

A WEAKLY 1-STABLE DISTRIBUTION FOR THE NUMBER OF RANDOM RECORDS AND CUTTINGS IN SPLIT TREES

CECILIA HOLMGREN,* *Uppsala University*

Abstract

In this paper we study the number of random records in an arbitrary split tree (or, equivalently, the number of random cuttings required to eliminate the tree). We show that a classical limit theorem for the convergence of sums of triangular arrays to infinitely divisible distributions can be used to determine the distribution of this number. After normalization the distributions are shown to be asymptotically weakly 1-stable. This work is a generalization of our earlier results for the random binary search tree in Holmgren (2010), which is one specific case of split trees. Other important examples of split trees include m -ary search trees, quad trees, medians of $(2k + 1)$ -trees, simplex trees, tries, and digital search trees.

Keywords: Random tree; split tree; cut; record; stable distribution; infinitely divisible distribution

2010 Mathematics Subject Classification: Primary 05C05; 05C80; 68W40; 68P10
Secondary 68R10; 60C05; 60F05; 68P05

1. Introduction

1.1. Preliminaries

We study the number of records in random split trees which were introduced by Devroye [7]. As shown by Janson [22], this number is equivalent (in distribution) to the number of cuts needed to eliminate this type of tree.

Given a rooted tree T , let each vertex v have a random value λ_v attached to it, and assume that these values are independent and identically distributed (i.i.d.) with a continuous distribution. We say that the value λ_v is a *record* if it is the smallest value in the path from the root to v . Let $X_v(T)$ denote the (random) number of records. Alternatively, we may attach random variables to the edges and let $X_e(T)$ denote the number of edges with record values. Only the order relations of the λ_v s are important, so the distribution of λ_v does not matter, i.e. we can choose any continuous distribution for λ_v .

The same random variables appear when we consider *cuttings* of the tree T as introduced by Meir and Moon [27] with the following definition. Make a random cut by choosing one vertex or edge at random. Delete this vertex or, respectively, edge so that the tree separates into several parts and keep only the part containing the root. Continue recursively until the root is cut or, respectively, only the root is left. Then the total (random) number of cuts made is $X_v(T)$ or, respectively, $X_e(T)$. More precisely, cuttings and records give random variables with the same distribution. The proof of this equivalence uses a natural coupling argument as shown in [21] and [22].

Received 11 May 2010; revision received 30 September 2010.

* Current address: Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay, France.

Email address: cecilia@math.uu.se

Expanding the earlier studies on selected conditioned Galton–Watson trees [27], [29], [30], Janson [22] found that the numbers of cuts (or the numbers of records) of conditioned Galton–Watson trees are asymptotically Rayleigh distributed. There the proof relied on the fact that the method of moments can be used. For the deterministic (nonrandom) complete binary tree, it is, however, not possible to use the method of moments to determine the asymptotic distribution of the number of cuts. To overcome this problem, Janson [21] introduced another strategy, which is to approximate $X_v(T)$ by a sum of independent random variables derived from λ_v , and then apply a classical limit theorem for triangular arrays; see, e.g. [24, Theorem 15.28]. We recently showed that Janson’s approach could also be applied to the random binary search tree [18]. Applications of the limit theorem for triangular arrays have also recently been exploited in [5] and [6] for the study of some two-sided fixed points of smoothing transforms.

In this paper we consider all types of (random) split tree defined by Devroye [7]; the binary search tree that we considered in [18] is one example of such trees. Some other important examples of split trees are m -ary search trees, quad trees, median of $(2k + 1)$ -trees, simplex trees, tries, and digital search trees. The split trees belong to the family of so-called $\log n$ trees, which are trees with height (maximal depth) asymptotically almost surely (a.a.s.) $\mathcal{O}(\log n)$. (For the notation a.a.s., see [23, p. 10].) These have similar properties to the deterministic complete binary tree with height $\lceil \log_2 n \rceil$ considered in [21]. In the complete binary tree (with high probability) most vertices are close to $\lceil \log_2 n \rceil$ (the height of the tree). In split trees on the other hand (with high probability) most vertices are close to depths of approximately $c \ln n$, where c is a constant; for the binary search tree that we investigated in [18], this depth is approximately $2 \ln n$ (see, e.g. [8]). Here, by using renewal theory we extend the methods used in [18] for the specific case of the binary search tree to show that, for split trees, it is also possible, in general, to apply a limit theorem (see, e.g. [24, Theorem 15.28]) for the convergence of sums of triangular arrays to infinitely divisible distributions to determine the asymptotic distribution of $X_v(T)$.

1.1.1. The split tree generating algorithm. The formal, comprehensive ‘split tree generating algorithm’ is as follows, with the following introductory notation; see [7] and [17]. A split tree is a finite subtree of a skeleton tree S_b , i.e. an infinite rooted tree in which each vertex has exactly b children. The split tree is constructed recursively by distributing balls one at a time to generate a subset of vertices of S_b . We say that the tree has cardinality n if n balls are distributed. Each vertex v of S_b is given an independent copy of the so-called random split vector $\mathcal{V} = (V_1, V_2, \dots, V_b)$ of probabilities, where $\sum_i V_i = 1$, $V_i \geq 0$.

Let n_v denote the total number of balls that the vertices in the subtree rooted at v hold together, and let C_v be the number of balls that are held by v itself. We say that a vertex v is a leaf if v itself holds at least one ball but no descendants of v hold any balls. Equivalently, we define v as a leaf if and only if $C_v = n_v > 0$. A vertex $v \in S_b$ is included in the split tree if and only if $n_v > 0$; if $n_v = 0$, the vertex v is not included and it is called useless.

Initially, there are no balls, i.e. $C_v = 0$ for each v . Choose an independent copy \mathcal{V}_v of \mathcal{V} for every $v \in S_b$. Add balls one by one to the root using the following recursive procedure for adding a ball to the subtree rooted at v .

Step 1. If v is not a leaf, choose child i with probability V_i and recursively add the ball to the subtree rooted at child i , by the rules given in steps 1, 2, and 3.

Step 2. If v is a leaf and $C_v = n_v < s$, then add the ball to v and stop. Thus, C_v and n_v increase by 1.

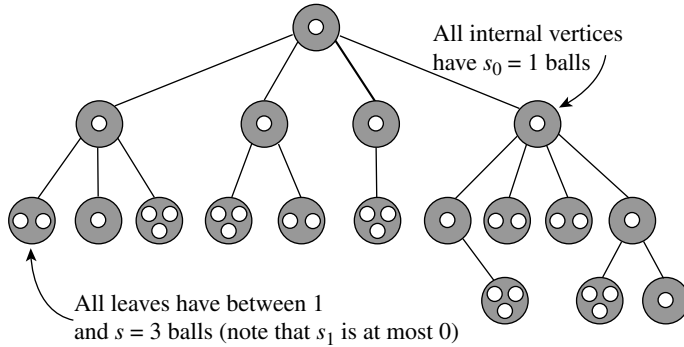


FIGURE 1: An illustration of a split tree with parameters $b = 4, s = 3, s_0 = 1,$ and $s_1 = 0$.

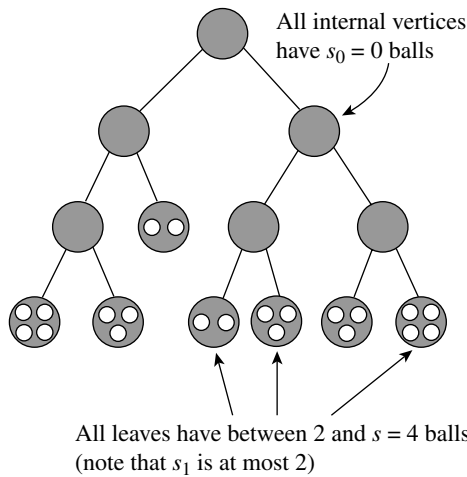


FIGURE 2: An illustration of a split tree with parameters $b = 2, s = 4, s_0 = 0,$ and $s_1 = 2$.

Step 3. If v is a leaf and $C_v = n_v = s$, the ball cannot be placed at v since it is occupied by the maximal number of balls it can hold. In this case, let $n_v = s + 1$ and $C_v = s_0$, by placing $s_0 \leq s$ randomly chosen balls at v and $s + 1 - s_0$ balls at its children. This is done by first giving s_1 randomly chosen balls to each of the b children. The remaining $s + 1 - s_0 - bs_1$ balls are placed by choosing a child for each ball independently according to the probability vector $\mathcal{V}_v = (V_1, V_2, \dots, V_b)$, and then using the algorithm described in steps 1, 2, and 3 applied to the subtree rooted at the selected child.

From step 3, it follows that the integers s_0 and s_1 have to satisfy the inequalities $0 \leq s_0 \leq s$ and $0 \leq bs_1 \leq s + 1 - s_0$. Step 3 also shows that all internal vertices hold s_0 balls and the leaves between $\max\{1, s_1\}$ and s balls. For illustrations of split trees, see Figures 1 and 2.

We can assume that the components V_i of the split vector \mathcal{V} are identically distributed. (If this were not the case, they could be made identically distributed by using a random permutation, as explained in [7].) Let V be a random variable with this distribution. In a binary search tree $b = 2$, the split vector $\mathcal{V} = (V_1, V_2)$ is $(U, 1 - U)$, where U is a uniform $U(0, 1)$ random variable, and in this specific case $N = n$. However, in most cases of split trees $n \neq N$ and N is random although n is deterministic.

1.2. Some important facts and results for split trees

1.2.1. *Results concerning depth analysis.* In [7, Theorem 1] Devroye presented a weak law and a central limit law for the depth D_n of the last ball in a split tree. Let

$$\mu := b E(-V \ln(V)), \quad \sigma^2 := b E(V \ln^2 V) - \mu^2. \tag{1.1}$$

If $P(V = 1) = 0$ and $P(V = 0) < 1$, then $D_n / \ln n \xrightarrow{P} \mu^{-1}$ (where ‘ \xrightarrow{P} ’ denotes convergence in probability) and $E(D_n) / \ln n \rightarrow \mu^{-1}$, and if $\sigma \in (0, \infty)$ also,

$$\frac{D_n - \mu^{-1} \ln n}{\sqrt{\sigma^2 \mu^{-3} \ln n}} \xrightarrow{D} N(0, 1),$$

where $N(0, 1)$ denotes the standard normal distribution and ‘ \xrightarrow{D} ’ denotes convergence in distribution. Supposing that $\sigma > 0$ is equivalent to assuming that V is not monoatomic, i.e. it is not the case that $V = 1/b$.

Let D_n^* be the average depth, and let D_k be the depth of the k th ball. In [17, Theorem 2.3], using the same assumptions for V as when proving the limit law for D_n , we showed that $E(D_n^*) / \ln n \rightarrow \mu^{-1}$ and that, for all $n / \ln n \leq k \leq n$, $\text{var}(D_k) / \ln n \rightarrow \sigma^2 \mu^{-3}$.

