned} &e^{-\theta(c)\cdot l_n(c)}e^{\tau_n(1-bEe^{-\theta(c)\cdot Z})}\tau_n^{d/2}\rho_{\tau_n}(l_n(c))\\ &\quad= \bigl(W_\infty(\theta(c))+o(1)\bigr)\\ &\qquad\times\left(\frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta(c)\cdot Z})/(b-1))}\frac{(Y/(b-1))^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}}{\sqrt{(2\pi)^d\det D^2A(-\theta(c))}}\right). \end{aligned}$$

With (18) we have

$$\begin{aligned} &e^{\tau_n(bEe^{-\theta(c)\cdot Z}-1)}\left(\frac{Y}{b-1}\right)^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}\\ &\qquad\sim_{\text{a.s.}} n^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}\bigl(1+o(1)\bigr), \end{aligned}$$

and we obtain almost surely with $\tau_n \sim_{\text{a.s.}} \frac{\log n}{b-1}$, $n \to \infty$,

$$\begin{aligned}
&\frac{e^{\tau_n(bEe^{-\theta(c)\cdot Z}-1)}(Y/(b-1))^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}}{e^{-\theta(c)\cdot l_n(c)}\sqrt{(2\pi\tau_n)^d \det D^2A(-\theta(c))}} \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta(c)\cdot Z})/(b-1))} \\
&\quad \sim_{\text{a.s.}} \frac{n^{(bEe^{-\theta(c)\cdot Z}-1)/(b-1)}}{e^{-\theta(c)\cdot l_n(c)}\sqrt{(2\pi\log n/(b-1))^d \det D^2A(-\theta(c))}} \\
&\quad\quad \times \frac{\Gamma(1/(b-1))}{\Gamma((bEe^{-\theta(c)\cdot Z})/(b-1))} \\
&\quad =: A_c(n).
\end{aligned}$$

□

REMARK 6.10. Consider the case $d = 1$ and let Z be bounded. Define

$$f(z) = 1 - bEz^Z + \log(z)bEZz^Z$$

for $z > 0$. Then it follows immediately that $\tilde{\Lambda} = \{\theta \in \mathbb{R} : f(e^{-\theta}) < 0\}$ [cf. Remark 5.7]. Since $f'(z) = \frac{1}{z}(\log(z)bEZ^2z^Z) = 0 \Leftrightarrow z = 1$ we have a single local minimum ($f''(z) > 0$) at the point $z = 1$ with $f(1) = 1 - b$.

Further, for $Z \geq 0$ a.s. we have

$$\lim_{z \searrow 0} f(z) = 1 - bp_0, \qquad \lim_{z\to\infty} f(z) = \infty.$$

Consequently, we have one root of f if $p_0 > \frac{1}{b}$ and otherwise two roots of f. In the first case, let $z_0 = 0$ and let z_1 be the root of f. In the second case, let $z_0 < z_1$ be the two roots of f. In both cases we have $\tilde{\Lambda} = (-\log(z_1), -\log(z_0))$ [where we set $-\log(0) := \infty$].

7. Examples. Note that most of the following examples are taken from Broutin and Devroye (2006).

In order to simplify notation and to work out various connections to known results, in the case $d = 1$ and for $z \in \mathbb{R}^+$, we use $M_\infty(z) := W_\infty(-\log(z))$ instead of $W_\infty(\lambda)$ [cf. (12)]. Set

$$\begin{aligned}
V &:= \{e^{-\lambda} : \lambda \in \Lambda\}, \\
\tilde{V} &:= V \cap \mathbb{R} \quad \text{and} \\
V^* &:= \{bE(Zz^Z) : z \in \tilde{V}\} = \Lambda^*.
\end{aligned}$$

In complete analogy to Corollary 6.8 we have the following:

COROLLARY 7.1. *Let $d = 1$ and let K be a compact subset of*

$$V^* := \{bE(Zz^Z) : z \in \tilde{V}\}.$$

Then almost surely

$$\lim_{n\to\infty} \sup_{l:(b-1)l/\log n\in K} \left|\frac{U_l(n)}{\hat{A}_{(b-1)l/\log n}(n)} - M_\infty(z_{l,n})\right| = 0,$$

where $z_{l,n} \in \tilde{V}$ *is chosen so that*

$$bEZz_{l,n}^Z = \frac{l(b-1)}{\log n},$$

$$\hat{A}_{(b-1)l/\log n}(n) := \frac{n^{(bEz_{l,n}^Z-1)/(b-1)}}{z_{l,n}^l\sqrt{2\pi \log nb/(b-1)E(Z^2z_{l,n}^Z)}}\frac{\Gamma(1/(b-1))}{\Gamma(bEz_{l,n}^Z/(b-1))}$$

and

$$M_\infty(z) = \sum_{j=1}^b z^{Z_j}\big(U^{(j)}\big)^{(bEz^Z-1)/(b-1)} M_{\infty,(j)}(z),$$

where $M_{\infty,(1)}(z), \dots, M_{\infty,(b)}(z)$ *are independent, distributed as* $M_\infty(z)$ *and independent of* $(U^{(j)})$ *where* $(U^{(j)})$ *is defined in* (26).

EXAMPLE 7.2 (Random binary search tree). A random binary search tree can be built incrementally. Let $U_1, \dots, U_n$ be independent random variables uniformly distributed over the unit interval. We start the tree by storing U_1 in the root node. If U_2 is greater than U_1, we add a right child to the root and store U_2 in that node. If U_2 is less than U_1, we add a left child to the root and store U_2 in that node. Then we repeat that procedure incrementally for $U_3, \dots, U_n$. The nodes where we stored some U_j for some j are called internal nodes. We refer to Devroye (1991), Devroye (1998) and the references given there for the construction of binary search trees. A summary of known results about binary search trees is given in Mahmoud (1992b) and Knuth (1998).

Let $\mathcal{T}_n$ be a random binary search tree with n (internal) nodes. We will only consider complete binary search trees. That means that we add $n+1$ external nodes to each binary search tree with n internal nodes in the following manner. If u is an internal node and has no offspring, we add two external nodes as its potential children to it. If it has already one child, then we add one external node to u as a second potential child. If u has already two children, we add nothing. Note that every external node corresponds to one of the free places available for the sorting of a new internal node and that each free place is likely to be chosen next with equal probability.

