

WEAK CONVERGENCE OF THE NUMBER OF VERTICES AT INTERMEDIATE LEVELS OF RANDOM RECURSIVE TREES

ALEXANDER IKSANOV,* *Taras Shevchenko National University of Kyiv*

ZAKHAR KABLUCHKO,** *Westfälische Wilhelms-Universität Münster*

Abstract

Let $X_n(k)$ be the number of vertices at level k in a random recursive tree with $n + 1$ vertices. We are interested in the asymptotic behavior of $X_n(k)$ for intermediate levels $k = k_n$ satisfying $k_n \rightarrow \infty$ and $k_n = o(\log n)$ as $n \rightarrow \infty$. In particular, we prove weak convergence of finite-dimensional distributions for the process $(X_n(\lfloor k_n u \rfloor))_{u>0}$, properly normalized and centered, as $n \rightarrow \infty$. The limit is a centered Gaussian process with covariance $(u, v) \mapsto (u + v)^{-1}$. One-dimensional distributional convergence of $X_n(k_n)$, properly normalized and centered, was obtained with the help of analytic tools by Fuchs *et al.* (2006). In contrast, our proofs, which are probabilistic in nature, exploit a connection of our model with certain Crump–Mode–Jagers branching processes.

Keywords: Crump–Mode–Jagers branching process; Gaussian process; intermediate level; random recursive tree; weak convergence

2010 Mathematics Subject Classification: Primary 60F05; 60J80

Secondary 60G50; 60C05

1. Introduction and main result

A (deterministic) *recursive tree* with n vertices is a rooted tree with vertices labeled with $1, 2, \dots, n$ that has the following property: the labels of the vertices on the unique path from the root (labeled with 1) to any other vertex (labeled with $m \in \{2, \dots, n\}$) form an increasing sequence. There are $(n - 1)!$ different recursive trees with n vertices, and we denote them $T_{1,n}, T_{2,n}, \dots, T_{(n-1)!,n}$. A random object \mathcal{T}_n is called *random recursive tree* with n vertices if it has uniform distribution on the set of recursive trees with n vertices, that is,

$$\mathbb{P}\{\mathcal{T}_n = T_{i,n}\} = \frac{1}{(n - 1)!}, \quad i = 1, 2, \dots, (n - 1)!$$

Let $X_n(k)$ be the number of vertices at level $k \in \mathbb{N}$ (that is, at distance k from the root) in the random recursive tree \mathcal{T}_{n+1} on $n + 1$ vertices. It is known that \mathcal{T}_{n+1} has logarithmic height (see [13, Theorem 1] and [4]); namely,

$$\frac{\max\{k \in \mathbb{N} : X_n(k) \neq 0\}}{\log n} \rightarrow e \quad \text{as } n \rightarrow \infty \text{ a.s.},$$

where we abbreviate almost surely to a.s. The asymptotic behavior of the occupation numbers $X_n(k)$ as $n \rightarrow \infty$ has been much studied for various asymptotic regimes of $k = k_n$ that is

Received 29 June 2018; revision received 23 October 2018.

* Postal address: Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine. Email address: iksan@univ.kiev.ua

** Postal address: Institut für Mathematische Statistik, Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany. Email address: zakh.kabluchko@uni-muenster.de

allowed to be a function of n . Fuchs *et al.* [6, Theorem 3] showed, by using analytic tools, that for any fixed $k \in \mathbb{N}$,

$$\frac{\sqrt{2k-1}(k-1)! (X_n(k) - (\log n)^k/k!)}{(\log n)^{k-1/2}} \xrightarrow{D} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty, \tag{1.1}$$

where $\mathcal{N}(0, 1)$ is the standardized normal distribution. Throughout, we write ‘ \xrightarrow{W} ’, ‘ \xrightarrow{D} ’, and ‘ $\xrightarrow{F.D.}$ ’ to denote weak convergence in a functional space, and weak convergence of one-dimensional and finite-dimensional distributions, respectively. Furthermore, the uniform in $k = 1, 2, \dots, o(\log n)$ rate of convergence in the uniform metric was obtained. On the other hand, in the regime where $\varepsilon \log n < k_n < (e - \varepsilon) \log n$ (with $\varepsilon > 0$ fixed), functional limit theorems with nonnormal limits were established in [2], [3], [5], and [12].

This paper is a follow-up of [10] in which a functional limit theorem was proved for the random process $(X_{[n^u]}(1), \dots, X_{[n^u]}(k))_{u \geq 0}$ for each $k \in \mathbb{N}$, properly normalized and centered, as $n \rightarrow \infty$. In particular, for $u = 1$ this result yields the following multivariate version of (1.1):

$$\left(\frac{(j-1)! (X_n(j) - (\log n)^j/j!)}{(\log n)^{j-1/2}} \right)_{j=1, \dots, k} \xrightarrow{D} (N_1, \dots, N_k) \quad \text{as } n \rightarrow \infty, \tag{1.2}$$

where (N_1, \dots, N_k) is a k -variate normal random vector with zero mean and covariances

$$\mathbb{E}N_i N_j = \frac{1}{i+j-1}, \quad 1 \leq i, j \leq k. \tag{1.3}$$

Let $(k_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers satisfying $k_n \rightarrow \infty$ and $k_n = o(\log n)$ as $n \rightarrow \infty$. Our purpose is to investigate weak convergence of the process $(X_n([k_n u]))_{u > 0}$, again properly normalized and centered, thereby providing information about occupancy of *intermediate levels* in a random recursive tree on $n + 1$ vertices. Our main result is given in Theorem 1.1.