1.2.2. *Results concerning the number of nodes.*

Assumption 1.1. Assume as in Section 1.2.1 that $P(V = 1) = 0$, and, as in [17], for simplicity, also assume that $P(V = 0) = 0$ and that $-\ln V$ is nonlattice.

Tries and digital search trees are special forms of split trees with a random permutation of deterministic components (p_1, p_2, \dots, p_b) and, therefore, are not as random as many other examples. Of the common split trees only for some special cases of tries and digital search trees (e.g. the symmetric ones, $p_1 = p_2 = \dots = p_b = 1/b$) does $-\ln V$ have a lattice distribution. By supposing that Assumption 1.1 holds we showed in [17, Theorem 2.1] that there is a constant α depending on the type of split tree such that, for the random number of nodes N ,

$$E(N) = \alpha n + o(n), \tag{1.2}$$

and $\text{var}(N) = o(n^2)$.

Let $d(v)$ denote the depth of a node. In [17, Theorem 2.2] we showed that the expected number of nodes, where $d(v) \leq \mu^{-1} \ln n - \ln^{0.5+\varepsilon} n$ or $d(v) \geq \mu^{-1} \ln n + \ln^{0.5+\varepsilon} n$ for some arbitrary $\varepsilon > 0$, is $\mathcal{O}(n / \ln^k n)$ for any constant k . In this paper we assume that this number is $\mathcal{O}(n / \ln^3 n)$. In [17, Remark 4.3] we noted that, for any constant r , there is a constant $C > 0$ such that the expected number of nodes with $d(v) \geq C \ln n$ is $\mathcal{O}(1/n^r)$; hence, the number of vertices with ‘large’ depths can be bounded by a small error term.

1.2.3. *Results concerning the total path length.* In the present study we consider the ‘total path length’ of a tree T as the sum of all depths of the vertices in T . Since the split tree is a random tree, the total path length is a random variable, which we denote by $\Upsilon(T)$. However, a more natural definition of the total path length is the sum of all depths of the balls in T , which we denote by $\Psi(T)$.

From the fact that, for the average depth, $E(D_n^*) / \ln n \rightarrow \mu^{-1}$, it follows that

$$E(\Psi(T^n)) = \mu^{-1} n \ln n + nq(n), \tag{1.3}$$

where $q(n) = o(\ln n)$ is a function that depends on the type of split tree. Using (1.2), it follows

from the profile result [17, Theorem 2.2], including its remark [17, Remark 4.3], that

$$E(\Upsilon(T^n)) = \mu^{-1}\alpha n \ln n + nr(n), \tag{1.4}$$

where α is the constant that occurs in (1.2) and the function $r(n) = o(\ln n)$.

Assumption 1.2. Assume that $q(n)$ in (1.3) converges to some constant ζ .

Examples of split trees where the limiting constant that $q(n)$ converges to is explicitly calculated are binary search trees (see, e.g. [11]), random m -ary search trees [25], quad trees [28], the random median of a $(2k + 1)$ -tree [31], tries, and Patricia tries [3]. In fact, Assumption 1.2 has now been shown to hold for general split trees; see [4, Theorem 3.1]. We keep it as an assumption for the sake of precision.

Assumption 1.3. Assume that the result in (1.2) can be improved such that $E(N) = \alpha n + f(n)$, where $f(n) = \mathcal{O}(n/\ln^{1+\varepsilon} n)$.

Assumption 1.3 is reasonable. For instance, it holds, e.g. for m -ary search trees [26]; for such random trees, $f(n)$ is $o(\sqrt{n})$ when $m \leq 26$ and $\mathcal{O}(n^{1-\varepsilon})$ when $m \geq 27$. Furthermore, as described in Section 1.2.2, tries are special cases of split trees which are not as random as other types of split tree. Flajolet and Vallée [13] recently showed that, for most tries (as long as $-\ln V$ is not too close to being lattice), Assumption 1.3 holds.

In [4, Corollary 5.1], under Assumption 1.3, we showed, by applying [4, Theorem 3.1], that $r(n)$ in (1.4) converges to some constant ζ . By applying [4, Corollary 5.1] we showed the following result, which we will apply in the proof of the main theorem below.

Lemma 1.1. Let $L = \lfloor \beta \log_b \ln n \rfloor$ for some large enough constant β . Assume that $-\ln V$ is nonlattice and that Assumption 1.3 holds. Then

$$\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \frac{n\zeta}{\mu^{-2} \ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right), \tag{1.5}$$

where ζ is the constant that $r(n)$ in (1.4) converges to.

Proof. See Appendix A.

1.3. The main theorem

The main theorem of this study is presented below.

Theorem 1.1. Suppose that Assumptions 1.1–1.3 hold. Then, as $n \rightarrow \infty$,

$$(X_v(T^n) - C_n) \bigg/ \frac{\alpha n}{\mu^{-2} \ln^2 n} \xrightarrow{D} -W, \tag{1.6}$$

where

$$C_n := \frac{\alpha n}{\mu^{-1} \ln n} + \frac{\alpha n \ln \ln n}{\mu^{-1} \ln^2 n} - \frac{\zeta n}{\mu^{-1} \ln^2 n}$$

for the constant ζ in (1.5), and W has a weakly 1-stable distribution, with characteristic function

$$E(e^{itW}) = \exp\left(-\frac{1}{2}\mu^{-1}\pi|t| + it(C - \mu^{-1} \ln |t|)\right). \tag{1.7}$$

Here $C := -\mu^{-1} \ln \mu^{-1} + 2\mu^{-1} - \mu^{-2}\sigma^2 - \mu^{-1}\gamma - (\sigma^2 - \mu^2)/2\mu^2$, μ and σ^2 are the constants in (1.1), α is the constant in (1.2), and γ is the Euler constant. The same result holds for $X_c(T^n)$.

Recall (see, e.g. [10, Section XVII.3]) that the characteristic function of an infinitely divisible distribution is

$$\exp\left(itb - \frac{a^2t}{2} + \int_{-\infty}^{\infty} (e^{itx} - 1 - itx\mathbf{1}[|x| < 1]) \, d\nu(x)\right) \tag{1.8}$$

for constants $a \geq 0$ and $b \in \mathbb{R}$, where ν is the so-called Lévy measure which satisfies $d\nu/dx = c_{\pm}/|x|^{\alpha+1}$ for $\alpha \in (0, 2)$ and constants c_{\pm} if the distribution is weakly α -stable.

The explicit constants in Theorem 1.1 for some types of split tree are as follows. For the binary search tree (see also [18]), $\alpha = 1$ (since $n = N$), $\mu = \frac{1}{2}$, $\sigma^2 = \frac{1}{4}$, and $\zeta = 2\gamma - 4$. For quad trees, $\alpha = 1$ (since $n = N$), $\mu = d/2$, $\sigma^2 = d^2/4$, and $\zeta = (3d - 3 + 4\gamma)/2d - 2^{d+1} \sum_{m=3}^{\infty} 1/m(m^d - 2^d)$. For m -ary search trees, $\alpha = 1/2(H_m - 1)$, $\mu = \sum_{i=2}^m 1/i$, and $\sigma^2 = \sum_{i=2}^m 1/i^2$. The constant ζ for m -ary search trees when $m \geq 3$ is not as easy to calculate since the total path length $\Upsilon(T)$ is the sum of the depths of the random number of nodes (here $n \neq N$) instead of the depths of the n balls. However, applying [12, Theorem 2.4(b)] (in which general functionals of m -ary search trees are considered) with $b_n = E(N) - 1$, it is possible to give an analytic expression for ζ . Let $b_n = E(N) - 1 = (n + 1)/2(H_m - 1) + h_n$. Then the constant ζ for m -ary search trees can be expressed as

$$\zeta := \sum_{j=0}^{\infty} \frac{h_j}{(j+1)(j+2)} + \frac{(H_m)^2 - 4H_m + H_m^{(2)} + 2}{4(H_m - 1)^2},$$

where H_m is the harmonic number and $H_m^{(2)} := \sum_{j=1}^m 1/m^2$ is the second-order harmonic number. An explicit expression for $E(N)$ for all N (which determines the h_j s) is given in [25].

Remark 1.1. Even without Assumptions 1.2 and 1.3, the normalized $X_v(T^n)$ (or $X_e(T^n)$) ought to still converge to a weakly 1-stable distribution with characteristic function as in (1.7) for some constant C . However, in this case C_n in (1.6) ought to be expressed as

$$\begin{aligned} & 2 \frac{E(N)}{\mu^{-1} \ln n} - 2E\left(\sum_{d(v)=L} \frac{E(N_v | n_v) \ln(n_v/n)}{\mu^{-1} \ln^2 n}\right) - \frac{E(N)L + \alpha n \ln \ln n}{\mu^{-1} \ln^2 n} \\ & - E\left(\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \mid n_i\right), \end{aligned}$$

where $\Upsilon(T_i)$ is the total path length of the subtrees T_i rooted at depth L .

Remark 1.2. In the proof of Theorem 1.1 we obtain

$$E(e^{itW}) = \exp\left(it(C + \mu^{-1}(\gamma - 1)) + \int_0^{\infty} (e^{itx} - 1 - itx\mathbf{1}[x < 1]) \, d\nu(x)\right), \tag{1.9}$$

where C is the constant in (1.7), γ is the Euler constant, and the Lévy measure ν is supported on $(0, \infty)$ and has density $d\nu(x)/dx = \mu^{-1}/x^2$. Thus, W has a weakly 1-stable distribution. The expression in (1.9) can be simplified to (1.7).

Remark 1.3. As in [21] and [18], most records occur close to the depth where most vertices are, i.e. approximately $\mu^{-1} \ln n$ for split trees. Also, in analogy with [21] and [18], from Lemma 2.4 and the proof of Theorem 2.1, it follows that most of the random fluctuations of $X_v(T^n)$ can be explained by the values at depths close to $\ln \ln n$.

Remark 1.4. As for the binary search tree [18, Remark 1.3], for other split trees,

$$E(X_e(T^n)) - E(X_v(T^n)) = E\left(\sum_{v \neq \sigma} \frac{1}{d(v)(d(v) + 1)}\right) - 1 \sim C_1 \frac{\alpha n}{\log^2 n}$$

for some constant $C_1 > 0$, while there is no similar difference in the limit distribution; see Theorem 1.1. As in [18], this behaviour suggests that it is impossible to use the method of moments to find the distribution of the number of cuts (or records) for split trees, unlike for the (nonlogarithmic) conditioned Galton–Watson trees in [22]. In [18] we instead used methods similar to those that Janson used for the complete binary tree in [21]. In this paper we generalize the proofs in [18] to consider general split trees.

Remark 1.5. The method used here should most likely work for other trees of logarithmic height as well, and, thus, the limiting distribution for these trees should also be infinitely divisible and probably also weakly 1-stable. This turns out to be the case for the random recursive tree (that is, a logarithmic tree), where the limiting distribution of $X_e(T)$ was recently found to be weakly 1-stable; see [9, Theorem 1.1] and [20, Theorem 1.1]. However, the methods used for the recursive trees in [9] and [20] differ completely from our methods. The advantage of studying split trees compared to the whole class of $\log n$ trees is that there is a common definition that describes all split trees, and this is the reason why we only consider these trees in this paper.