It is well known that for the random binary search tree $\mathcal{T}_n$ we have

$$\lim_{n\to\infty}\frac{|\min\{D_u : u \in \partial\mathcal{T}_n\}|}{\log n} = \alpha_-, \qquad \lim_{n\to\infty}\frac{|\max\{D_u : u \in \partial\mathcal{T}_n\}|}{\log n} = \alpha_+,$$

where α_-, α_+ are the only nonnegative solutions of the equation $x \log \frac{x}{2} - x + 2 = 1$ [see, e.g., Devroye (1986, 1987, 1998) and references given there]. Chauvin et al. (2005) proved the following result, which is covered by Theorem 6.6:

THEOREM 7.3 [Chauvin et al. (2005)]. *Almost surely, for any compact subset K of* (α_-, α_+),

$$\lim_{n\to\infty} \sup_{l:l/\log n \in K} \left(\frac{U_l(n)}{EU_l(n)} - M_\infty\left(\frac{l}{2\log n}\right)\right) = 0.$$

The profile of the binary search tree was also studied with other methods. Let α denote the limit ratio of the level and the logarithm of the tree size. Then, Fuchs, Hwang and Neininger (2006) proved convergence in distribution for $\alpha \in V^* = (\alpha_-, \alpha_+)$ and for $\alpha \in [1, 2]$ convergence of all moments to prove their results. They used the contraction method and the method of moments.

Drmota, Svante and Neininger (2008) treated a class of generalized m-ary search trees including binary search trees. For those trees they proved that in a certain range the normalized profile converges in distribution (Theorem 1.1). They used arguments based on the contraction method in order to prove convergence in distribution of several random analytic functions in a complex domain.

EXAMPLE 7.4 (Random recursive tree). A random recursive tree is built inductively. The tree $\tilde{\mathcal{T}}_1$ consist of a single node $\emptyset$, the root. Let $\tilde{\mathcal{T}}_n$ already exist and consist of the nodes $\{v_1, \ldots, v_n\}$. To grow the tree choose a node v_j out of the set $\{v_1, \ldots, v_n\}$ uniformly and at random, and attach the new node v_{n+1} as a child to node v_j [cf. Smythe and Mahmoud (1994) and references given there].

Fuchs, Hwang and Neininger (2006) showed that the profile of the random recursive tree normalized by its mean converges in distribution if the limit ratio α of the level and the logarithm of the tree size lies in $[0, e)$. They also showed convergence of all moments to hold for $\alpha \in [0, 1]$. Furthermore, they proved that inside the interval $(1, e)$ only convergence of a finite number of moments is possible. Drmota and Hwang (2005a) showed that the variance of the profile $U_l(n)$ of the random recursive tree asymptotically undergoes four phase transitions and exhibits a bimodal behavior in contrast to the unimodality of the expected value of the profiles (cf. comments made on this topic in the Introduction). For l around the most numerous level (where the width is attained), the value of the martingale shall be a.s. constant; more precisely one has $M_\infty(l/\log n) = 1$ almost surely [cf. Drmota and Hwang (2005a, 2005b)]. In the sequel, Drmota and Hwang (2005b) sketched that $U_l(n) \sim_{\text{a.s.}} M_\infty(\alpha) EU_l(n)$ almost surely, where $\alpha = \lim_n \frac{l}{\log n} \in [0, 1)$, using a martingale argument of Chauvin, Drmota and Jabbour-Hattab (2001) and Cauchy's integral formula.

We will show below as an application of Theorem 6.6 that the profile of the random recursive tree normalized by its mean converges almost surely if the limit ratio α of the level and the logarithm of the tree size lies in $(0, e)$. Additionally the profile converges uniformly for α in any compact subset of $(0, e)$.

First note that it is possible to interpret a random recursive tree with n internal nodes as a weighted binary tree $T_{\tau_{n-1}}$ with $n-1$ internal nodes by weighting the edges with independent copies of $Z \stackrel{d}{=} \text{Bernoulli}(\frac{1}{2})$ and finally by interpreting the n external nodes of the latter as the internal nodes of the former. We have to choose $Z_2 = 1 - Z_1$. This follows immediately since every external node in the weighted binary tree is equally likely to be the next one to die and to get two external children where in the recursive tree each internal node is equally likely to be the next one to produce an offspring. For more details on the construction we refer to Broutin and Devroye (2006), Section 4.2.

With this construction it is clear that not only is the height treated in Broutin and Devroye (2006), but also the distribution of the profile is kept by this construction. By embedding the random recursive tree $(\tilde{\mathcal{T}}_n)_{n\geq 1}$ in the weighted tree process and by identifying $(T_{\tau_{n-1}})_{n\geq 1}$ with $(\tilde{\mathcal{T}}_n)_{n\geq 1}$, we deduce the following:

THEOREM 7.5. *Let $K \subset (0, e)$ be a compact subset. Then almost surely*

$$\lim_{n\to\infty} \sup_{l:l/\log n\in K} \left| \frac{U_l(n)}{EU_l(n)} - M_\infty\left(\frac{l}{\log n}\right) \right| = 0,$$

and for $z \in (0, e)$,

$$M_\infty(z) = zU^z M_{\infty,(1)}(z) + (1-U)^z M_{\infty,(2)}(z),$$

where $M_{\infty,(i)}(z) \stackrel{d}{=} M_\infty(z)$, U is uniform$[0, 1]$ *random variable and $M_\infty(z)$, $M_{\infty,(1)}(z)$, $M_{\infty,(2)}(z)$ and U are independent.*

PROOF. Obviously, using Corollary 7.1, we have:

(1) $\tilde{V} = \{z > 0 : 1 - (1+z) < -\log(z)z\} = (0, e)$;
(2) $V^* = \{2E(Zz^Z) = z : z \in \tilde{V}\} = (0, e)$;

and we have $z_{l,n} = \frac{l}{\log n}$. Then with

$$\hat{A}_{l/\log n}(n) = \frac{n^{l/\log n}}{\Gamma(1 + l/\log n)(l/\log n)^l \sqrt{2\pi l}}$$

and

$$B_l(n) := \frac{(\log n)^l}{\Gamma(1 + l/\log n)\sqrt{2\pi l}} \left(\frac{e}{l}\right)^l,$$

we easily obtain

$$\frac{\hat{A}_{l_n/\log n}(n)}{B_{l_n}(n)} = 1 + o(1).$$

Finally, we note that Hwang (1995) showed that [see also Fuchs, Hwang and Neininger (2006), equation (3), page 2]

$$EU_l(n) = \frac{(\log n)^l}{l!\Gamma(1 + l/\log n)}\bigl(1 + o(1)\bigr) \sim_{\text{a.s.}} B_l(n)\bigl(1 + o(1)\bigr)$$

which yields the theorem by using the Stirling formula.