Theorem 1.1. *Let $(k_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers satisfying $k_n \rightarrow \infty$ and $k_n = o(\log n)$ as $n \rightarrow \infty$. The following limit theorem holds for the intermediate levels of a random recursive tree with $n + 1$ vertices:*

$$\left(\frac{[k_n]^{1/2} ([k_n u] - 1)! (X_n([k_n u]) - (\log n)^{[k_n u]}/[k_n u]!)}{(\log n)^{[k_n u]-1/2}} \right)_{u > 0} \xrightarrow{F.D.} \left(\int_{[0, \infty)} e^{-uy} dB(y) \right)_{u > 0} \quad \text{as } n \rightarrow \infty, \tag{1.4}$$

where $(B(v))_{v \geq 0}$ is a standard Brownian motion.

Remark 1.1. The limit process in Theorem 1.1 can be defined via integration by parts as

$$T(u) := \int_{[0, \infty)} e^{-uy} dB(y) = u \int_0^\infty e^{-uy} B(y) dy, \quad u > 0.$$

The process T is a.s. continuous on $(0, \infty)$. However, it cannot be defined by continuity at $u = 0$ because of the oscillating behavior of the Brownian motion at ∞ . This explains that the limit theorem holds for $u > 0$ rather than $u \geq 0$.

It can be checked (details can be found in [8, Section 2]) that $T(u)$ has the same distribution as $B(1)/\sqrt{2u}$ for each $u > 0$. Therefore, we recover (1.1) when taking $u = 1$ in (1.4). Note also that

$$\mathbb{E}T(u)T(v) = (u + v)^{-1}, \quad u, v > 0.$$

As a consequence of

$$\mathbb{E}(e^u T(e^{2u})e^v T(e^{2v})) = \frac{e^{u+v}}{e^{2u} + e^{2v}} = \frac{1}{2 \cosh(u - v)}, \quad u, v \in \mathbb{R},$$

a transformed process $(e^u T(e^{2u}))_{u \in \mathbb{R}}$ is stationary Gaussian. Finally, observe that on the formal level Theorem 1.1 is consistent with (1.2). Indeed, taking $i = [k_n u]$ and $j = [k_n v]$ in (1.3), we obtain the covariance $1/([k_n u] + [k_n v] - 1) \sim (u + v)^{-1} k_n^{-1}$ as $n \rightarrow \infty$, where $x_n \sim y_n$ as $n \rightarrow \infty$ means that $\lim_{n \rightarrow \infty} (x_n/y_n) = 1$.

2. Connection to a Crump–Mode–Jagers branching process

To prove Theorem 1.1 we shall use the same approach as in [10]. The core of this approach is distributional equality (2.1) which shows that the process $(X_n([k_n u]))_{u > 0}$ is naturally embedded into an appropriate Crump–Mode–Jagers (CMJ) branching process.

The random recursive tree can be constructed in continuous time as follows. At time 0, start with a tree consisting of one vertex labeled by 1 (the root). After an exponential time with unit mean, add to this vertex an offspring labeled by 2. Any time a new vertex with label n is added to a tree, assign to each vertex of the tree a unit exponential clock that is independent of everything else. Each time some clock rings, add an offspring to the corresponding vertex and repeat the procedure. We denote by τ_n the time at which the vertex with label $n + 1$ was added to the tree. Then the tree obtained at time τ_n has the same probability law as the random recursive tree \mathcal{T}_{n+1} . Note that $\tau_0 = 0$, and for each $n \in \mathbb{N}$ the difference $\tau_n - \tau_{n-1}$ is an exponential random variable with mean $1/n$. Moreover, all such differences are independent.

By construction, the times at which the root of the tree generates offspring form arrival times of a Poisson process with unit intensity. A similar statement holds for any vertex in the tree: if a vertex was born at time t then the differences between the times at which this vertex generates offspring and t , form arrival times of a Poisson process with unit intensity. In the following, we shall generalize this construction by replacing exponential interarrival times with arbitrary positive interarrival times.

Let $(\xi_k)_{k \in \mathbb{N}}$ be independent copies of a positive random variable ξ . Let $S := (S_n)_{n \in \mathbb{N}}$ be the ordinary random walk with jumps ξ_n for $n \in \mathbb{N}$, that is, $S_n = \xi_1 + \dots + \xi_n$, $n \in \mathbb{N}$. The corresponding renewal process $(N(t))_{t \in \mathbb{R}}$ is defined by

$$N(t) := \sum_{k \geq 1} \mathbf{1}_{\{S_k \leq t\}}, \quad t \in \mathbb{R}.$$

Let $U(t) := \mathbb{E}N(t)$, where $t \in \mathbb{R}$, be the renewal function. For $t \leq 0$, we have $N(t) = 0$ a.s. and $U(t) = 0$.