1.4. Renewal theory applications for studies of split trees

1.4.1. *Subtrees.* For the split tree where the number of balls $n > s$, there are s_0 balls in the root and the cardinalities of the b subtrees are distributed as (s_1, \dots, s_1) plus a multinomial vector $(n - s_0 - bs_1, V_1, \dots, V_b)$. Thus, conditioning on the random \mathcal{V} vector that belongs to the root, the subtrees rooted at the children have cardinalities close to nV_1, \dots, nV_b . This fact is often used in applications of random binary search trees; in particular, we used this frequently in [18].

Conditioning on the split vectors, n_v at depth d is in the stochastic sense bounded by the following random variables:

$$\text{binomial}\left(n, \prod_{r=1}^d W_{r,v}\right) - sd \leq n_v \leq \text{binomial}\left(n, \prod_{r=1}^d W_{r,v}\right) + s_1d, \tag{1.10}$$

where the $W_{r,v}$, $r \in \{1, \dots, d\}$, are i.i.d. random variables given by the split vectors associated with the nodes in the unique path from v to the root; see [7] and [17]. This means in particular that $W_{r,v} \stackrel{D}{=} V$, where ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution. It follows from an application of the Chebyshev inequality that n_v for v at depth d is close to

$$M_v^n := nW_{1,v}W_{2,v} \cdots W_{d,v}; \tag{1.11}$$

see [17]. Since the n_v s for all v at the same depth are identically distributed, we sometimes omit the vertex index of $W_{r,v}$ in (1.10) and just write W_r .

1.4.2. *Results obtained using renewal theory.* In [17] we introduced renewal theory in the context of split trees, and in this study we further demonstrate its usefulness in the proof of Theorem 1.1.

For each vertex v , where $W_{r,v} \stackrel{D}{=} V$ are the i.i.d. random variables defined in Section 1.4.1, let $Y_{k,v} := -\sum_{r=1}^k \ln W_{r,v}$. Below we omit the vertex index and just write Y_k , since, for vertices v at the same depth k , the $Y_{k,v}$ s are identically distributed.

Recall from (1.11) that the subtree size n_v for v at depth k is close to $M_v^n = ne^{-Y_k}$. In [18], where the binary search tree is considered, we defined $Y_k := -\sum_{r=1}^k \ln U_r$ for uniform $U(0, 1)$ random variables U_r ; recall that the split vector is $(U, 1 - U)$, where U is uniform for this type of split tree. The sum $\sum_{r=1}^k \ln U_r$ is distributed as a $-\Gamma(k, 1)$ random variable, whereas, for general split trees, for which we do not know the common distribution function of Y_k , renewal theory can be used instead. (For an introduction to renewal theory, see, e.g. [14, Chapter II] or [1, Chapter IV, pp. 105–124].) We define the exponential renewal function

$$U(t) := \sum_{k=1}^{\infty} b^k \mathbb{P}(Y_k \leq t) = \sum_{k=1}^{\infty} F_k. \tag{1.12}$$

In view of the assumption that $\mathbb{P}(V = 1) = 0$ it is easy to show that $U(t)$ is finite for all $t > 0$; see [2, Theorem 1] or [19, Theorem 2.1]. Recall the definitions of the constants μ and σ in (1.1). In [17, Lemma 3.1] we stated the following result, which is fundamental for the proof of Theorem 1.1. The function $U(t)$ exhibits the asymptotics

$$U(t) = (\mu^{-1} + o(1))e^t \quad \text{as } t \rightarrow \infty. \tag{1.13}$$

The asymptotics in (1.13) follow as a consequence of [1, Theorem VI.5.1], which deals with nonprobability measures. In [17] we also defined $W(x) := \int_0^x e^{-t}(U(t) - \mu^{-1}e^t) dt$, and in [17, Corollary 3.2] we showed that

$$W(x) = \frac{\sigma^2 - \mu^2}{2\mu^2} - \mu^{-1} + o(1) \quad \text{as } x \rightarrow \infty. \tag{1.14}$$

2. Proofs

2.1. Notation

Most of our notation is similar to that used in [18], where the binary search tree is considered.

We use the notation \log_b for the b -logarithm (recall that a split tree with parameter b is a b -ary tree) and \ln for the e -logarithm. We treat the case $X_v(T^n)$ in Theorem 1.1 in detail and then indicate why the same result holds for $X_e(T^n)$. From now on, since it is clear that we consider the vertex model, we just write $X(T^n)$. First let $X(T^n)_y$ be $X(T^n) - 1$ conditioned on the root label $\lambda_\sigma = y$.

We say that $Y_n = o_p(a_n)$ if a_n is a positive number and Y_n is a random variable such that $Y_n/a_n \xrightarrow{p} 0$ as $n \rightarrow \infty$. We say that $Y_n = \mathcal{O}_{L^p}(a_n)$ if a_n is a positive number and Y_n is a random variable such that $(\mathbb{E}(Y_n^p))^{1/p} \leq C a_n$ for some constant C .

We sometimes use the notation $m = \mu^{-1} \ln n$. In the sequel we write T instead of T^n . For a vertex $v \in T$, we let T_v be the subtree of T rooted at v . Recall that n_v is the number of balls, and, similarly, let N_v be the number of nodes in T_v .

We write $\text{Exp}(\theta)$ for an exponential distribution with parameter θ , i.e. the density function $f(x) = e^{-x/\theta}/\theta$. Without loss of generality, we can assume that the labels λ_v have an exponential distribution $\text{Exp}(1)$. As mentioned above, this does not affect the distribution of $X(T^n)$.

Let $d(v)$ denote the depth of v , i.e. the distance to the root. Recall that V is a random variable distributed as the identically distributed components in the split vector $\mathcal{V} = (V_1, \dots, V_b)$. Also,

recall that, for each v , we let $Y_{k,v} := -\sum_{r=1}^k \ln W_{r,v}$, where $W_{r,v} \stackrel{D}{=} V$ are the i.i.d. random variables defined in Section 1.4.1. Since the $Y_{k,v}$ s are identically distributed for vertices at the same depth, we often omit the vertex index and write Y_k . Recall from (1.12) the renewal function $U(t) := \sum_{k=1}^\infty b^k P(Y_k \leq t)$.

Let Λ_{v_i} be the minimum of λ_v along the path $P(v_i) = \sigma, \dots, v_i$, from the root σ of T to v_i , $1 \leq i \leq b^L$, where the v_i are the vertices at depth $L = \lfloor \beta \log_b \ln n \rfloor$ for some constant β . Thus, the assumption that $\lambda_v \stackrel{D}{=} \text{Exp}(1)$ gives $\Lambda_{v_i} \stackrel{D}{=} \text{Exp}(1/(L + 1))$.

For simplicity, we write $T_i := T_{v_i}$, $n_i := n_{v_i}$, $N_i := N_{v_i}$, and $\Lambda_i := \Lambda_{v_i}$. We write $d_i(v) := d(v) - L$ (i.e. the depth in the subtree T_i , $i \in \{1, \dots, b^L\}$, of a vertex $v \in T_i$). We say that a vertex v in T^n is ‘good’ if

$$\mu^{-1} \ln n - \ln^{0.6} n \leq d(v) \leq \mu^{-1} \ln n + \ln^{0.6} n,$$

and ‘bad’ otherwise. In particular, a vertex $v \in T_i$ is ‘good’ if

$$\mu^{-1} \ln n_i - \ln^{0.6} n_i \leq d_i(v) \leq \mu^{-1} \ln n_i + \ln^{0.6} n_i. \tag{2.1}$$

We define $\varphi(T_i, \Lambda_i) := E(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$ and $\psi(T_i, \Lambda_i) := \text{var}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$. The conditional expected value given n_i is denoted by $E_{n_i}(Z) := E(Z \mid n_i)$.

We write $\xi_v := n_v \mu^{-1} \ln n e^{-\lambda_v \mu^{-1} \ln n} / n$, which is used in the later part of the proof. We define Ω_L as the σ -field generated by $\{n_v, d(v) \leq L\}$. Finally, we write \mathcal{G}_j as the σ -field generated by the \mathcal{V} vectors for all v with $d(v) \leq j$. Equivalently, this is the σ -field generated by $\{W_{r,v}, r \in \{1, 2, \dots, j\}\}$ for all v with $d(v) = j$. In particular, we use the fact that $\{n_v, d(v) \leq L\}$ up to small errors is determined by the σ -field \mathcal{G}_L ; this follows because of the representation of subtree sizes in Section 1.4.1.

2.2. Expressing the number of records as a sum of triangular arrays

Recall from Section 2.1 that $\varphi(T_i, \Lambda_i) := E(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$.

Lemma 2.1. *For all subtrees T_i rooted at v_i with $d(v_i) = L$, conditioned on n_i ,*

$$\begin{aligned} \varphi(T_i, \Lambda_i) &= \frac{N_i}{\mu^{-1} \ln n_i} (1 - e^{-(\mu^{-1} \ln n_i) \Lambda_i}) - \frac{\Upsilon(T_i) - \mu^{-1} N_i \ln n_i}{\mu^{-2} \ln^2 n_i} \\ &+ \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} + \mathcal{O}_{L^1} \left(\frac{n_i}{\ln^{2.2} n_i} \right), \end{aligned} \tag{2.2}$$

where $\Upsilon(T_i)$ is the total path length of T_i , and the ‘good’ $v \in T_i$ are those with $d_i(v)$ satisfying (2.1).