From Theorem 5.8, part (2)(b), using $Z_1 = 1 - Z_2$, $U^{(1)} + U^{(2)} = 1$ and by setting $U := U^{(1)}$, we obtain

$$M_\infty(z) = zU^z M_{\infty,(1)}(z) + (1 - U)^z M_{\infty,(2)}(z)$$

and that the claimed independence relations also hold. For the distribution of U note that if E_1, E_2 are independent, exponentially distributed random variables, then $\frac{E_1}{E_1+E_2}$ is uniform[0, 1] distributed. Now Y, defined after Lemma 5.2, is Gamma distributed with parameters (1, 1) which is the same as being exponentially distributed with expectation one. It follows that $U^{(i)}$, $i = 1, 2$, is uniformly [0, 1] distributed. □

Note that $V^* = (0, e)$ is the natural range for convergence, since, Devroye (1987) and Pittel (1994) showed for the height H_n of $\tilde{\mathcal{T}}_n$ that

$$\lim_{n\to\infty} \frac{H_n}{\log n} \xrightarrow{P} e.$$

So e should be the upper bound for any range of convergence of the profile.

EXAMPLE 7.6 (Random lopsided trees). Prefix-free codes are particularly interesting because they can be decoded directly by following a path in a tree and output a character corresponding to the codeword when reaching a leaf. Each node u represents a prefix p and its children represent the words that can be built by appending a symbol to p. When reaching a leaf, one obtains a character corresponding directly to the codeword.

Some codes have encoding length depending on the symbols. These codes are called Varn codes [cf. Varn (1971)] and naturally lead to lopsided trees. Lopsided trees are trees with edges having nonequal length. We refer to Broutin and Devroye (2006) for further details, especially on the height of such trees and for further references. There are no results about the asymptotic behavior of the profile of random lopsided trees yet.

Let $c_1 \le c_2 \le \cdots \le c_b$ be fixed positive integers. A tree is said to be lopsided if it is b-ary rooted and for each node, the edge to its jth child has length c_j, $1 \le j \le b$.

A random lopsided tree can be constructed incrementally in the following way: The tree $\tilde{\mathcal{T}}_1$ consists of a single internal node ø, the root. Additionally, we attach b external children to the root node. If $\tilde{\mathcal{T}}_n$ already exists, take an external node uniformly and at random and replace it by an internal node. The weights of the edges from that internal node to its b external children are $c_1, \dots, c_b$. It is clear that Z in the weighted b-ary tree framework has to be chosen as $Z \stackrel{d}{=} c_W$ where W is a uniform distributed random variable on the set $\{1, \dots, b\}$.

Then, with Corollary 7.1 and Remark 6.10, by embedding the lopsided trees in the b-ary tree model, and by identifying $(T_{\tau_n}) = (\tilde{\mathcal{T}}_n)$ we have the following result for the profile $(U_l(n))$:

THEOREM 7.7. *If K is a compact subset of $V^* := \{\sum_{j=1}^b c_j z^{c_j} : z \in \tilde{V}\}$ with $\tilde{V} = (z_0, z_1)$ where z_0, z_1 are defined in Remark 6.10, then, almost surely,*

$$\lim_{n\to\infty} \sup_{l:(b-1)l/\log n \in K} \left| \frac{U_l(n)}{\hat{A}_{(b-1)l/\log n}(n)} - M_\infty(z_{l,n}) \right| = 0,$$

where $z_{l,n} \in \tilde{V}$ is the solution z of $\sum_{j=1}^b c_j z^{c_j} = \frac{(b-1)l}{\log n}$, and

$$\hat{A}_{(b-1)l/\log n}(n) = \frac{n^{(\sum_{j=1}^b z_{l,n}^{c_j} - 1)/(b-1)}}{z_{l,n}^l \sqrt{2\pi \log(n) 1/(b-1) \sum_{j=1}^b c_j^2 z_{l,n}^{c_j}}} \frac{\Gamma(1/(b-1))}{\Gamma((\sum_{j=1}^b z_{l,n}^{c_j})/(b-1))}.$$

We have

$$M_\infty(z) = \sum_{j=1}^b z^{c_j} \bigl(U^{(j)}\bigr)^{1/(b-1)(\sum_{r=1}^b z^{c_r} - 1)} M_{\infty,(j)}(z)$$

with $M_\infty(z) \stackrel{d}{=} M_{\infty,(j)}(z)$, $j = 1, \dots, b$, $U^{(j)} = \frac{Y_j}{\sum_{r=1}^b Y_r}$, where Y_j are i.i.d. random Gamma$(\frac{1}{b-1}, \frac{1}{b-1})$ *distributed random variables, and $M_\infty(z)$, $M_{\infty,(1)}(z), \dots, M_{\infty,(b)}(z)$, $(U^{(1)}, \dots, U^{(b)})$ are independent.*

EXAMPLE 7.8. Consider the following tree model. Start with a single internal node. At each step the tree is expanded by choosing uniformly and at random an internal node out of the tree and then by replacing it with a given deterministic tree T^*. This model can be described by the model of lopsided trees. Assume $|T^*| = k$. Imagine a lopsided tree in which each replaced node gives birth to k children with edge weights equal to the distances of the nodes in the tree T^* to the root of T^*. The internal profile can now be calculated using the external profile of the corresponding lopsided tree.