We are now ready to recall the construction of the CMJ branching process relevant to us. We are interested only in the special case when the CMJ process is generated by the random walk S . At time $\tau_0 = 0$, there is one individual, called the ancestor. The ancestor produces offspring (the first generation) with birth times given by a point process $\mathcal{Z} = \sum_{n \geq 1} \delta_{S_n}$ on $\mathbb{R}_+ := [0, \infty)$. The first generation produces the second generation. The shifts of birth times

of the second generation individuals with respect to their mothers' birth times are distributed according to independent copies of the same point process \mathcal{Z} . The second generation produces the third one, and so on. All individuals act independently of each other.

For $k \in \mathbb{N}$, denote by $Y_k(t)$ the number of the k th generation individuals with birth times less than or equal to t . For example, $Y_1(t) = N(t)$ for $t \geq 0$. For $n \in \mathbb{N}$, denote by τ_n the birth time of the n th individual (in the chronological order of birth times, excluding the ancestor).

Now we are ready to state the basic observation for the proof of Theorem 1.1. In the special case when ξ has exponential distribution with unit mean, the individuals of the CMJ process correspond to vertices of the random recursive tree, the ancestor at time 0 corresponds to the root, and the generation of the individual corresponds to the distance to the root. It follows that for a sequence (k_n) with $\lim_{n \rightarrow \infty} k_n = \infty$, we have

$$(X_n([k_n u]))_{u>0} \stackrel{D}{=} (Y_{[k_n u]}(\tau_n))_{u>0}, \quad n \in \mathbb{N}. \tag{2.1}$$

The basic decomposition we need reads as

$$Y_k(t) = \sum_{i \geq 1} Y_{k-1}^{(i)}(t - S_i), \quad t \geq 0, k \geq 2,$$

where $Y_j^{(i)}(t)$ is the number of successors in the $(j + 1)$ th generation of the first generation individual born at time S_i that are born in the interval $[S_i, t + S_i]$. By the definition of the CMJ process, $(Y_j^{(1)}(t))_{t \geq 0}, (Y_j^{(2)}(t))_{t \geq 0}, \dots$ are independent copies of $(Y_j(t))_{t \geq 0}$ which are independent of S . Note that, for $k \geq 2$, $(Y_k(t))_{t \geq 0}$ is a particular instance of a random process with immigration at the epochs of a renewal process which is a renewal shot-noise process with random and independent response functions (the term was introduced in [11]; see also [9] for a review).

For $t \geq 0$ and $k \in \mathbb{N}$, we define $U_k(t) := \mathbb{E}Y_k(t)$. Then $U_1(t) = U(t)$ and

$$U_k(t) = \int_{[0,t]} U_{k-1}(t - y) dU(y), \quad k \geq 2, t \geq 0.$$

In the special case when the distribution of ξ is exponential with unit mean, we have $U_1(t) = t$ for $t \geq 0$ and, more generally,

$$U_k(t) = \frac{t^k}{k!}, \quad k \in \mathbb{N}, t \geq 0, \tag{2.2}$$

which follows from the recursive formula $U_k(t) = \int_0^t U_{k-1}(y) dy$ for $k \geq 2$.

Theorem 1.1 will be obtained as a consequence of the following two results.

Theorem 2.1. *Let $k(t)$ be any positive function satisfying $k(t) \rightarrow \infty$ and $k(t) = o(t)$ as $t \rightarrow \infty$. Assume that the distribution of ξ is exponential with unit mean. Then*

$$\left(\frac{[k(t)]^{1/2} ([k(t)u] - 1)!}{t^{[k(t)u]-1/2}} \sum_{j \geq 1} (Y_{[k(t)u]-1}^{(j)}(t - S_j) - U_{[k(t)u]-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)_{u>0} \xrightarrow{\text{F.D.}} 0 \quad \text{as } t \rightarrow \infty,$$

where, recalling (2.2), $U_{[k(t)u]-1}(t) = t^{[k(t)u]-1} / ([k(t)u] - 1)!$ for $u > 0$ and $t > 0$.

In what follows we denote by $D(0, \infty)$ (respectively, $D[0, \infty)$) the Skorokhod space of right-continuous functions defined on $(0, \infty)$ (respectively, $[0, \infty)$) with finite limits from the left at positive points.

Theorem 2.2. *Let $k(t)$ be any positive function satisfying $k(t) \rightarrow \infty$ and $k(t) = o(t)$ as $t \rightarrow \infty$. Assume that $\sigma^2 := \text{var } \xi \in (0, \infty)$ (the distribution of ξ is not assumed exponential). Then*

$$\left(\frac{[k(t)]^{1/2}([k(t)u] - 1)!}{\sqrt{\sigma^2 \mu^{-2[k(t)u]-1} t^{2[k(t)u]-1}}} \left(\sum_{j \geq 1} \frac{(t - S_j)^{[k(t)u]-1} \mathbf{1}_{\{S_j \leq t\}}}{([k(t)u] - 1)! \mu^{[k(t)u]-1}} - \frac{t^{[k(t)u]}}{([k(t)u])! \mu^{[k(t)u]}} \right) \right)_{u>0} \xrightarrow{w} (T(u))_{u>0} \text{ as } t \rightarrow \infty \tag{2.3}$$

in the J_1 -topology on $D(0, \infty)$, where $\mu = \mathbb{E}\xi < \infty$. In particular, we have in (2.3) weak convergence of the finite-dimensional distributions.