Proof. For each vertex $v \in T_i$, let I_v be the indicator that λ_v is the minimum value (from v to v_i) given T_i and Λ_i . We obtain $\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} E(I_v)$. If $d_i(v) = j$ in T_i , let $v_i, v_{i1}, \dots, v_{ij} = v$ be the vertices in the path from the root v_i to v . Then, $I_v = 1$ if and only if $\lambda_{v_{ij}} < \Lambda_i$ and $\lambda_{v_{ik}} > \lambda_{v_{ij}}$ for $k \in \{1, \dots, j - 1\}$. Since the λ_v s are independent $\text{Exp}(1)$ random variables,

$$E(I_v) = \int_0^{\Lambda_i} \prod_{k=1}^{j-1} P(\lambda_{v_{ik}} > x) e^{-x} dx = \int_0^{\Lambda_i} e^{-jx} dx = \frac{1 - e^{-j\Lambda_i}}{j}.$$

Thus, $\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} (1 - e^{-d_i(v)\Lambda_i})/d_i(v)$. Expanding $1/d_i(v)$ for arbitrary ‘good’ $v \in T_i$ gives

$$\frac{1}{d_i(v)} = \frac{1}{\mu^{-1} \ln n_i} - \frac{d_i(v) - \mu^{-1} \ln n_i}{\mu^{-2} \ln^2 n_i} + \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} + \mathcal{O}\left(\frac{|(d_i(v) - \mu^{-1} \ln n_i)^3|}{\ln^4 n_i}\right).$$

Recall from Section 1.2.2 that the number of ‘bad’ vertices in T_i , i.e. those that are not in the strip in (2.1), is $\mathcal{O}_{L^1}(n_i / \ln^3 n_i)$ and can thus be ignored. Hence,

$$\begin{aligned} \sum_{v \neq v_i} \frac{1}{d_i(v)} &= \sum_{\text{good } v \neq v_i} \frac{1}{d_i(v)} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^3 n_i}\right) \\ &= \frac{N_i}{\mu^{-1} \ln n_i} - \frac{\Upsilon(T_i) - \mu^{-1} N_i \ln n_i}{\mu^{-2} \ln^2 n_i} \\ &\quad + \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \end{aligned} \tag{2.3}$$

Now we prove that

$$\sum_{v \neq v_i} \frac{1}{d_i(v)} (e^{-d_i(v)} - e^{-(\mu^{-1} \ln n_i)\Lambda_i}) = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right), \tag{2.4}$$

which obviously implies that

$$\varphi(T_i, \Lambda_i) = (1 - e^{-(\mu^{-1} \ln n_i)\Lambda_i}) \sum_{v \neq v_i} \frac{1}{d_i(v)} + \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right). \tag{2.5}$$

For simpler calculations, we show the bound in (2.4) by considering $e^{-\Lambda_i \lfloor \mu^{-1} \ln n_i \rfloor}$ instead of $e^{-(\mu^{-1} \ln n_i)\Lambda_i}$. That we can do this follows from the fact that multiplying the Taylor estimate in (2.3) by $e^{-\Lambda_i \lfloor \mu^{-1} \ln n_i \rfloor}$ gives the same expression up to the error term $\mathcal{O}_{L^1}(n_i / \ln^{2.2} n_i)$ as multiplying by $e^{-(\mu^{-1} \ln n_i)\Lambda_i}$. For $j > 0$,

$$\begin{aligned} e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + j)\Lambda_i} &= e^{-\Lambda_i \lfloor \mu^{-1} \ln n_i \rfloor} + e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + j)\Lambda_i}, \\ e^{(-\lfloor \mu^{-1} \ln n_i \rfloor - j)\Lambda_i} &= e^{-\Lambda_i \lfloor \mu^{-1} \ln n_i \rfloor} + e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + j)\Lambda_i} (e^{-2j\Lambda_i} - e^{-j\Lambda_i}). \end{aligned}$$

Since we only have to consider the ‘good’ vertices, it is enough to show that

$$Q_1 + Q_2 = \mathcal{O}_{L^1}\left(\frac{n_i}{\ln^{2.2} n_i}\right), \tag{2.6}$$

where

$$\begin{aligned} Q_1 &:= \sum_{j=1}^{\lfloor \ln^{0.6} n_i \rfloor} \sum_{d_i(v)=j} e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + j)\Lambda_i} (1 - e^{-j\Lambda_i}) \frac{1}{\lfloor \mu^{-1} \ln n_i \rfloor - j}, \\ Q_2 &:= \sum_{j=1}^{\lfloor \ln^{0.6} n_i \rfloor} \sum_{d_i(v)=j} e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + j)\Lambda_i} (e^{-2j\Lambda_i} - e^{-j\Lambda_i}) \frac{1}{\lfloor \mu^{-1} \ln n_i \rfloor + j}. \end{aligned}$$

We have

$$\begin{aligned} Q_1 &\leq N_i e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + \ln^{0.6} n_i) \Lambda_i} (1 - e^{-\ln^{0.6} n_i \Lambda_i}) \frac{1}{\lfloor \mu^{-1} \ln n_i \rfloor - \ln^{0.6} n_i} \\ &= N_i \mathcal{O} \left(\frac{\ln^{0.6} n_i \Lambda_i}{\mu^{-1} \ln n_i} \right) e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + \ln^{0.6} n_i) \Lambda_i}, \end{aligned}$$

and, similarly,

$$Q_2 = N_i \mathcal{O} \left(\frac{\ln^{0.6} n_i \Lambda_i}{\mu^{-1} \ln n_i} \right) e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + \ln^{0.6} n_i) \Lambda_i}.$$

Since Λ_i is an $\text{Exp}(1/(L + 1))$ random variable, using integration by parts, we obtain

$$\begin{aligned} E(\Lambda_i e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + \ln^{0.6} n_i) \Lambda_i}) &= \int_0^\infty (L + 1) y e^{(-\lfloor \mu^{-1} \ln n_i \rfloor + \ln^{0.6} n_i) y} e^{-y(L+1)} dy \\ &= \frac{L + 1}{(\lfloor \mu^{-1} \ln n_i \rfloor - \ln^{0.6} n_i + L + 1)^2}. \end{aligned}$$

Thus, (2.6) holds and it follows that (2.5) is satisfied.

Now we show that (2.5) implies (2.2) in Lemma 2.1. We have

$$e^{-(\mu^{-1} \ln n_i) \Lambda_i} = \mathcal{O}_{L^1} \left(\frac{L}{\ln n_i} \right).$$

Hence,

$$e^{-(\mu^{-1} \ln n_i) \Lambda_i} \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} = \mathcal{O}_{L^1} \left(\frac{n_i}{\ln n_i} \right).$$

Recall from Section 1.2.2 that the number of ‘bad’ nodes in T_i is $\mathcal{O}_{L^1}(n_i / \ln^3 n_i)$ and that, for any constant r , there is a constant $C > 0$ such that the number of nodes with $d(v) \geq C \ln n$ is $\mathcal{O}_{L^1}(1/n^r)$. Using these facts, we obtain an obvious upper bound for the total path length, i.e. $|\Upsilon(T_i) - \mu^{-1} N_i \ln n_i| \leq N_i \ln^{0.6} n_i + \mathcal{O}_{L^1}(n_i / \ln n_i)$. Hence,

$$\frac{(\Upsilon(T_i) - \mu^{-1} N_i \ln n_i) e^{-(\mu^{-1} \ln n_i) \Lambda_i}}{\mu^{-2} \ln^2 n_i} = \mathcal{O}_{L^1} \left(\frac{n_i}{\ln^{2.2} n_i} \right),$$

and Lemma 2.1 follows.

Recall that $\psi(T_i, \Lambda_i) := \text{var}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$ and that $E_{n_i}(\cdot) := E(\cdot \mid n_i)$.

Lemma 2.2. For all vertices v_i with $d(v_i) = L$, conditioned on n_i ,

$$E_{n_i}(\psi(T_i, \Lambda_i)) = \mathcal{O} \left(\frac{n_i^2}{\ln^3 n_i} \right).$$

Proof. We say that a pair (v, w) , with $h(v) = j$ and $h(w) = k$, is ‘good’ if j and k satisfy

$$\mu^{-1} \ln n_i - \ln^{0.6} n_i \leq j, k \leq \mu^{-1} \ln n_i + \ln^{0.6} n_i,$$

and ‘bad’ otherwise. In analogy with [18] and [21], using the same indicator I_v as in the proof of Lemma 2.1, we show that, for a ‘good’ pair,

$$1 \geq \text{cov}(I_v, I_w) = \frac{1}{jk} e^{-(j+k-d)\Lambda_i} (1 - e^{-d\Lambda_i}) + \mathcal{O} \left(\frac{d}{\ln^3 n_i} \right) = \mathcal{O}_{L^1} \left(\frac{d}{\ln^3 n_i} \right). \tag{2.7}$$

(Compare this with [18, Equations (2.13)–(2.14)].) Since the number of ‘bad’ vertices is $\mathcal{O}_{L^1}(n_i/\ln^3 n_i)$, it follows that the number of ‘bad’ pairs is $\mathcal{O}_{L^1}(n_i^2/\ln^3 n_i)$. Hence, the sum of covariances for the ‘bad’ pairs is $\mathcal{O}(n_i^2/\ln^3 n_i)$. Thus,

$$E_{n_i}(\psi(T_i, \Lambda_i)) = E_{n_i}\left(\sum_{\text{good } (v,w) \in T_i} \text{cov}(I_v, I_w)\right) + \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right). \tag{2.8}$$

Recall that \mathcal{G}_j is the σ -field generated by the split vectors for all vertices v with $d(v) \leq j$. Let n_{i_v} denote the number of balls in the subtrees rooted at v for $v \in T_i$. From (1.10) we have, for v , where $d_i(v) = d$,

$$E_{n_i}(n_{i_v} \mid \mathcal{G}_{L+d}) \leq n_i \prod_{r=1}^d W_r + s_1 d.$$

Thus,

$$E_{n_i}(n_{i_v}) \leq n_i \prod_{r=1}^d E(W_r) + s_1 d = \frac{n_i}{b^d} + s_1 d.$$

Again, by using (1.10) we obtain

$$E_{n_i}(n_{i_v}^2 \mid \mathcal{G}_{L+d}) = n_i^2 \prod_{r=1}^d W_r^2 + \mathcal{O}\left(n_i d \prod_{r=1}^d W_r\right) + \mathcal{O}(d^2).$$

Thus,

$$E_{n_i}(n_{i_v}^2) \leq n_i^2 \prod_{r=1}^d E(W_r^2) + \mathcal{O}\left(\frac{n_i d}{b^d}\right) + \mathcal{O}(d^2). \tag{2.9}$$

Note that $E(W_r^2) < E(W_r) = 1/b$ since $W_r \in [0, 1]$ and W_r is nondegenerate. Hence, there is an $\varepsilon > 0$ such that the right-hand side of (2.9) is bounded by $n_i^2/(b + \varepsilon)^d + \mathcal{O}(n_i d/b^d) + \mathcal{O}(d^2)$. Hence, from (2.8), using (2.7) and (2.9),

$$E_{n_i}(\psi(T_i, \Lambda_i)) = \mathcal{O}\left(\sum_d \frac{n_i^2 b^d d}{(b + \varepsilon)^d \ln^3 n_i}\right) + \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right) = \mathcal{O}\left(\frac{n_i^2}{\ln^3 n_i}\right).$$

The estimate in Lemma 2.2 is used in the proof of the following result.