EXAMPLE 7.9 (Plane oriented and linear recursive trees). Plane oriented trees (PORTs) are rooted trees in which the children of every node are oriented. The depths of nodes in random PORTs have been studied by Mahmoud (1992a) and their height by Pittel (1994). PORTs can be built recursively; start with one single node, the root. If $\tilde{\mathcal{T}}_n$ already exists, add node v_{n+1} uniformly and at random in one of the slots available. The slots are the positions in the tree that lead to different new trees. One can think of the slots as external nodes that are placed before, between and after internal nodes. So a node with $k \geq 1$ children has $k+1$ external nodes attached to it, always one external node between two (internal) children and one in front of the first (internal) child as well as one after the last (internal) child. If an internal node has no children, then we attach one external node to it as a potential child.

A more general model of recursive trees is based on Pittel (1994). In these recursive trees each node u has a weight w_u. When growing this kind of tree, a new node is added as a child of node u with probability proportional to w_u. Now $w_u := 1 + \beta \deg(u)$, where $\deg(u)$ denotes the number of children of u and $\beta \geq 0$, is called the parameter. When β is an integer, we can use the general tree model of Broutin and Devroye (2006) to describe those trees. Let $\beta \in \mathbb{N}$ and $\mathcal{T}_n^\beta$ be such a random recursive tree with parameter β and with n internal nodes where $\mathcal{T}_1^\beta = \{\varnothing\}$ consists of a single node, the root. The tree is expanded by adding a child to node u with probability proportional to $1 + \beta \deg(u)$. Alternatively we can choose an external node uniformly and at random where we attached to each internal node u $\deg(u)\beta + 1$ external nodes. So when we pick an external node at level d and replace it by an internal node, we attach $\beta + 2$ new external nodes to the tree, $\beta + 1$ on level d and one at level $d + 1$.

Now consider $(\beta + 2)$-ary weighted trees $(T_t)_{t \geq 0}$ where the tree process is stopped when having n internal nodes. When choosing an external node uniformly and at random from the set of all external nodes and when replacing it by an internal node, we add $\beta + 2$ external nodes to that new internal node with weights $Z = (Z_1, \dots, Z_{\beta+2})$ where $Z_{(j)} = 0, 1 \leq j \leq \beta + 1, Z_{(\beta+2)} = 1$ (the brackets in the index means that the weights are ordered by increasing values). The external profile of that tree has a similar distribution as the external profile of the random recursive tree with parameter β. Let $U_l(n)^\beta$ be the number of external nodes in the tree $\mathcal{T}_n^\beta$ on level l and $U_l(n)$ be the number of external nodes in the tree $\mathcal{T}_n = T_{\tau_n}$, the corresponding weighted $(\beta + 2)$-ary tree. Then $U_{l+1}(n + 1)^\beta \stackrel{d}{=} U_l(n)$. Note that for $\beta = 0$ we obtain the random recursive tree of Example 7.4 and for $\beta = 1$ the so called PORTs. We can choose $Z \stackrel{d}{=} \text{Bernoulli}(\frac{1}{\beta+2})$ with $|\{j \leq \beta + 2 : Z_j = 0\}| = \beta + 1$.

Before formulating the convergence theorem for these recursive trees with parameter $\beta \in \mathbb{N}$ we remark that the profile of plane-oriented recursive trees ($\beta = 1$) was analysed by Hwang (2005). For $\alpha \in [0, \frac{1}{2}]$ he obtained convergence in distribution and of all moments of the normalized profile, where the limit is uniquely characterized by its moment sequence. Hwang (2005) presented no solution for the problem of convergence for $\alpha \notin [0, \frac{1}{2}]$ since for $\alpha \notin [0, \frac{1}{2}]$ only convergence of a finite number of moments is possible. As a consequence, the characterization of the limit via moments is not possible. In addition, no fixed point equation had been known until now. Hwang (2005) anticipated convergence in distribution of the normalized profile for $\alpha \in (\frac{1}{2}, \alpha^*)$ where α^* is the solution of $\frac{1}{2} + z + z\log(2z) = 0$.

We show here even more; namely, we prove uniform almost sure convergence for α in any compact subset of $(0, \alpha^*)$. Note also that our construction also shows how the tree could be split into subtrees in order to use the contraction method. We identify $(\mathcal{T}^\beta)_{n\geq 0} = (T^*_{\tau_{n-1}})_{n\geq 0}$ where the trees $(T^*_t)_{t\geq 0}$ are defined as the trees $(T_t)_{t\geq 0}$ with the root node (resp. the imaginary edge e_0 to the root node) having itself the weight 1. Then we obtain

$$\rho^*_t(l+1) := \left|\left\{u \in \partial T^*_t : D_u = \sum_{e\in\pi(u)} Z_e = l+1\right\}\right| = \rho_t(l).$$

Note that $e_0 \in \pi(u)$ for all u. Finally, it follows that

$$U_{l+1}(n+1)^\beta = \rho^*_{\tau_n}(l+1) = \rho_{\tau_n}(l).$$

THEOREM 7.10. *Let $K \in (0, z_0)$ be a compact set where z_0 is the only solution of $z\log(z) - z - \beta = 0$. Then almost surely*

$$\lim_{n\to\infty} \sup_{l:(\beta+1)l/\log n\in K} \left|\frac{U_{l_n+1}(n+1)^\beta}{\hat{A}_{(b-1)l/\log n}(n)} - M_\infty(z_{l,n})\right| = 0$$

with

$$\hat{A}_{(\beta+1)l/\log n}(n) = \frac{n^{(\beta+z_{l,n})/(\beta+1)}}{z^l_{l,n}\sqrt{2\pi l}} \frac{\Gamma(1/(\beta+1))}{\Gamma(1+z_{l,n}/(\beta+1))},$$

where $z_{l,n} := (\beta+1)\frac{l}{\log n}$. Further

$$M_\infty(z) = \sum_{j=1}^{\beta+1} \big(U^{(j)}\big)^{(\beta+z)/(\beta+1)} M_{\infty,(j)}(z) + \left(1 - \sum_{j=1}^{\beta+1} U^{(j)}\right) z M_{\infty,(\beta+2)}(z),$$

where $M_{\infty,(1)}(z), \dots, M_{\infty,(\beta+2)}(z)$ are independent, distributed as $M_\infty(z)$ and independent of $(U^{(1)}, \dots, U^{(\beta+1)})$ with $U^{(j)} = \frac{Y_j}{\sum_{r=1}^{\beta+2} Y_r}$ where $Y_j \stackrel{d}{=} \mathrm{Gamma}(\frac{1}{\beta+1}, \frac{1}{\beta+1})$ i.i.d.