3. Proof of Theorem 1.1

Throughout the proof we assume that ξ is exponentially distributed with unit mean. In particular, we have $\mu = \sigma^2 = 1$ in the notation of Theorem 2.2. Keeping this in mind, a combination of Theorems 2.1 and 2.2 yields

$$\left(\frac{[k(t)]^{1/2}([k(t)u] - 1)!}{t^{[k(t)u]-1/2}} \left(Y_{[k(t)u]}(t) - \frac{t^{[k(t)u]}}{([k(t)u])!} \right) \right)_{u>0} \xrightarrow{\text{F.D.}} (T(u))_{u>0} \text{ as } t \rightarrow \infty \tag{3.1}$$

for any positive function $k(t)$ satisfying $k(t) \rightarrow \infty$ and $k(t) = o(t)$ as $t \rightarrow \infty$.

Given a sequence (k_n) as in the statement of Theorem 1.1, define the function $k = k(t) = k_{\lfloor e^t \rfloor}$. Note that $k(\log n) = k_n$, $k(t) \rightarrow \infty$, and $k(t) = o(t)$ as $t \rightarrow \infty$. Choose any positive $\ell = \ell(t)$ satisfying $\lim_{t \rightarrow \infty} (\ell^2(t)/(tk(t))) = \infty$ and $\ell(t) = o(t)$ as $t \rightarrow \infty$. For instance, one can take $\ell(t) = t^{3/4} k^{1/4}(t)$. For $n \in \mathbb{N}$, set $a_n = \log n / \ell(\log n)$.

Recall from the previous section that τ_n is the sum of n independent exponentially distributed random variables with means $1, \frac{1}{2}, \dots, 1/n$. This implies that $(\tau_n - (1 + \frac{1}{2} + \dots + 1/n))_{n \in \mathbb{N}}$ is a square integrable (hence, convergent) martingale with respect to the natural filtration. As a consequence, $\tau_n - \log n$ converges a.s., whence $\lim_{n \rightarrow \infty} (\tau_n - \log n) / a_n = 0$ a.s. Equivalently, given $\varepsilon > 0$ there exists an a.s. finite $N > 0$ such that $\log n - \varepsilon a_n \leq \tau_n \leq \log n + \varepsilon a_n$ whenever $n \geq N$. In what follows, for ease of notation we write 1 for ε .

Fix any positive and finite $T_1 < T_2$. For $u > 0$ and $n \in \mathbb{N}$, set

$$K_n(u) := \frac{[k(\log n)]^{1/2}([k(\log n)u] - 1)!}{(\log n)^{[k(\log n)u]-1/2}} \left(Y_{[k(\log n)u]}(\tau_n) - \frac{(\log n)^{[k(\log n)u]}}{([k(\log n)u])!} \right).$$

In view of (2.1), it suffices to show that

$$(K_n(u))_{u>0} \xrightarrow{\text{F.D.}} (T(u))_{u>0} \text{ as } n \rightarrow \infty.$$

Obviously, for all $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{u \in [T_1, T_2]} K_n(u) \mathbf{1}_{\{N > n\}} > \delta \right\} = 0.$$

Since $Y_{[k(t)u]}(t_1) \leq Y_{[k(t)u]}(t_2)$ whenever $t_1 < t_2$, we infer

$$\begin{aligned}
 K_n(u) \mathbf{1}_{\{n \geq N\}} &\leq \frac{[k(\log n)]^{1/2} ([k(\log n)u] - 1)!}{(\log n + a_n)^{[k(\log n)u] - 1/2}} \left(Y_{[k(\log n)u]}(\log n + a_n) \right. \\
 &\quad \left. - \frac{(\log n + a_n)^{[k(\log n)u]}}{([k(\log n)u])!} \right) \\
 &\quad \times \left(1 + \frac{1}{\ell(\log n)} \right)^{[k(\log n)u] - 1/2} \mathbf{1}_{\{n \geq N\}} \\
 &\quad + \frac{(k(\log n) \log n)^{1/2}}{[k(\log n)u]} \left(\left(1 + \frac{1}{\ell(\log n)} \right)^{[k(\log n)u]} - 1 \right) \\
 &=: I_n(u) \times \eta_n(u) + J_n(u).
 \end{aligned}$$

Setting in (3.1) $t = \log n$ when t is an argument of the function k and $t = \log n + a_n$, otherwise we infer $(I_n(u))_{u>0} \xrightarrow{\text{F.D.}} (T(u))_{u>0}$ as $n \rightarrow \infty$. Further, our choice of ℓ entails $\lim_{n \rightarrow \infty} (\ell(\log n)/k(\log n)) = \infty$, whence

$$\lim_{n \rightarrow \infty} \sup_{u \in [0, T_2]} |\eta_n(u) - 1| = 0 \quad \text{a.s.}$$

Finally, for large enough n ,

$$\sup_{T_1 \leq u \leq T_2} J_n(u) \leq \frac{2(k(\log n) \log n)^{1/2}}{\ell(\log n)}.$$

The right-hand side converges to 0 as $n \rightarrow \infty$ by our choice of ℓ . Combining these results, we obtain $(I_n(u) \times \eta_n(u) + J_n(u))_{u>0} \xrightarrow{\text{F.D.}} (T(u))_{u>0}$ as $n \rightarrow \infty$. The same conclusion for the lower bound of $K_n(u) \mathbf{1}_{\{n \geq N\}}$ can be derived similarly. The proof of Theorem 1.1 is complete.