Lemma 2.3. *In a split tree T^n , let $v_i, 1 \leq i \leq b^L$, be the vertices at depth $L = \lfloor \beta \log_b \ln n \rfloor$, choosing $\beta > 1/(-\log_b E(V^2) - 1)$. Then*

$$X(T^n) = \sum_{i=1}^{b^L} \varphi(T_i, \Lambda_i) + o_p\left(\frac{n}{\ln^2 n}\right).$$

Proof. We write the number of records as $\{P^* + P_1 + \dots + P_{b^L}\}$, where P^* is the number of records with depth at most L and P_i is the number of records in the subtree T_i rooted at depth L , except for the root v_i . Let \mathcal{F}_L be the σ -field generated by $\{\lambda_v : d(v) \leq L\}$, and let \mathcal{F}_L^* be the σ -field generated by T^n and \mathcal{F}_L . We also note that $E(P_i \mid \mathcal{F}_L^*) = \varphi(T_i, \Lambda_i)$. By the same calculation as in [18, Equations (2.17)–(2.18)],

$$E\left(\left(X(T^n) - P^* - \sum_{i=1}^{b^L} \varphi(T_i, \Lambda_i)\right)^2\right) = \sum_{i=1}^{b^L} E\psi(T_i, \Lambda_i). \tag{2.10}$$

Obviously, for large enough k ,

$$\sum_{i \in \{1, \dots, b^L\}, n_i \leq nb^{-kL}} n_i \leq b^L \frac{n}{b^{kL}} = \mathcal{O}\left(\frac{n}{\ln^3 n}\right). \tag{2.11}$$

(By choosing k large enough in (2.11), the power of the logarithm can be taken to be arbitrarily large.) Lemma 2.2 and (2.11) give $\sum_{i=1}^{b^L} E_{n_i}(\psi(T_i, \Lambda_i)) = \mathcal{O}(\sum_{i=1}^{b^L} n_i^2 / \ln^3 n)$; compare with [18, Equation (2.20)]. The expected value of this sum is equal to the expected value of the left-hand side of (2.10). From the calculations in (2.9), for $i \in \{1, \dots, b^L\}$,

$$E(n_i^2) \leq n^2(E(V^2))^L + \mathcal{O}(nL). \tag{2.12}$$

Hence, choosing $\beta > 1/(-\log_b E(V^2) - 1)$ we obtain $\sum_{i=1}^{b^L} E(n_i^2) = o(n^2 / \ln n)$, and, thus, the left-hand side of (2.10) is $o(n^2 / \ln^4 n)$. Thus, Lemma 2.3 follows from the Markov inequality.

Applying Lemmas 2.1 and 2.3 we obtain, for $\beta > 1/(-\log_b E(V^2) - 1)$,

$$\begin{aligned} X(T^n) &= \sum_{i=1}^{b^L} \left(\frac{2N_i}{\mu^{-1} \ln n_i} - \frac{N_i e^{-(\mu^{-1} \ln n_i) \Lambda_i}}{\mu^{-1} \ln n_i} + \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} - \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \right) \\ &\quad + o_p\left(\frac{n}{\ln^2 n}\right), \end{aligned} \tag{2.13}$$

where we have used the fact that the Markov inequality gives $\mathcal{O}_{L^1}(n / \ln^{2.2} n) = o_p(n / \ln^2 n)$.

In [17, Corollary 2.2] we proved that

$$\sum_{i=1}^{b^L} \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} = \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.14}$$

We obtain, for $n_i \geq n/b^{kL}$,

$$E\left(\left| \frac{N_i}{\mu^{-1} \ln n_i} e^{-(\mu^{-1} \ln n_i) \Lambda_i} - \frac{N_i}{\mu^{-1} \ln n} e^{-(\mu^{-1} \ln n) \Lambda_i} \right|\right) = \mathcal{O}\left(\frac{L^2 n}{b^L \ln^3 n}\right); \tag{2.15}$$

compare with [18, p. 404].

Again, we use the bound in (2.11) for the $n_i < n/b^{kL}$ (for large enough k), so that we can ignore them in the sums in (2.13). Thus, by (2.14) and (2.15), with another application of the Markov inequality, the approximation in (2.13) can be simplified to

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2N_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} - \sum_{i=1}^{b^L} \frac{N_i e^{-(\mu^{-1} \ln n) \Lambda_i}}{\mu^{-1} \ln n} + \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.16}$$

(Compare this with [18, Equation (2.22)].)

By choosing large enough β , from (2.12) we can achieve

$$\sum_{i=1}^{b^L} E(n_i^2) = o\left(\frac{n^2}{\ln^k n}\right) \tag{2.17}$$

for arbitrarily large k . Applying (2.17), the variance result of N in Section 1.2.2, that is,

$\text{var}(N) = o(n^2)$, and supposing that Assumption 1.3 holds, Chebyshev’s inequality results in

$$\sum_{i=1}^{b^L} \frac{N_i}{\ln n_i} = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\ln n_i} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.18}$$

The third sum in (2.16) is treated similarly. For simplicity, we change the notation N_i , $1 \leq i \leq b^L$, to N_v , $d(v) = L$, and, similarly, for n_i , $1 \leq i \leq b^L$. Hence, from (2.16), for large enough β ,

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2\alpha n_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} - \sum_{i=1}^{b^L} \frac{\alpha n_i e^{-(\mu^{-1} \ln n)\Lambda_i}}{\mu^{-1} \ln n} + \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{2.19}$$

Lemma 2.4. *Let $L = \lfloor \beta \log_b \ln n \rfloor$ for some constant β . Then*

$$\sum_{i=1}^{b^L} n_i e^{-(\mu^{-1} \ln n)\Lambda_i} = \sum_{d(v) \leq L} n_v e^{-(\mu^{-1} \ln n)\lambda_v} + o_p\left(\frac{n}{\ln n}\right).$$

Thus, choosing $\beta > 1/(-\log_b E(V^2) - 1)$, from (2.19),

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2\alpha n_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} - \sum_{d(v) \leq L} \frac{\alpha n_v e^{-(\mu^{-1} \ln n)\lambda_v}}{\mu^{-1} \ln n} + \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right).$$

Proof. Recall that $m := \mu^{-1} \ln n$ and that Λ_i is the minimum of the $L + 1$ i.i.d. random variables λ_v , $v \in P(v_i)$, defined in Section 2.1. Thus, $e^{-m\Lambda_i}$ is the maximum. Now we define Λ_i^j as the j th smallest value in $\{\lambda_v, v \in P(v_i)\}$, so that $e^{-m\Lambda_i^j}$ is the j th maximum. Note, in particular, that $\Lambda_i^1 = \Lambda_i$. Choosing $a = 2 \ln m/m$ means that, for some i , the probability that at least $\lfloor \beta \rfloor + 1$ of the λ_v s, $v \in P(v_i)$, are less than a is $\mathcal{O}(b^L (La)^{\lfloor \beta \rfloor + 1}) = o(1)$. Thus, with probability tending to 1, there are at most $\lfloor \beta \rfloor$ values λ_v less than a in each $P(v_i)$, giving, for each i ,

$$0 \leq \sum_{v \in P(v_i)} e^{-m\lambda_v} - \sum_{j=1}^{\lfloor \beta \rfloor} e^{-m\Lambda_i^j} \leq (L - \lfloor \beta \rfloor) e^{-ma} = \frac{L - \lfloor \beta \rfloor}{m^2}.$$

Hence, using the fact that $n_v - sb^L \leq \sum_{\{i: v \in P(v_i)\}} n_i \leq n_v$,

$$\begin{aligned} \sum_{i=1}^{b^L} n_i \sum_{j=1}^{\lfloor \beta \rfloor} e^{-m\Lambda_i^j} &= \sum_{i=1}^{b^L} n_i \sum_{v \in P(v_i)} e^{-m\lambda_v} + o_p\left(\frac{n}{\ln n}\right) \\ &= \sum_{d(v) \leq L} e^{-m\lambda_v} \sum_{\{i: v \in P(v_i)\}} n_i + o_p\left(\frac{n}{\ln n}\right) \\ &= \sum_{d(v) \leq L} n_v e^{-m\lambda_v} + o_p\left(\frac{n}{\ln n}\right). \end{aligned}$$

Observing that the second smallest value Λ_i^2 in $\{i : v \in P(v_i)\}$ is at most x if at least two λ_v are at most x , and using the fact that the λ_v s are i.i.d., we calculate $P(\Lambda_i^2 \leq x)$ as

$$1 - P(\lambda_v > x)^L - L P(\lambda_v > x)^{L-1} P(\lambda_v \leq x) = 1 - e^{-Lx} - L e^{-(L-1)x} (1 - e^{-x}).$$

Hence, $E(e^{-m\Lambda_i^2})$ can be expressed as

$$\int_0^\infty e^{-mx} ((L - L^2)e^{-Lx} + L(L - 1)e^{-(L-1)x}) dx = \frac{L - L^2}{m + L} + \frac{L^2 - L}{m + L - 1} = \mathcal{O}\left(\frac{L^2}{m^2}\right),$$

implying that $\sum_{i=1}^{b^L} n_i \sum_{j=2}^{\lfloor \beta \rfloor} e^{-m\Lambda_i^j} = \mathcal{O}_{L^1}(nL^2/m^2)$. Thus, the Markov inequality gives

$$\sum_{i=1}^{b^L} n_i e^{-(\mu^{-1} \ln n)\Lambda_i} = \sum_{d(v) \leq L} n_v e^{-(\mu^{-1} \ln n)\lambda_v} + o_p\left(\frac{n}{\ln n}\right).$$

From Lemma 2.4 (where β is chosen large enough), by applying (2.18) and the total path length result in (1.5), we thus obtain

$$\begin{aligned} X(T^n) &= \sum_{d(v)=L} \frac{\alpha n_v}{\mu^{-1} \ln n_v} - \sum_{d(v) \leq L} \frac{\alpha n_v e^{-(\mu^{-1} \ln n)\lambda_v}}{\mu^{-1} \ln n} - \frac{\zeta n}{\mu^{-2} \ln^2 n} + \frac{\alpha n \sigma^2}{\ln^2 n} \\ &\quad + o_p\left(\frac{n}{\ln^2 n}\right). \end{aligned} \tag{2.20}$$

As in [18] and [21], our aim is to express $X(T^n)$ as a sum of triangular arrays. Recall that $\xi_v := mn_v e^{-m\lambda_v}/n$. Normalizing $X(T^n)$ gives, using (2.20),

$$\begin{aligned} &\frac{\mu^{-2} \ln^2 n}{\alpha n} \left(X(T^n) - \frac{\alpha n}{\mu^{-1} \ln n} - \frac{\alpha n \ln \ln n}{\mu^{-1} \ln^2 n} + \frac{\zeta n}{\mu^{-2} \ln^2 n} \right) \\ &= - \sum_{d(v) \leq L} \xi_v + \frac{\mu^{-2} \ln^2 n}{n} \sum_{d(v)=L} \frac{n_v}{\mu^{-1} \ln n_v} - \mu^{-1} \ln \ln n - \mu^{-1} \ln n \\ &\quad + \mu^{-2} \sigma^2 + o_p(1). \end{aligned} \tag{2.21}$$

Let

$$D := \frac{\mu^{-2} \ln^2 n}{n} \sum_{d(v)=L} \frac{n_v}{\mu^{-1} \ln n_v} - \mu^{-1} \ln \ln n - \mu^{-1} \ln n + \mu^{-2} \sigma^2 \tag{2.22}$$

and $\xi'_i = -D/n$. Then

$$\begin{aligned} &\frac{\mu^{-2} \ln^2 n}{\alpha n} \left(X(T^n) - \frac{\alpha n}{\mu^{-1} \ln n} - \frac{\alpha n \ln \ln n}{\mu^{-1} \ln^2 n} + \frac{\zeta n}{\mu^{-2} \ln^2 n} \right) \\ &= - \sum_{d(v) \leq L} \xi_v - \sum_{i=1}^n \xi'_i + o_p(1). \end{aligned} \tag{2.23}$$

As in [18], since the n_v s in (2.21) are not independent, $\{\xi_v\} \cup \{\xi'_i\}$ is not a triangular array. Recall that Ω_L is the σ -field generated by $\{n_v, d(v) \leq L\}$. Conditioned on Ω_L , $\{\xi_v\} \cup \{\xi'_i\}$ is a triangular array with ξ'_i conditioned on Ω_L deterministic.