REMARK 7.11. For $\beta = 1$, z_0 is the only solution of $\frac{1}{2} + \frac{z}{2} - \frac{z}{2}\log(z) = 0$. So l has to be chosen so that $\frac{l}{\log n} \in (0, z_0/2)$. Obviously $z_0/2$ is the only solution of $\frac{1}{2} + z - z\log(2z) = 0$, $z_0/2 = \alpha^*$.

PROOF. We have $b = \beta + 2$, $Z \stackrel{d}{=} \mathrm{Bernoulli}(\frac{1}{\beta+2})$, $bEz^Z = \beta + 1 + z$. It follows that

$$\begin{aligned}\tilde{V} &= \{z \geq 0 : 1 - bEz^Z < -\log(z) bE(Zz^Z)\}\\ &= \{z \geq 0 : z\log(z) - z - \beta < 0\}.\end{aligned}$$

Since $bE(Zz^Z) = z$ we have $\tilde{V} = V^*$. Define $f(z) := z\log(z) - z - \beta$. Then

$$f'(z) = \log(z), \qquad f''(z) > 0,$$

and $(1, -1-\beta)$ is a local minimum of f. Since

$$\lim_{z \searrow 0} f(z) = -\beta, \qquad \lim_{z\to\infty} f(z) = \infty,$$

there is only one solution of $f(z) = 0$ that we will call z_0. In the interval $[0, z_0)$ f is negative other than that nonnegative. The rest follows from Corollary 7.1. □

Note that independently of this work Sulzbach (2008) proved a functional limit theorem for the profile of plane oriented recursive trees using the martingale method.

EXAMPLE 7.12 (Changes of direction in a binary search tree). Let $\tilde{\mathcal{T}}_n$ be a random binary search tree with n internal nodes, and let $u \in \tilde{\mathcal{T}}_n$. Define $D_n(u) := D_n(\pi(u))$ as the number of changes of direction in $\pi(u)$ where $\pi(u)$ is the path from the root to node u. Now let 0 and 1 encode a move down to the left and to the right, respectively. For example the path encoded by 1001010110 will have $D = 7$, that is, a count of each occurrence of the patterns 01 and 10.

We are interested in $D_l(n) := |\{u \in \partial\tilde{\mathcal{T}}_n : D_n(u) = l\}|$. Broutin and Devroye (2006) introduced the following labelling of the edges: for each level $l \geq 2$ of edges form the word $(0110)^{l-1}$, and map the binary characters to the edges from left to right. Call this weighted binary tree $\mathcal{T}_n$. Then, by embedding, we find that $D_l(n) = |\{u \in \partial\mathcal{T}_n : D_u = l\}|$. Consequently choose $Z \stackrel{d}{=} \text{binomial}(\frac{1}{2})$, $Z_2 = 1 - Z_1$, and obtain the following as in the random recursive tree case:

THEOREM 7.13. *Let $K \subset (0, e)$ be a compact subset. Then almost surely*

$$\lim_{n\to\infty} \sup_{l:l/\log n \in K} \left| \frac{D_l(n)}{ED_l(n)} - M_\infty\left(\frac{l}{\log n}\right) \right| = 0,$$

where

$$M_\infty(c) \stackrel{d}{=} cU^c M_\infty(c) + (1-U)^c M_\infty^*(c),$$

where $M_\infty(c) \stackrel{d}{=} M_\infty^*(c)$, U is a uniform[0, 1] *random variable and $M_\infty(c)$, $M_\infty^*(c)$ and U are independent.*

EXAMPLE 7.14 (Random l-colouring of the edges in a tree). Take a random binary search tree and randomly color the edges with one of l different colors. We can think of different problems in that framework. For instance, we could be interested in the question how many nodes u have exactly l red edges in $\pi(u)$ if color red appears with probability p. For this problem we have to choose $Z \stackrel{d}{=}$ Bernoulli(p), since we count only red edges which we mark with $Z = 1$ and all other colored edges with $Z = 0$. Let $D_n(l)$ be the number of nodes in the tree $\mathcal{T}_n$ with exactly l red edges in $\pi(u)$.

THEOREM 7.15. *Let $K \subset \{2zp : z \in \tilde{V}\}$ with $\tilde{V} = \{z > 0 : 2pz\log(z) - 2zp + 2p - 1 < 0\}$. Then almost surely*

$$\lim_{n\to\infty} \sup_{l:l/\log n \in K} \left| \frac{D_n(l)}{\hat{A}_{l/\log n}(n)} - M_\infty(z_{l,n}) \right| = 0,$$

where we set $z = z_{l,n} := \frac{l}{2p\log n}$ and

$$\hat{A}_{l/\log n}(n) := \frac{n^{1-2p+2pz_{l,n}}}{z_{l,n}^l \sqrt{2\pi l}} \frac{1}{\Gamma(2(1-p)+2pz_{l,n}))},$$

$$M_\infty(z) = z^{Z_1} U^{1-2p+2pz} M_{\infty,(1)}(z) + z^{Z_2}(1-U)^{1-2p+2pz} M_{\infty,(2)}(z),$$

where $M_{\infty,(1)}(z)$, $M_{\infty,(2)}(z)$ are independent, distributed as $M_\infty(z)$, independent of Z_1, Z_2, U where $U \stackrel{d}{=}$ uniform[0, 1] and Z_1, Z_2 are independent, identically distributed with Bernoulli(p) *distribution.*

PROOF. Since $b = 2, bEZz^Z = 2zp, bEz^Z = 2(1-p+zp)$, we have $\tilde{V} = \{z > 0 : 2pz\log(z) - 2zp + 2p - 1 < 0\}$ and $V^* = \{2zp : z \in \tilde{V}\}$. $\square$

For the random recursive tree the number of nodes with paths having exactly l red edges can be analyzed taking $Z \stackrel{d}{=}$ Bernoulli$(p) \times$ Bernoulli$(1/2)$, thus having $Z \stackrel{d}{=}$ Bernoulli$(p/2)$ [cf. Example 7.4]. The random recursive tree can be interpreted as a weighted binary tree. Now randomly color the edges of this tree. The probability of having a red edge is now p and the probability of having an edge with weight 1 is $1/2$. This model could alternatively be analyzed in a 2-dimensional weighted model where $Z = (Z^{(1)}, Z^{(2)})$, $Z^{(1)} \stackrel{d}{=}$ Bernoulli$(1/2)$, $Z^{(2)} \stackrel{d}{=}$ Bernoulli(p) where $Z^{(1)}, Z^{(2)}$ are independent.