4. Proof of Theorem 2.1

We first prove the following lemma.

Lemma 4.1. *Assume that the distribution of ξ is exponential with unit mean and let $k = k(t) \rightarrow \infty$ through integers and $k(t) = o(t)$ as $t \rightarrow \infty$. Then*

$$\mathbb{E} \left(\sum_{j \geq 1} (Y_{k-1}^{(j)}(t - S_j) - U_{k-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)^2 \sim \frac{1}{4} \frac{t^{2k}}{(k!)^2} \left(\frac{k}{t} \right)^2, \quad t \rightarrow \infty.$$

Proof. Without any restrictions on the distribution of a positive random variable ξ , the following formulas were obtained in [10, Lemma 4.2]: for $k \geq 2$ and $t \geq 0$,

$$\begin{aligned}
 D_k(t) := \text{var } Y_k(t) &= \mathbb{E} \left(\sum_{j \geq 1} (Y_{k-1}^{(j)}(t - S_j) - U_{k-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)^2 \\
 &\quad + \mathbb{E} \left(\sum_{j \geq 1} U_{k-1}(t - S_j) \mathbf{1}_{\{S_j \leq t\}} - U_k(t) \right)^2, \tag{4.1}
 \end{aligned}$$

$$\mathbb{E} \left(\sum_{j \geq 1} (Y_{k-1}^{(j)}(t - S_j) - U_{k-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)^2 = \int_{[0, t]} D_{k-1}(t - y) dU(y), \tag{4.2}$$

and

$$\begin{aligned} \mathbb{E} \left(\sum_{j \geq 1} U_{k-1}(t - S_j) \mathbf{1}_{\{S_j \leq t\}} - U_k(t) \right)^2 &= 2 \int_{[0,t]} U_{k-1}(t - y) U_k(t - y) dU(y) \\ &\quad + \int_{[0,t]} U_{k-1}^2(t - y) dU(y) - U_k^2(t). \end{aligned} \tag{4.3}$$

Assume now that the distribution of ξ is exponential with unit mean. Invoking (2.2) and (4.3), we obtain

$$\begin{aligned} \mathbb{E} \left(\sum_{j \geq 1} U_{k-1}(t - S_j) \mathbf{1}_{\{S_j \leq t\}} - U_k(t) \right)^2 &= 2 \int_0^t U_{k-1}(y) U_k(y) dy + \int_0^t U_{k-1}^2(y) dy - U_k^2(t) \\ &= \frac{t^{2k-1}}{((k-1)!)^2(2k-1)}, \quad k \geq 2, t \geq 0. \end{aligned}$$

Using the latter formula together with (4.1) and (4.2), we have

$$\begin{aligned} D_k(t) &= \int_0^t D_{k-1}(y) dy + \mathbb{E} \left(\sum_{j \geq 1} U_{k-1}(t - S_j) \mathbf{1}_{\{S_j \leq t\}} - U_k(t) \right)^2 \\ &= \int_0^t D_{k-1}(y) dy + \frac{t^{2k-1}}{((k-1)!)^2(2k-1)}. \end{aligned}$$

This, in combination with the boundary condition $D_1(t) = t$, immediately yields

$$D_k(t) = \sum_{i=0}^{k-1} \frac{t^{k+i}}{(i!)^2} \frac{(2i)!}{(k+i)!}, \quad k \in \mathbb{N}, t \geq 0,$$

whence, recalling (4.2),

$$\begin{aligned} \mathbb{E} \left(\sum_{j \geq 1} (Y_{k-1}^{(j)}(t - S_j) - U_{k-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)^2 &= \int_0^t D_{k-1}(y) dy \\ &= \sum_{i=0}^{k-2} \frac{t^{k+i}}{(i!)^2} \frac{(2i)!}{(k+i)!}, \quad k \geq 2, t \geq 0. \end{aligned}$$

We claim that the left-hand side is asymptotic to the $(k - 2)$ th term of the last sum, which is

$$\frac{t^{2k-2}}{((k-2)!)^2} \frac{(2k-4)!}{(2k-2)!} \sim \frac{1}{4} \frac{t^{2k}}{(k!)^2} \left(\frac{k}{t}\right)^2, \quad t \rightarrow \infty.$$

To prove this, it suffices to show that

$$\lim_{t \rightarrow \infty} \sum_{i=1}^{k-3} \frac{A(i, k, t)}{t^{k-i-2}} = 0,$$

where

$$A(i, k, t) := \frac{(k!)^2(2i)!}{(i!)^2(k+i)!k^2}.$$

Using the inequality

$$(2\pi n)^{1/2}(ne^{-1})^n \leq n! \leq e(2\pi n)^{1/2}(ne^{-1})^n, \quad n \in \mathbb{N},$$

which is a consequence of the Stirling formula in the form

$$n! = (2\pi n)^{1/2}(ne^{-1})^n e^{\theta_n/(12n)}, \quad n \in \mathbb{N},$$

where $\theta_n \in (0, 1)$, we obtain

$$\frac{1}{2^{1/2}e} A(i, k, t) \leq \frac{4^i}{i^{1/2}} \frac{k^{2k-1}}{(k+i)^{k+i+1/2} e^{k-i-2}} \leq 4^i k^{1/2} \left(\frac{k}{e}\right)^{k-i-2}. \tag{4.4}$$