2.3. Applying a limit theorem for sums of triangular arrays

2.3.1. *Theorem 2.1, which proves Theorem 1.1.* As in [18] and [21], the proof of Theorem 1.1 will be completed by a classical theorem for convergence of sums of triangular arrays to infinitely divisible distributions; see, e.g. [24, Theorem 15.28]. For the sake of independence, we intend to condition on the n_v s in the sums in (2.23). We show that, conditioned on the n_v s, we obtain convergence in distribution for the normalized $X(T^n)$ to a random variable W with an infinitely divisible distribution, which is not dependent on the n_v s we conditioned on. Then it follows in the same way as in [18] that, also unconditioned, the normalized $X(T^n)$ converges in distribution to W . The main theorem, Theorem 1.1, is proven by Theorem 2.1 below.

Theorem 2.1. *For any constant $c > 0$, large enough β , and $L = \lfloor \beta \log_b \ln n \rfloor$, the following statements hold: as $n \rightarrow \infty$,*

- (i) $\sup_{v, d(v) \leq L} P(\xi_v > x \mid \Omega_L) \xrightarrow{\text{a.s.}} 0$ for every $x > 0$,
- (ii) $\Delta_1 := \sum_{d(v) \leq L} P(\xi_v > x \mid \Omega_L) \xrightarrow{P} \nu(x, \infty) := \frac{\mu^{-1}}{x}$ for every $x > 0$,
- (iii) $\Delta_2 := \sum_{d(v) \leq L} E(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) - \frac{\mu^{-2} \ln^2 n}{n} \sum_{d(v)=L} \frac{n_v}{\mu^{-1} \ln n_v} + \mu^{-1} \ln \ln n$
 $+ \mu^{-1} \ln n - \mu^{-2} \sigma^2$
 $\xrightarrow{P} -\mu^{-1} \ln \mu^{-1} + \mu^{-1} - \mu^{-2} \sigma^2 - \frac{\sigma^2 - \mu^2}{2\mu^2} + \mu^{-1} \ln c$,
- (iv) $\Delta_3 := \sum_{d(v) \leq L} \text{var}(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) \xrightarrow{P} \mu^{-1} c$.

Compare this theorem with [18, Theorem 2.5]. Before proving Theorem 2.1 we show how it proves Theorem 1.1. Recall D from (2.22). We apply [24, Theorem 15.28], with the constants

$$a = 0 \quad \text{and} \quad b = -\mu^{-1} \ln \mu^{-1} + \mu^{-1} - \mu^{-2} \sigma^2 - \frac{\sigma^2 - \mu^2}{2\mu^2},$$

to $\sum_{d(v) \leq L} \xi_v + \sum_{i=1}^n \xi'_i$ conditioned on Ω_L with $\xi'_i = -D/n$ deterministic. The constants a and b are those that occur in (1.8). Note that $D/n \rightarrow 0$; thus, because of (i), conditioned on Ω_L , $\{\xi_v\} \cup \{\xi'_i\}$ is a null array.

We define $S(n) := \sum_{d(v) \leq L} \xi_v + \sum_{i=1}^n \xi'_i$. Let $\nu(x) =: 1 - \nu(x, \infty)$. From (ii) we have that $d\nu(x)/dx = \mu^{-1}/x^2$; hence,

$$\int_0^c x^2 d\nu(x) = \int_0^c \mu^{-1} dx = \mu^{-1} c \quad \text{and} \quad \int_c^1 x d\nu(x) = \int_c^1 \frac{\mu^{-1}}{x} dx = -\mu^{-1} \ln c.$$

Thus, the right-hand sides of (iii) and (iv) are $b - \int_c^1 x d\nu(x)$ and $\int_0^c x^2 d\nu(x)$, respectively, where b is the constant just defined. In analogy to the binary search tree in [18], we can also apply [24, Theorem 15.28] to general split trees, implying that, conditioned on Ω_L , $S(n) \xrightarrow{D} W$ as $n \rightarrow \infty$, where W has an infinitely divisible distribution (in particular, a weakly 1-stable distribution) with characteristic function as in (1.9). Since the conditioning does not affect the distribution of W , it follows that, also unconditioned, $S(n) \xrightarrow{D} W$; for a formal proof of

this fact, see [18, p. 408]. Thus, unconditioned, the normalized $X(T^n)$ in (2.21) converges in distribution to $-W$. Thus, the proof of Theorem 1.1 for $X_v(T)$ is completed. Now it simply follows, by the same type of argument as for the binary search tree [18, p. 409], that the result holds for $X_e(T)$ too.

Remark 2.1. Theorem 15.28 of [24] in fact requires usual convergence, i.e. standard pointwise convergence of sequences with no probability involved, whereas the convergence in Theorem 2.1 is in the probabilistic sense. However, in [18, pp. 408–409] we proved in two ways (firstly by using subsequences and secondly by using Skorokhod’s coupling theorem) that convergence in probability is actually sufficient for $S(n) \xrightarrow{D} W$ to hold. By analogy, these proofs also work for general split trees.

The idea of the proof of Theorem 2.1 is, as for the binary search tree [18, Theorem 2.5], to use Chebyshev’s inequality. In [18] we frequently used the fact that the sum $\sum_{r=1}^k \ln U_r$, where the U_r are uniform $U(0, 1)$ random variables, is distributed as a $-\Gamma(k, 1)$ random variable. For general split trees, the exponential renewal function $U(t)$ defined in (1.12) is fundamental.

2.3.2. *Lemmas for the proof of Theorem 2.1.* Recall that Ω_j is the σ -field generated by $\{n_v, d(v) \leq j\}$ and that \mathcal{G}_j is the σ -field generated by $\{W_{r,v}, r \in \{1, 2, \dots, j\}, d(v) = j\}$. Recall that $m := \mu^{-1} \ln n$, and write

$$\hat{n}_v := n \prod_{r=1}^k W_{r,v}, \quad \hat{\xi}_v := \frac{m \hat{n}_v}{n} e^{-m \lambda_v}. \tag{2.24}$$

Note that \mathcal{G}_j is equivalent to the σ -field generated by $\{\hat{n}_v : d(v) \leq j\}$.

We now present four crucial lemmas, which we use to prove Theorem 2.1.

Lemma 2.5. *For any constant $c > 0$, large enough β , and $L = \lfloor \beta \log_b \ln n \rfloor$, the following statements hold: as $n \rightarrow \infty$,*

$$\begin{aligned} \sum_{d(v) \leq L} P(\xi_v > x \mid \Omega_L) &= \sum_{d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L) + o_p(1), \\ \sum_{d(v) \leq L} E(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) &= \sum_{d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o_p(1), \\ \sum_{d(v)=L} \frac{n_v}{\mu^{-1} \ln n_v} &= \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{\hat{n}_v \ln(\hat{n}_v/n)}{\mu^{-1} \ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right), \\ \sum_{d(v) \leq L} \text{var}(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) &= \sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o_p(1). \end{aligned}$$

For simplicity, we sometimes use the shorthand notation

$$\Phi_L := \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{\hat{n}_v \ln(\hat{n}_v/n)}{\mu^{-1} \ln^2 n}, \tag{2.25}$$

$$R_1 := \sum_{d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L), \tag{2.26}$$

$$R_2 := \sum_{d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) - \frac{\mu^{-2} \ln^2 n}{n} \Phi_L, \tag{2.27}$$

$$R_3 := \sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L). \tag{2.28}$$

Lemma 2.6. *For any constant $c > 0$, large enough β , and $L = \lfloor \beta \log_b \ln n \rfloor$, the following statements hold: as $n \rightarrow \infty$,*

$$E(R_1) = \frac{\mu^{-1}}{x} + o(1) = v(x, \infty) + o(1) \text{ for every } x > 0,$$

$$E(R_2) = -\mu^{-1} \ln n - \mu^{-1} \ln \ln n + \mu^{-1} - \mu^{-1} \ln \mu^{-1} + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1),$$

$$E(R_3) = \mu^{-1} c + o(1).$$

Let $l := \lfloor \log_b \ln n / 2 \rfloor$, and, for short, write

$$\begin{aligned} S_1 &:= \sum_{l \leq d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L), \\ S_2 &:= \sum_{l \leq d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) - \frac{\mu^{-2} \ln^2 n}{n} \Phi_L, \\ S_3 &:= \sum_{l \leq d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L). \end{aligned} \tag{2.29}$$

Lemma 2.7. *For any constant $c > 0$, large enough β , $L = \lfloor \beta \log_b \ln n \rfloor$, and $l = \lfloor \log_b \ln n / 2 \rfloor$, it holds that, for $i \in \{1, 2, 3\}$, as $n \rightarrow \infty$,*

$$\text{var}(E(S_i \mid \mathcal{G}_l)) \rightarrow 0. \tag{2.30}$$

Lemma 2.8. *For any constant $c > 0$, large enough β , $L = \lfloor \beta \log_b \ln n \rfloor$, and $l = \lfloor \log_b \ln n / 2 \rfloor$, it holds that, for $i \in \{1, 2, 3\}$, as $n \rightarrow \infty$,*

$$E(\text{var}(S_i \mid \mathcal{G}_l)) \rightarrow 0. \tag{2.31}$$

Before proving these lemmas we show how they lead to the proof of Theorem 2.1.

2.3.3. *Proof of Theorem 2.1.* For any $x > 0$ and v with $d(v) \leq L$, we have

$$\begin{aligned} P(\xi_v > x \mid \Omega_L) &= P\left(e^{-m\lambda_v} > \frac{nx}{mn_v} \mid \Omega_L\right) \\ &= P\left(\lambda_v < \frac{1}{m} \ln \frac{mn_v}{nx} \mid \Omega_L\right) \\ &= 1 - \exp\left(-\frac{1}{m} \ln^+ \frac{mn_v}{nx}\right). \end{aligned} \tag{2.32}$$

Thus, for every $x > 0$,

$$P(\xi_v > x \mid \Omega_L) \leq \frac{1}{m} \ln^+ \frac{mn_v}{nx} \leq \frac{1}{m} \ln^+ \frac{m}{x} \rightarrow 0, \tag{2.33}$$

which proves (i).