EXAMPLE 7.16 (The left minus right exceedance). Let $\mathcal{T}_n$ be a binary search tree with n internal nodes and let $u \in \mathcal{T}_n$. Define $D_u := \sum_{e\in\pi(u)}(L(e) - R(e))$ where $L(e)$ is the indicator function of e being a left edge and $R(e)$ is analogously the indicator of e being a right edge. We are interested in

$$U_l(n) := |\{u \in \partial\mathcal{T}_n : D_u = l\}|,$$

namely, the number of external nodes in our binary search tree which have exactly l more left edges than right edges in the path from the root leading to that external node. Naturally in the framework of weighted b-ary trees we choose $b = 2$ and mark all right edges with -1 and weight all left edges with 1. Since right and left edges are equally likely to be chosen, we use Z with $P(Z = 1) = \frac{1}{2} = P(Z = -1)$.

THEOREM 7.17. *Let K be a compact subset of* $(z_0 - \frac{1}{z_0}, z_1 - \frac{1}{z_1})$ *where* $0 < z_0 < z_1$ *are the two positive solutions of*

$$1 - z - \frac{1}{z} + z\log(z) - \frac{1}{z}\log(z) = 0.$$

Then almost surely,

$$\lim_{n\to\infty} \sup_{l: l/\log n \in K} \left(\frac{U_l(n)}{\hat{A}_{l/\log n}(n)} - M_\infty(z_{l,n}) \right) = 0,$$

where we define

$$\hat{A}_{l/\log n}(n) := \frac{1}{\Gamma(z_{l,n} + 1/z_{l,n})} \frac{n^{z_{l,n}+1/z_{l,n}-1}}{z_{l,n}^l \sqrt{2\pi (z_{l,n} + 1/z_{l,n}) \log(n)}}$$

with

$$z_{l,n} := \frac{l}{2\log n} + \sqrt{\left(\frac{l}{2\log n}\right)^2 + 1}.$$

Further we have

$$M_\infty(z) = z(U)^{z+1/z-1} M_{\infty,(1)}(z) + \frac{1}{z}(1-U)^{z+1/z-1} M_{\infty,(2)}(z),$$

where $M_{\infty,(1)}(z)$, $M_{\infty,(2)}(z)$ *are independent, distributed as* $M_\infty(z)$ *and independent of* $U \overset{d}{=} \mathrm{uniform}[0, 1]$.

PROOF. First note that

$$\begin{aligned}\tilde{V} &= \{z > 0 : 1 - bEz^Z + \log(z) bE(Zz^Z) < 0\} \\ &= \left\{ z > 0 : 1 - \left(z + \frac{1}{z}\right) + \log(z)\left(z - \frac{1}{z}\right) < 0 \right\}.\end{aligned}$$

Easily we obtain for $z > 0$ with $g(z) := 1 - (z + \frac{1}{z}) + \log(z)(z - \frac{1}{z})$

$$\begin{aligned} &g'(z) = \log(z)\left(1 + \frac{1}{z^2}\right) = 0 \quad \Leftrightarrow \quad z = 1, \\ &g''(1) > 0, \end{aligned}$$

that g has a single local minimum at $z = 1$ with $g(1) = -1$ and since g is continuous on $(0, \infty)$ with

$$\lim_{z \searrow 0} g(z) = \infty, \qquad \lim_{z\to\infty} g(z) = \infty,$$

there are exactly two roots of the function g on $(0, \infty)$. Call them $0 < z_0 < z_1$. Now $V^* := \{z - \frac{1}{z} : z \in (z_0, z_1)\} = (z_0 - \frac{1}{z_0}, z_1 - \frac{1}{z_1})$. If $c := \frac{l}{\log(n)}$, then choose $z = z(c) > 0 : z - \frac{1}{z} = c \Leftrightarrow z^2 - cz - 1 = 0 \Leftrightarrow z = \frac{c}{2} + \sqrt{(\frac{c}{2})^2 + 1}$. From this the proof follows. $\square$

EXAMPLE 7.18 (Stochastic models for the web graph). We give a new example not contained in Broutin and Devroye (2006). The web may be viewed as a directed graph in which each vertex is a static HTML web page, and each edge is a hyperlink from one web page to another. Kumar et al. (2000) proposed and analyzed a class of random graph models inspired by empirical observations on the web graph. These observations suggested that the web is not well modeled by traditional graph models.

The linear growth copy model of Kumar et al. (2000) is parameterized by a copy factor $\alpha \in (0, 1)$ and a constant outdegree $d \geq 1$. Only the choice $d = 1$ results in a random forest that might be turned into a tree and studied using our framework.

We start with one single vertex. Assume that the random forest $\tilde{\mathcal{T}}_n$ with n internal nodes has already been created. At each time step, one vertex u is added by the following procedure: from the tree $\tilde{\mathcal{T}}_n$ choose a vertex uniformly and at random. Call this vertex v. With probability α we attach node u as a child to node v. With probability $1 - \alpha$ the node u becomes a brother of node v; that means that we attach node u as a child to the father of node v. If node v is a root with no ancestors, we let u be an isolated node, namely the root of a new tree consisting of that single node. We could now ask how many nodes are roots, nodes on level $1, 2, \dots$ and so on in that random forest.