This yields

$$\frac{1}{2^{1/2}e} \sum_{i=1}^{[k/2]-1} \frac{A(i, k, t)}{t^{k-i-2}} \leq k^{1/2} \sum_{i=k-[k/2]-1}^{k-3} \left(\frac{4k}{et}\right)^i \leq k^{1/2} \left(\frac{4k}{et}\right)^{k-[k/2]-1} \left(1 - \frac{4k}{et}\right)^{-1}$$

having utilized $4^i \leq 4^{k-i-2}$ which holds for $1 \leq i \leq [k/2] - 1$. The right-hand side goes to 0 as $t \rightarrow \infty$. Another appeal to (4.4) yields

$$\frac{1}{2^{1/2}e} \sum_{i=[k/2]}^{k-3} \frac{A(i, k, t)}{t^{k-i-2}} \leq k^{1/2} \left(\frac{k}{et}\right)^{k-[k/2]-2} \sum_{i=[k/2]}^{k-3} 4^i \leq \frac{1}{3} k^{1/2} 4^{k-2} \left(\frac{k}{et}\right)^{k-[k/2]-2}.$$

The right-hand side converges to 0 as $t \rightarrow \infty$ which completes the proof of the lemma. □

We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. In view of the Cramér–Wold device and Markov’s inequality, weak convergence of the finite-dimensional distributions to the zero vector is a consequence of

$$\begin{aligned} & \frac{(([ku] - 1)!)^2 k}{t^{2[ku]-1}} \mathbb{E} \left(\sum_{j \geq 1} (Y_{[ku]-1}^{(j)}(t - S_j) - U_{[ku]-1}(t - S_j)) \mathbf{1}_{\{S_j \leq t\}} \right)^2 \\ & \sim \frac{(([ku] - 1)!)^2 k}{t^{2[ku]-1}} \frac{1}{4} \frac{t^{2[ku]}}{(([ku]!)^2)} \left(\frac{[ku]}{t}\right)^2 \\ & \sim \frac{k}{4t} \\ & \rightarrow 0 \quad \text{for each } u > 0 \text{ as } t \rightarrow \infty. \end{aligned}$$

Here, we have used Lemma 4.1 for the first asymptotic equivalence. □

5. Proof of Theorem 2.2

First, we use the Cramér–Wold device to prove weak convergence of finite-dimensional distributions in (2.3). Namely, it is sufficient to check that for any $j \in \mathbb{N}$, any real $\alpha_1, \dots, \alpha_j$, and any $0 < u_1 < \dots < u_j < \infty$, we have

$$\sum_{i=1}^j \alpha_i \frac{k^{1/2}([ku_i] - 1)! Z(ku_i, t)}{\sqrt{\sigma^2 \mu^{-2[ku_i]-1} t^{2[ku_i]-1}}} \xrightarrow{D} \sum_{i=1}^j \alpha_i u_i \int_0^\infty B(y) e^{-u_i y} dy \quad \text{as } t \rightarrow \infty, \tag{5.1}$$

where

$$Z(ku, t) := \sum_{j \geq 1} \left(\frac{(t - S_j)^{[ku]-1}}{([ku] - 1)! \mu^{[ku]-1}} \mathbf{1}_{\{S_j \leq t\}} - \frac{t^{[ku]}}{([ku])! \mu^{[ku]}} \right).$$

For ease of notation, here and hereafter, we write k for $k(t)$.

We have for any $u, T > 0$, and sufficiently large t ,

$$\begin{aligned} \frac{k^{1/2}([ku] - 1)! Z(ku, t)}{\sqrt{\sigma^2 \mu^{-2[ku]-1} t^{2[ku]-1}}} &= \frac{k^{1/2}}{\sqrt{\sigma^2 \mu^{-3} t^{2[ku]-1}}} \int_{[0,t]} (t - y)^{[ku]-1} d(N(y) - \mu^{-1}y) \\ &= \frac{k^{1/2}([ku] - 1)}{\sqrt{\sigma^2 \mu^{-3} t^{2[ku]-1}}} \left(\int_0^{T k/t} (N(y) - \mu^{-1}y)(t - y)^{[ku]-2} dy \right. \\ &\quad \left. + \int_{T k/t}^t (N(y) - \mu^{-1}y)(t - y)^{[ku]-2} dy \right) \\ &= \frac{[ku] - 1}{k} \int_0^T \frac{N((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3} t/k}} \left(1 - \frac{y}{k}\right)^{[ku]-2} dy \\ &\quad + \frac{k^{1/2}([ku] - 1)}{\sqrt{\sigma^2 \mu^{-3} t^{2[ku]-1}}} \int_{T k/t}^t (N(y) - \mu^{-1}y)(t - y)^{[ku]-2} dy. \end{aligned}$$

From [7, Theorem 3.1, p. 162],

$$\frac{N(t \cdot) - \mu^{-1}(\cdot)}{\sqrt{\sigma^2 \mu^{-3} t}} \xrightarrow{w} B(\cdot) \quad \text{as } t \rightarrow \infty$$

in the J_1 -topology on $D[0, \infty)$. By Skorokhod’s representation theorem, there exist versions \widehat{N} and \widehat{B} such that

$$\lim_{t \rightarrow \infty} \sup_{0 \leq y \leq T} \left| \frac{\widehat{N}(ty) - \mu^{-1}ty}{\sqrt{\sigma^2 \mu^{-3} t}} - \widehat{B}(y) \right| = 0 \quad \text{a.s. for all } T > 0. \tag{5.2}$$