Recall the definitions of R_1 , R_2 , and R_3 in (2.26), (2.27), and (2.28). Lemma 2.5 shows that the left-hand sides of Theorem 2.1(ii), (iii), and (iv) have the following asymptotics:

$$\begin{aligned} \Delta_1 &= R_1 + o_p(1), \\ \Delta_2 &= R_2 + \mu^{-1} \ln \ln n + \mu^{-1} \ln n - \mu^{-2} \sigma^2 + o_p(1) := \hat{R}_2 + o_p(1), \\ \Delta_3 &= R_3 + o_p(1). \end{aligned}$$

Lemma 2.6 shows that the expected values of R_1 , \hat{R}_2 , and R_3 converge to the right-hand sides of Theorem 2.1(ii), (iii), and (iv).

We complete the proof of Theorem 2.1 by showing that

$$\text{var}(R_1) \rightarrow 0 \quad \text{for every } x > 0, \quad \text{var}(R_2) \rightarrow 0, \quad \text{and} \quad \text{var}(R_3) \rightarrow 0. \tag{2.34}$$

Then, by Chebyshev’s inequality, Theorem 2.1(ii), (iii), and (iv) follow. Thus, it remains to show how (2.34) follows from Lemma 2.6 and Lemma 2.7. By using (2.33), we easily obtain

$$\begin{aligned} \sum_{d(v) \leq L} \mathbb{P}(\hat{\xi}_v > x \mid \mathcal{G}_L) &= \sum_{l \leq d(v) \leq L} \mathbb{P}(\hat{\xi}_v > x \mid \mathcal{G}_L) + o(1), \\ \sum_{d(v) \leq L} \mathbb{E}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) &= \sum_{l \leq d(v) \leq L} \mathbb{E}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o(1), \\ \sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) &= \sum_{l \leq d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o(1). \end{aligned}$$

Hence,

$$R_1 = S_1 + o(1), \quad R_2 = S_2 + o(1), \quad R_3 = S_3 + o(1). \tag{2.35}$$

To show (2.34), we use a variance formula that is easy to establish (see, e.g. [15, Exercise 10.17-2]) $\text{var}(X) = \mathbb{E}(\text{var}(X \mid \mathcal{G})) + \text{var}(\mathbb{E}(X \mid \mathcal{G}))$, where X is a random variable and \mathcal{G} is a sub- σ -field. Consequently, by applying the variance formula, from Lemma 2.7 and Lemma 2.8, we obtain, as $n \rightarrow \infty$, for $i \in \{1, 2, 3\}$,

$$\text{var}(S_i) = \mathbb{E}(\text{var}(S_i \mid \mathcal{G}_i)) + \text{var}(\mathbb{E}(S_i \mid \mathcal{G}_i)) \rightarrow 0,$$

and, thus, (2.34) follows from (2.35). Hence, Lemmas 2.5–2.8 prove Theorem 2.1, and, thus, also Theorem 1.1.

2.3.4. *Proofs of Lemmas 2.5–2.8.* Finally, we present the proofs of Lemmas 2.5–2.8.

Proof of Lemma 2.5. Since a binomial (k, p) random variable has expected value kp and variance $kp(1 - p)$, by applying (1.10), the Chebyshev inequality results in

$$\mathbb{P}\left(\left|n_v - n \prod_{r=1}^k W_{r,v}\right| > n^{0.6} \mid \Omega_L\right) \leq \frac{1}{n^{0.19}}. \tag{2.36}$$

This motivates the notation $\hat{n}_v := n \prod_{r=1}^k W_{r,v}$ in (2.24). Also, recall $\hat{\xi}_v$ from (2.24) and the σ -field \mathcal{G}_j from Section 2.1. By using (2.32) and (2.33), we obtain (compare with [18, Equation (2.50)]),

$$\sum_{d(v) \leq L} \mathbb{P}(\xi_v > x \mid \Omega_L) = \sum_{k=1}^L \sum_{d(v)=k} \frac{1}{m} \ln^+\left(\frac{mn_v}{nx}\right) \left(1 + \mathcal{O}\left(\frac{\ln m}{m}\right)\right), \tag{2.37}$$

and, similarly,

$$\sum_{d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L) = \sum_{k=1}^L \sum_{d(v)=k} \frac{1}{m} \ln^+ \left(\frac{m\hat{n}_v}{nx} \right) \left(1 + o \left(\frac{\ln m}{m} \right) \right). \tag{2.38}$$

Using (2.37), (2.38), and (2.36), we obtain

$$\sum_{d(v) \leq L} P(\xi_v > x \mid \Omega_L) = \sum_{d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L) + o_p(1). \tag{2.39}$$

Similarly, (2.36) implies that

$$\sum_{d(v) \leq L} E(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) = \sum_{d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o_p(1),$$

where

$$\sum_{d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) = \sum_{d(v) \leq L} \frac{m\hat{n}_v}{n(m+1)} \exp \left(-\frac{m+1}{m} \ln^+ \left(\frac{m\hat{n}_v}{nc} \right) \right). \tag{2.40}$$

(Compare with [18, Equations (2.56)–(2.57)], where $\sum_{d(v) \leq L} E(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L)$ is estimated.)

Using (2.36) (compare with [18, Equations (2.62)–(2.63)]), we obtain

$$\sum_{d(v)=L} \frac{n_v}{\mu^{-1} \ln n_v} = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{\hat{n}_v \ln(\hat{n}_v/n)}{\mu^{-1} \ln^2 n} + o_p \left(\frac{n}{\ln^2 n} \right).$$

By applying (2.36) and using similar calculations as in (2.37)–(2.39), we obtain

$$\sum_{d(v) \leq L} \text{var}(\xi_v \mathbf{1}[\xi_v \leq c] \mid \Omega_L) = \sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) + o_p(1),$$

where (compare with [18, Equation (2.65)])

$$\sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) = \sum_{d(v) \leq L} \frac{m^2 \hat{n}_v^2}{2mn^2} \exp \left(-\frac{2m+1}{m} \ln^+ \left(\frac{m\hat{n}_v}{nc} \right) \right) + o(1).$$

Proof of Lemma 2.6. Recall that $Y_k = -\sum_{r=1}^k \ln W_r$ and that $R_1 = \sum_{d(v) \leq L} P(\hat{\xi}_v > x \mid \mathcal{G}_L)$. As in the calculations in [18, Equation (2.51)], from (2.38) and using integration by parts, we obtain

$$\begin{aligned} E(R_1) &= (1 + o(1)) \sum_{k=1}^L b^k E \left(\frac{\ln(m/x) - Y_k}{m} \mathbf{1} \left[Y_k \leq \ln \left(\frac{m}{x} \right) \right] \right) \\ &= (1 + o(1)) \frac{1}{m} \int_0^{\ln(m/x)} \sum_{k=1}^L b^k P(Y_k \leq t) dt. \end{aligned} \tag{2.41}$$

We want to show that

$$\frac{1}{m} \int_0^{\ln(m/x)} \sum_{k=L+1}^{\infty} b^k \mathbb{P}(Y_k \leq t) dt = o(1). \tag{2.42}$$

To show this, we use large deviations. Choose an arbitrary $s > 0$. By applying the Markov inequality and using the fact that the $W_{r,v}$, $r \in \{1, \dots, k\}$, are i.i.d., we obtain

$$\mathbb{P}(Y_k \leq t) = \mathbb{P}(-Y_k \geq -t) = \mathbb{P}(e^{-sY_k} \geq e^{-st}) \leq (\mathbb{E}(V^s))^k e^{st}. \tag{2.43}$$

In the definition of $L = \lfloor \beta \log_b \ln n \rfloor$, the constant β can be chosen arbitrarily large. It is enough to show that $(\ln m/m) \sum_{k=L+1}^{\infty} b^k \mathbb{P}(Y_k \leq \ln(m/x))$ is $o(1)$ for proving (2.42). Choosing $s > 1$, we obtain $\mathbb{E}(V^s) < \mathbb{E}(V) = 1/b$. Thus, we can find a $\delta > 0$ such that $\mathbb{E}(V^s) \leq 1/b^{1+\delta}$. By applying this fact together with (2.43), we obtain

$$\sum_{k=L+1}^{\infty} b^k \mathbb{P}\left(Y_k \leq \ln\left(\frac{m}{x}\right)\right) \leq \sum_{k=L+1}^{\infty} \frac{b^k}{b^{k+\delta k}} \frac{m^s}{x^s} = \mathcal{O}(m^{-\delta\beta} m^s). \tag{2.44}$$

Thus, choosing $\beta > (s - 1)/\delta$ in L gives (2.42). Now it follows from the asymptotics of $U(t)$ in (1.13) that the principal term in (2.41) has the following asymptotics:

$$\frac{1}{m} \int_0^{\ln(m/x)} U(t) dt + o(1) = \frac{\mu^{-1} + o(1)}{m} \int_0^{\ln(m/x)} e^t dt + o(1) = \nu(x, \infty) + o(1). \tag{2.45}$$

Hence, $\mathbb{E}(R_1) = \nu(x, \infty) + o(1)$. Similar to [18, Equation (2.58)], using (2.40), we obtain

$$\mathbb{E}\left(\sum_{d(v) \leq L} \mathbb{E}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L)\right) := E_1 + E_2, \tag{2.46}$$

where

$$\begin{aligned} E_1 &= \mathbb{E} \sum_{d(v) \leq L} b^k \frac{m}{m+1} e^{-Y_k} \exp\left(-\frac{m+1}{m} \left(\ln\left(\frac{m}{c}\right) - Y_k\right)\right) \mathbf{1}\left[Y_k \leq \ln\left(\frac{m}{c}\right)\right], \\ E_2 &= \mathbb{E} \sum_{d(v) \leq L} b^k \frac{m}{m+1} e^{-Y_k} \mathbf{1}\left[Y_k > \ln\left(\frac{m}{c}\right)\right]. \end{aligned} \tag{2.47}$$

Using integration by parts, we obtain

$$\begin{aligned} E_1 &= \frac{m}{m+1} \exp\left(-\frac{m+1}{m} \ln\left(\frac{m}{c}\right)\right) \left(\left| \sum_{k=1}^L b^k e^{t/m} \mathbb{P}(Y_k \leq t) \right|_0^{\ln(m/c)} \right. \\ &\quad \left. - \int_0^{\ln(m/c)} \sum_{k=1}^L \frac{b^k}{m} e^{t/m} \mathbb{P}(Y_k \leq t) dt \right), \end{aligned} \tag{2.48}$$

which is asymptotic to $\mu^{-1} + o(1)$ in view of (1.13) and (2.44). By similar calculations as in (2.48) we obtain

$$\begin{aligned} E_2 &= \frac{m}{m+1} L - \frac{m}{m+1} \int_0^{\ln(m/c)} \sum_{k=1}^L b^k e^{-t} d\mathbb{P}(Y_k \leq t) \\ &= L - \mu^{-1} - \frac{m}{m+1} \int_0^{\ln(m/c)} \sum_{k=1}^L b^k e^{-t} \mathbb{P}(Y_k \leq t) dt + o(1). \end{aligned} \tag{2.49}$$

From (2.44) with $\beta > s/\delta$, it follows that

$$\begin{aligned} \int_0^{\ln(m/c)} \sum_{k=1}^L b^k e^{-t} \mathbf{P}(Y_k \leq t) dt &= \int_0^{\ln(m/c)} e^{-t} U(t) dt + o(1) \\ &= \int_0^{\ln(m/c)} e^{-t} (U(t) - \mu^{-1} e^t) dt + \mu^{-1} \ln\left(\frac{m}{c}\right) + o(1). \end{aligned} \tag{2.50}$$

Applying the solution of $W(x) := \int_0^x e^{-t} (U(t) - \mu^{-1} e^t) dt$ in (1.14), from (2.49) we obtain $E_2 = L - \mu^{-1} \ln(m/c) - (\sigma^2 - \mu^2)/2\mu^2 + o(1)$.