We can interpret this random forest as a binary tree with weighted edges. When raising the forest we may instead raise the binary tree as follows. In the random forest a new node u is attached by choosing uniformly and at random an internal node v out of the existing forest $\tilde{\mathcal{T}}_n$. In the binary tree $\mathcal{T}_{n-1} = T_{\tau_{n-1}}$ we will instead choose an external node, call it $\tilde{v}$, uniformly and at random from one of the n external nodes. With the probability α, the new node in the forest will be a child of node v and located one level below v. We transmit this by making the external node $\tilde{v}$ in the binary tree an internal one and attach two new external nodes to $\tilde{v}$, one with edge weight 0, representing v, and the other with edge-weight 1, representing u. With the probability $1 - \alpha$, node u becomes a brother of v, that means it stays on the same level as node v. Then we will replace the external node $\tilde{v}$ in the binary tree by an internal node and attach two new external nodes to it, one with edge weight 0, representing v, and the other with edge-weight 0, representing u. Then an arbitrary edge has weighted one with probability $\alpha/2$ and otherwise it has weight zero. Choose $Z \stackrel{d}{=} \text{Bernoulli}(\frac{\alpha}{2})$ and the weights Z_1, Z_2 attached to the root of the binary tree as follows. Let $Z_1 \stackrel{d}{=} Z$ and $Z_2 = \mathbb{1}_{\{Z_1=0\}} Y$ with $Y \stackrel{d}{=} \text{Bernoulli}(\frac{\alpha}{2-\alpha})$ and being independent of Z_1. Then $Z_2 \stackrel{d}{=} Z$ and the resulting tree $\mathcal{T}_{n-1}$ grows as the tree described above. By embedding we obtain the following:

THEOREM 7.19. *Let K be a compact subset of $V^* := (\alpha z_0, \alpha z_1)$ with z_0, z_1 the two roots of the function $f(z) = -1 + \alpha + \alpha z(1 + \log(z))$. Then almost surely*

$$\lim_{n \to \infty} \sup_{l : l/\log n \in K} \left| \frac{U_l(n)}{\hat{A}_{l/\log n}(n)} - M_\infty\left(\frac{l}{\alpha \log n}\right) \right| = 0$$

with

$$\hat{A}_{l/\log n}(n) := \frac{n^{1-\alpha+l/\log n}}{(l/(\alpha\log n))^l\sqrt{2\pi l}}\frac{1}{\Gamma(2-\alpha+l/\log n)}$$

and

$$M_\infty(z) = z^{Z_1}(U)^{1-\alpha+c}M_{\infty,(1)}(z) + z^{\mathbb{1}\{Z_1\}Y}(1-U)^{1-\alpha+c}M_{\infty,(2)}(z),$$

where $M_{\infty,(1)}(z)$, $M_{\infty,(2)}(z)$ *are independent, distributed as* $M_\infty(z)$ *and independent of* U *where* $U \stackrel{d}{=} \text{uniform}[0,1]$.

EXAMPLE 7.20 (Combination of weights). Higher dimensional weights can be used to describe all the trees studied earlier with additional weights attached to the nodes or, alternatively, edges. For example we can study a 2-ary tree with $Z = (Z^{(1)}, Z^{(2)})$, $Z^{(1)} = 1$ and $Z^{(2)} \stackrel{d}{=} \text{binomial}(\frac{1}{2})$ which refers to a random binary search tree with edges marked with zero or one. We can think of a situation where we use the second weight for identifying if the ancestor passes some attribute on to its child ($=1$) or not ($=0$). Let

$$U_{(\tilde{l},l)}(n) = \left|\left\{u \in \partial\mathcal{T}_n : D_u = \left(\sum_{e\in\pi(u)} Z_e^{(1)}, \sum_{e\in\pi(u)} Z_e^{(2)}\right) = (\tilde{l}, l)\right\}\right|.$$

Applying Theorem 6.6 (resp. Corollary 6.8) we can then describe the asymptotics of that numbers.

REFERENCES

ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

ALDOUS, D. and SHIELDS, P. (1988). A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields* **79** 509–542. MR966174

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

BERTOIN, J. (2003). The asymptotic behavior of fragmentation processes. *J. Eur. Math. Soc.* (*JEMS*) **5** 395–416. MR2017852

BERTOIN, J. and ROUAULT, A. (2003). Additive martingales and probability tilting for homogeneous fragmentations. Available at http://www.proba.jussieu.fr/mathdoc/textes/PMA-808.pdf.

BERTOIN, J. and ROUAULT, A. (2005). Discretization methods for homogeneous fragmentations. *J. London Math. Soc.* (2) **72** 91–109. MR2145730

BIGGINS, J. D. (1977). Martingale convergence in the branching random walk. *J. Appl. Probab.* **14** 25–37. MR0433619

BIGGINS, J. D. (1991). Uniform convergence of martingales in the one-dimensional branching random walk. In *Selected Proceedings of the Sheffield Symposium on Applied Probability* (*Sheffield, 1989*). *IMS Lecture Notes—Monograph Series* **18** 159–173. IMS, Hayward, CA. MR1193068

BIGGINS, J. D. (1992). Uniform convergence of martingales in the branching random walk. *Ann. Probab.* **20** 137–151. MR1143415

BROUTIN, N. and DEVROYE, L. (2006). Large deviations for the weighted height of an extended class of trees. *Algorithmica* **46** 271–297. MR2291957

BROUTIN, N., DEVROYE, L. and MCLEISH, E. (2008). Weighted height of random trees. *Acta Inform*. **45** 237–277. MR2403296

CALIEBE, A. and RÖSLER, U. (2004). Fixed points of a smoothing transformation with finite expectation: Closing a gap. Preprint. Available at http://www.math.uni-kiel.de/stochastik/roesler/research.html.