Using (5.2) with t/k replacing t in combination with

$$\lim_{t \rightarrow \infty} \sup_{0 \leq y \leq T} \left| \left(1 - \frac{y}{k(t)}\right)^{[k(t)u]-2} - e^{-uy} \right| = 0,$$

we infer

$$\lim_{t \rightarrow \infty} \sup_{u \in [0, T]} \left| \frac{\widehat{N}((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3} t/k}} \left(1 - \frac{y}{k}\right)^{[ku]-2} - \widehat{B}(y)e^{-uy} \right| = 0 \quad \text{a.s.}$$

This shows that

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_{i=1}^j \alpha_i \frac{[ku_i] - 1}{k} \int_0^T \frac{\widehat{N}((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3} t/k}} \left(1 - \frac{y}{k}\right)^{[ku_i]-2} dy \\ = \sum_{i=1}^j \alpha_i u_i \int_0^T \widehat{B}(y)e^{-u_i y} dy \quad \text{a.s.} \end{aligned}$$

and thereupon

$$\begin{aligned} & \sum_{i=1}^j \alpha_i \frac{[ku_i] - 1}{k} \int_0^T \frac{N((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3}t/k}} \left(1 - \frac{y}{k}\right)^{[ku_i]-2} dy \\ & \xrightarrow{D} \sum_{i=1}^j \alpha_i u_i \int_0^T B(y) e^{-u_i y} dy \quad \text{as } t \rightarrow \infty. \end{aligned}$$

Since $\lim_{T \rightarrow \infty} \sum_{i=1}^j \alpha_i u_i \int_0^T B(y) e^{-u_i y} dy = \sum_{i=1}^j \alpha_i u_i \int_0^\infty B(y) e^{-u_i y} dy$ a.s., it remains to prove that

$$\begin{aligned} & \lim_{T \rightarrow \infty} \limsup_{t \rightarrow \infty} \mathbb{P} \left\{ \left| \sum_{i=1}^j \alpha_i \frac{k^{1/2}([ku_i] - 1)}{t^{[ku_i]-1/2}} \int_{Tk/t}^t (N(y) - \mu^{-1}y)(t - y)^{[ku_i]-2} dy \right| > \varepsilon \right\} \\ & = 0 \quad \text{for all } \varepsilon > 0. \end{aligned}$$

In view of Markov’s inequality and the fact that $\mathbb{E}|N(y) - \mu^{-1}y| \sim \sigma \mu^{-3/2} \mathbb{E}|B(1)|y^{1/2}$ as $y \rightarrow \infty$ (see [7, Theorem 8.4, p. 98]), the latter is a consequence of

$$\lim_{T \rightarrow \infty} \limsup_{t \rightarrow \infty} \frac{k^{1/2}([ku] - 1)}{t^{[ku]-1/2}} \int_{Tk/t}^t y^{1/2}(t - y)^{[ku]-2} dy = 0 \quad \text{for } u > 0.$$

To justify it, observe that

$$\begin{aligned} \frac{k^{1/2}([ku] - 1)}{t^{[ku]-1/2}} \int_{Tk/t}^t y^{1/2}(t - y)^{[ku]-2} dy &= \frac{[ku] - 1}{k} \int_T^k y^{1/2} \left(1 - \frac{y}{k}\right)^{[ku]-2} dy \\ &\rightarrow u \int_T^\infty y^{1/2} e^{-uy} dy \quad \text{as } t \rightarrow \infty \end{aligned}$$

by Lebesgue’s dominated convergence theorem. The proof of (5.1) is complete. For later use, we note that exactly the same argument leads to

$$\begin{aligned} & \int_0^k \left| \frac{N((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3}t/k}} \right| \left(1 - \frac{y}{k}\right)^{[ku]-2} (1 + y) dy \\ & \xrightarrow{D} \int_0^\infty |B(y)| e^{-uy} (1 + y) dy \quad \text{as } t \rightarrow \infty \text{ for } u > 0. \end{aligned} \tag{5.3}$$

It remains to prove tightness in (2.3). From [1, Theorem 15.5], it suffices to show that for any $0 < a < b < \infty$, $\varepsilon > 0$, and $\gamma \in (0, 1)$, there exist $t_0 > 0$ and $\delta > 0$ such that

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{\substack{a \leq u, v \leq b, \\ |u-v| \leq \delta}} \left| \frac{k^{1/2}([ku] - 1)! Z(ku, t)}{\sqrt{\sigma^2 \mu^{-2}[ku]-1} t^{2[ku]-1}} - \frac{k^{1/2}([kv] - 1)! Z(kv, t)}{\sqrt{\sigma^2 \mu^{-2}[kv]-1} t^{2[kv]-1}} \right| > \varepsilon \right\} \\ & \leq \gamma \quad \text{for all } t \geq t_0. \end{aligned} \tag{5.4}$$

As a preparation for the proof of (5.4), we note that for $a \leq u, v \leq b$ such that $|u - v| \leq \delta$, $y \in [0, k]$, and large enough k , we have

$$\begin{aligned} & \left| \frac{[ku] - 1}{k} \left(1 - \frac{y}{k}\right)^{[ku]-2} - \frac{[kv] - 1}{k} \left(1 - \frac{y}{k}\right)^{[kv]-2} \right| \\ &= \left(1 - \frac{y}{k}\right)^{[k(u \wedge v)]-2} \left| \frac{[k(u \vee v)] - [k(u \wedge v)]}{k} \left(1 - \frac{y}{k}\right)^{[k(u \vee v)] - [k(u \wedge v)]} \right. \\ & \quad \left. - \frac{[k(u \wedge v)] - 1}{k} \left(1 - \left(1 - \frac{y}{k}\right)^{[k(u \vee v)] - [k(u \wedge v)]}\right) \right| \\ &\leq \left(1 - \frac{y}{k}\right)^{[ka]-2} \left(\frac{[k(u \vee v)] - [k(u \wedge v)]}{k} + b \frac{[k(u \vee v)] - [k(u \wedge v)]}{k} y \right) \\ &\leq C|u - v| \left(1 - \frac{y}{k}\right)^{[ka]-2} (1 + y) \\ &\leq C\delta \left(1 - \frac{y}{k}\right)^{[ka]-2} (1 + y) \end{aligned}$$

for an appropriate constant $C > 0$. With this at hand,

$$\begin{aligned} & \sup_{a \leq u, v \leq b, |u-v| \leq \delta} \left| \frac{k^{1/2}([ku] - 1)! Z(ku, t)}{\sqrt{\sigma^2 \mu^{-2[ku]-1} t^{2[ku]-1}}} - \frac{k^{1/2}([kv] - 1)! Z(kv, t)}{\sqrt{\sigma^2 \mu^{-2[kv]-1} t^{2[kv]-1}}} \right| \\ &= \sup_{a \leq u, v \leq b, |u-v| \leq \delta} \left| \int_0^k \frac{N((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3} t/k}} \left(\frac{[ku] - 1}{k} \left(1 - \frac{y}{k}\right)^{[ku]-2} \right. \right. \\ & \quad \left. \left. - \frac{[kv] - 1}{k} \left(1 - \frac{y}{k}\right)^{[kv]-2} \right) dy \right| \\ &\leq C\delta \int_0^k \left| \frac{N((t/k)y) - \mu^{-1}(t/k)y}{\sqrt{\sigma^2 \mu^{-3} t/k}} \right| \left(1 - \frac{y}{k}\right)^{[ka]-2} (1 + y) dy. \end{aligned}$$

Recalling (5.3) and choosing sufficiently small δ , we arrive at (5.4). The proof of Theorem 2.2 is complete.

6. Open problem

It is an interesting open problem as to whether weak convergence of the finite-dimensional distributions in Theorem 1.1 can be strengthened to weak convergence on $D(0, \infty)$. To ensure this it is sufficient to show that the left-hand side of the centered formula in Theorem 2.1 converges weakly to the zero function on $D(0, \infty)$. Indeed, if the latter were true, the proof of Theorem 1.1 would only require an inessential modification. However, we have been able neither prove nor disprove the aforementioned functional version of Theorem 2.1.

Acknowledgement

Part of this work was done while A. Iksanov was visiting Münster in January 2018. A. I. gratefully acknowledges hospitality and the financial support from DFG SFB 878 ‘Geometry, Groups and Actions’.

References

- [1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- [2] CHAUVIN, B., DRMOTA, M. AND JABBOUR-HATTAB, J. (2001). The profile of binary search trees. *Ann. Appl. Prob.* **11**, 1042–1062.
- [3] CHAUVIN, B., KLEIN, T., MARCKERT, J.-F. AND ROUAULT, A. (2005). Martingales and profile of binary search trees. *Electron. J. Prob.* **10**, 420–435.
- [4] DEVROYE, L. (1987). Branching processes in the analysis of the heights of trees. *Acta Inform.* **24**, 277–298.
- [5] DRMOTA, M., JANSON, S. AND NEININGER, R. (2008). A functional limit theorem for the profile of search trees. *Ann. Appl. Prob.* **18**, 288–333.
- [6] 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.
- [7] GUT, A. (2009). *Stopped Random Walks. Limit Theorems and Applications*, 2nd edn. Springer, New York.
- [8] IKSANOV, A. (2013). Functional limit theorems for renewal shot noise processes with increasing response functions. *Stoch. Process. Appl.* **123**, 1987–2010.
- [9] IKSANOV, A. (2016). *Renewal Theory for Perturbed Random Walks and Similar Processes*. Birkhäuser, Cham.
- [10] IKSANOV, A. AND KABLUCHKO, Z. (2018). A functional limit theorem for the profile of random recursive trees. *Electron. Commun. Prob.* **23**, 1–13.
- [11] IKSANOV, A., MARYNYCH, A. AND MEINERS, M. (2017). Asymptotics of random processes with immigration I: scaling limits. *Bernoulli* **23**, 1233–1278.
- [12] KABLUCHKO, Z., MARYNYCH, A. AND SULZBACH, H. (2017). General Edgeworth expansions with applications to profiles of random trees. *Ann. Appl. Prob.* **27**, 3478–3524.
- [13] PITTEL, B. (1994). Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms* **5**, 337–347.