CHAUVIN, B., DRMOTA, M. and JABBOUR-HATTAB, J. (2001). The profile of binary search trees. *Ann. Appl. Probab.* **11** 1042–1062. MR1878289

CHAUVIN, B., KLEIN, T., MARCKERT, J.-F. and ROUAULT, A. (2005). Martingales and profile of binary search trees. *Electron. J. Probab.* **10** 420–435 (electronic). MR2147314

DEVROYE, L. (1986). A note on the height of binary search trees. *J. Assoc. Comput. Mach.* **33** 489–498. MR849025

DEVROYE, L. (1987). Branching processes in the analysis of the heights of trees. *Acta Inform*. **24** 277–298. MR894557

DEVROYE, L. (1991). Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2** 303–316. MR1109697

DEVROYE, L. (1998). Branching processes and their applications in the analysis of tree structures and tree algorithms. In *Probabilistic Methods for Algorithmic Discrete Mathematics*. *Algorithms and Combinations* **16** 249–314. Springer, Berlin. MR1678582

DEVROYE, L. and HWANG, H.-K. (2006). Width and mode of the profile for some random trees of logarithmic height. *Ann. Appl. Probab.* **16** 886–918. MR2244436

DRMOTA, M. and GITTENBERGER, B. (1997). On the profile of random trees. *Random Structures Algorithms* **10** 421–451. MR1608230

DRMOTA, M. and HWANG, H.-K. (2005a). Bimodality and phase transitions in the profile variance of random binary search trees. *SIAM J. Discrete Math*. **19** 19–45 (electronic). MR2178182

DRMOTA, M. and HWANG, H.-K. (2005b). Profiles of random trees: Correlation and width of random recursive trees and binary search trees. *Adv. in Appl. Probab.* **37** 321–341. MR2144556

DRMOTA, M., JANSON, S. and NEININGER, R. (2008). A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** 288–333. MR2380900

FLAJOLET, P. and ODLYZKO, A. M. (1990). Singularity analysis of generating functions. *SIAM J. Discrete Math*. **3** 216–240. MR1039294

FUCHS, M., HWANG, H.-K. and NEININGER, R. (2006). Profiles of random trees: Limit theorems for random recursive trees and binary search trees. *Algorithmica* **46** 367–407. MR2291961

HARRIS, T. E. (1963). *The Theory of Branching Processes*. *Die Grundlehren der Mathematischen Wissenschaften* **119**. Springer, Berlin. MR0163361

HWANG, H.-K. (1995). Asymptotic expansions for the Stirling numbers of the first kind. *J. Combin. Theory Ser. A* **71** 343–351. MR1342456

HWANG, H.-K. (2005). Profiles of random trees: Plane-oriented recursive trees (extended abstract). In 2005 *International Conference on Analysis of Algorithms*. *Discrete Mathematics and Theoretical Computer Science Proceedings* **AD** 193–200 (electronic). Association of Discrete Mathematics and Theoretical Computer Science, Nancy. MR2193119

JABBOUR-HATTAB, J. (2001). Martingales and large deviations for binary search trees. *Random Structures Algorithms* **19** 112–127. MR1848787

JOFFE, A., LE CAM, L. and NEVEU, J. (1973). Sur la loi des grands nombres pour des variables aléatoires de Bernoulli attachées à un arbre dyadique. *C. R. Acad. Sci. Paris Sér. A-B* **277** A963–A964. MR0341570

KERSTING, G. (1998). The height profile of a conditioned Galton–Watson tree. Available at http://citeseer.ist.psu.edu/409331.html.

KINGMAN, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Probab.* **3** 790–801. MR0400438

KNUTH, D. E. (1998). *The Art of Computer Programming* **3**, 2nd ed. Addison-Wesley, Reading, MA. MR0378456

KUMAR, R., RAGHAVAN, P., RAJAGOPALAN, S., SIVAKUMAR, D., TOMKINS, A. and UPFAL, E. (2000). Stochastic models for the web graph. In *41st Annual Symposium on Foundations of Computer Science* (*Redondo Beach*, *CA*, *2000*) 57–65. IEEE Computer Society Press, Los Alamitos, CA. MR1931804

MAHMOUD, H. M. (1992a). Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.* **41** 237–245. MR1181723

MAHMOUD, H. M. (1992b). *Evolution of Random Search Trees*. Wiley, New York. MR1140708

MEIR, A. and MOON, J. W. (1978). On the altitude of nodes in random trees. *Canad. J. Math.* **30** 997–1015. MR506256

NEVEU, J. (1987). Multiplicative martingales for spatial branching processes. In *Seminar on Stochastic Processes* (*Princeton*, *NJ*, *1987*). *Progress in Probability and Statistics* **15** 223–242. Birkhäuser, Boston, MA. MR1046418

PARK, G. (2006). Profile of tries. Ph.D. thesis, Purdue Univ.

PARK, G., HWANG, H.-K., NICODÈME, P. and SZPANKOWSKI, W. (2009). Profiles of tries. *SIAM J. Comput.* **38** 1821–1880. MR2476278

PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **82**. Springer, New York. MR0388499

PITMAN, J. (1999). The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.* **27** 261–283. MR1681110

PITTEL, B. (1994). Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms* **5** 337–347. MR1262983

SMYTHE, R. T. and MAHMOUD, H. M. (1994). A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.* **51** 1–29 (in Ukrainian). MR1445048

SULZBACH, H. (2008). A functional limit law for the profile of plane-oriented recursive trees. In *Fifth Colloquium on Mathematics and Computer Science*. *Discrete Mathematics and Theoretical Computer Science Proceedings* **AI** 339–350. Association of Discrete Mathematics and Theoretical Computer Science, Nancy. MR2508798

UCHIYAMA, K. (1982). Spatial growth of a branching process of particles living in $\mathbb{R}^d$. *Ann. Probab.* **10** 896–918. MR672291

VARN, B. F. (1971). Optimal variable length codes (arbitrary symbol cost and equal code word probabilities). *Information and Control* **19** 289–301. MR0316166

WANG, F. J. S. (1980). The convergence of a branching Brownian motion used as a model describing the spread of an epidemic. *J. Appl. Probab.* **17** 301–312. MR568941

WATANABE, S. (1967). Limit theorem for a class of branching processes. In *Markov Processes and Potential Theory* 205–232. Wiley, New York. MR0237007

WILLIAMS, D. (1979). *Diffusions*, *Markov Processes*, *and Martingales* **1**. Wiley, Chichester. MR531031

DEPARTMENT OF MATHEMATICAL STOCHASTICS
UNIVERSITY OF FREIBURG
ECKERSTR. 1
79104 FREIBURG
GERMANY
E-MAIL: eva.schopp@gmx.de