Recalling (2.46) and applying the approximations of E_1 and E_2 , we obtain

$$\mathbf{E}\left(\sum_{d(v) \leq L} \mathbf{E}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L)\right) = L + \mu^{-1} - \mu^{-1} \ln\left(\frac{m}{c}\right) - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1),$$

which has the following asymptotics:

$$K := L + \mu^{-1} - \mu^{-1} \ln \ln n - \mu^{-1} \ln \mu^{-1} + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1). \tag{2.51}$$

By the definition of \hat{n}_v in (2.24),

$$\begin{aligned} \Phi_L &:= \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{\hat{n}_v \ln(\hat{n}_v/n)}{\mu^{-1} \ln^2 n} \\ &= \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{n \prod_{r=1}^L W_{r,v} \sum_{r=1}^L \ln W_{r,v}}{\mu^{-1} \ln^2 n}. \end{aligned} \tag{2.52}$$

Hence, using the definition of μ in (1.1), we obtain $\mathbf{E}(\Phi_L) = n/\mu^{-1} \ln n + nL/\mu^{-2} \ln^2 n$. Recalling R_2 from (2.27) and K in (2.51), we obtain $\mathbf{E}(R_2) = K - \mu^{-1} \ln n - L$.

Recall that $R_3 = \sum_{d(v) \leq L} \text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L)$. Using integration by parts and (2.43) (choosing $1 < s < 2$), we obtain, by similar calculations as in (2.41)–(2.45), $\mathbf{E}(R_3) = \mu^{-1} c + o(1)$. (Compare with [18, Equations (2.66)–(2.69)].) For a complete proof of this fact, we refer the reader to [16].

Proof of Lemma 2.7. We show only (2.30) for $i = 2$. Then (2.30) for $i = 1$ and $i = 3$ are shown by similar but simpler calculations. For a complete proof, we refer the reader to [16]. Also, compare with the similar calculations in [18, Proof of Lemma 2.7].

For a given vertex $v_i \in T$ with $d(v_i) = l$, there are at most b^{j-l} choices of v at depth j with ancestor v_i . Recall that $Y_{j,v} := -\sum_{r=1}^j \ln W_{r,v}$. For v with $d(v) = j$, we also write $Z_{j-l,v} := Y_{j,v} - Y_{l,v_i} = -\sum_{r=l+1}^j \ln W_{r,v}$.

Recall the definition of S_2 from (2.29) and the definition of Φ_L from (2.25). Glancing at the calculations in (2.46)–(2.47), we obtain $\mathbf{E}(\sum_{l \leq d(v) \leq L} \mathbf{E}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l) = F_1 + F_2$,

where

$$F_1 := E\left(\sum_{l \leq d(v) \leq L} \frac{m}{m+1} \exp(-Y_{l,v_i} - Z_{j-l,v}) \exp\left(-\frac{m+1}{m}(\ln m - \ln c - Y_{l,v_i} - Z_{j-l,v})\right) \times \mathbf{1}\left[Y_{l,v_i} + Z_{j-l,v} \leq \ln\left(\frac{m}{c}\right)\right] \mid \mathcal{G}_l\right),$$

$$F_2 := E\left(\sum_{l \leq d(v) \leq L} \frac{m}{m+1} \exp(-Y_{l,v_i} - Z_{j-l,v}) \mathbf{1}\left[Y_{l,v_i} + Z_{j-l,v} > \ln\left(\frac{m}{c}\right)\right] \mid \mathcal{G}_l\right).$$

Then, by similar calculations as in (2.48),

$$F_1 = \exp\left(-\frac{m+1}{m} \left(\ln\left(\frac{m}{c}\right)\right)\right) \frac{m}{m+1} \sum_{i=1}^{b^l} \int_0^{\ln(m/c) - Y_{l,v_i}} \sum_{j=l+1}^L b^{j-l} e^{t/m} dP(Z_{j-l,v} \leq t)$$

$$= \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^l W_{r,v_i} + o(1)$$

$$= \mu^{-1} + o(1). \tag{2.53}$$

By similar calculations as in (2.48)–(2.50), we obtain

$$F_2 = \sum_{i=1}^{b^l} \prod_{r=1}^l W_{r,v_i} \left((L-l) - \int_0^{\ln(m/c) - Y_{l,v_i}} \frac{m}{m+1} \sum_{j=l+1}^L b^{j-l} e^{-t} dP(Z_{j-l,v} \leq t) \right) + o(1)$$

$$= \sum_{i=1}^{b^l} \prod_{r=1}^l W_{r,v_i} \left(L-l - \mu^{-1} \ln\left(\frac{m}{c}\right) - \frac{\sigma^2 - \mu^2}{2\mu^2} - \mu^{-1} \sum_{r=1}^l \ln W_{r,v_i} \right) + o(1)$$

$$= L-l - \mu^{-1} \ln\left(\frac{m}{c}\right) - \frac{\sigma^2 - \mu^2}{2\mu^2} - \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^l W_{r,v_i} \sum_{r=1}^l \ln W_{r,v_i} + o(1). \tag{2.54}$$

Thus, by applying the approximations of F_1 in (2.53) and F_2 in (2.54), we obtain

$$E\left(\sum_{l \leq d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l\right) = \mu^{-1} + L-l - \mu^{-1} \ln\left(\frac{m}{c}\right) - \frac{\sigma^2 - \mu^2}{2\mu^2}$$

$$- \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^l W_{r,v_i} \sum_{r=1}^l \ln W_{r,v_i} + o(1). \tag{2.55}$$

Let v_i be a vertex at depth l , and let v be a vertex at depth L . Similarly as in (2.52) (compare with [18, Equations (2.73)–(2.74)]), we obtain

$$E(\Phi_L \mid \mathcal{G}_l) = \frac{n}{\mu^{-1} \ln n} + \frac{n(L-l)}{\mu^{-2} \ln^2 n} - \sum_{i=1}^{b^l} \frac{n \prod_{r=1}^l W_{r,v_i} \sum_{r=1}^l \ln W_{r,v_i}}{\mu^{-1} \ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \tag{2.56}$$

From (2.55) and (2.56), $\text{var}(E(S_2 \mid \mathcal{G}_l))$ is $o(1)$.

Proof of Lemma 2.8. The calculations are similar to those of [18, Proof of Lemma 2.8]. We show only (2.31) for $i = 2$ and $i = 3$; then by similar calculations we can show (2.31) for $i = 1$. For a complete proof, we refer the reader to [16].

First we consider

$$\begin{aligned} & \text{var} \left(\sum_{l \leq d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l \right) \\ &= \sum_{\substack{l \leq d(v) \leq L \\ l \leq d(w) \leq L}} \text{cov}(E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L), E(\hat{\xi}_w \mathbf{1}[\hat{\xi}_w \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l). \end{aligned} \tag{2.57}$$

We can suppose that the closest ancestor u for v and w is at depth $d \geq l$, since the other terms are just 0 because of independence. For $d \geq l$,

$$\begin{aligned} & E(\text{cov}(E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L), E(\hat{\xi}_w \mathbf{1}[\hat{\xi}_w \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l)) \\ & \leq E(E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) E(\hat{\xi}_w \mathbf{1}[\hat{\xi}_w \leq c] \mid \mathcal{G}_L)). \end{aligned}$$

For a vertex v with $d(v) = j$,

$$E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) = \frac{m \hat{n}_v}{n(m+1)} \exp \left(-\frac{m+1}{m} \ln^+ \left(\frac{m \hat{n}_v}{nc} \right) \right) \leq \frac{\hat{n}_v}{n} = \prod_{r=1}^j W_{r,v}.$$

Denote by (v_u, w_u) a pair of vertices with closest ancestor u . Consider one such pair (v_u, w_u) , and let $d(u) = d$, $d(v) = j$, and $d(w) = k$. Since $E(W_{r,v}^2) < 1/b^{1+\delta}$ for some $\delta > 0$, it follows that

$$\begin{aligned} E(E(\hat{\xi}_{v_u} \mathbf{1}[\hat{\xi}_{v_u} \leq c] \mid \mathcal{G}_L) E(\hat{\xi}_{w_u} \mathbf{1}[\hat{\xi}_{w_u} \leq c] \mid \mathcal{G}_L)) & \leq C_1 E \prod_{r=1}^d W_{r,u}^2 b^{-(j-d)-(k-d)} \\ & \leq C_1 \left(\frac{1}{b^{1+\delta}} \right)^d b^{-(j-d)-(k-d)}, \end{aligned} \tag{2.58}$$

where C_1 is a constant depending on $E(\hat{W}_{u,v} \hat{W}_{u,w})$, $\hat{W}_{u,v}$ is the component in the split vector of vertex u that corresponds to the child v_u of u , and $\hat{W}_{u,w}$ has an analogous definition. Thus, by using (2.57)–(2.58), letting $v, w: v \wedge w = u$ denote that the vertices v, w have closest ancestor u , we find that

$$E \left(\text{var} \left(\sum_{l \leq d(v) \leq L} E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \mid \mathcal{G}_l \right) \right)$$

is bounded by

$$\begin{aligned} & \sum_{d=l}^L \sum_{\{u: d(u)=d\}} \sum_{\{v,w \in T_u: v \wedge w = u\}} E(E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) E(\hat{\xi}_w \mathbf{1}[\hat{\xi}_w \leq c] \mid \mathcal{G}_L)) \\ & \leq C_1 \sum_{d=l}^L b^{-\delta d} \sum_{j=d}^L b^{j-d-(j-d)} \sum_{k=d}^L b^{k-d-(k-d)} \\ & \leq C_2 L^2 b^{-\delta l} \\ & \rightarrow 0, \end{aligned} \tag{2.59}$$

where C_2 is a constant. (Compare with the calculations in [18, Equation (2.82)].) Similarly,

we show that

$$E\left(\text{var}\left(\frac{\mu^{-2} \ln^2 n}{n} \Phi_L \mid \mathcal{G}_l\right)\right) \rightarrow 0.$$

By applying the conditional Hölder inequality we deduce that $E(\text{var}(S_2 \mid \mathcal{G}_l))$ is $o(1)$.

We can now easily show (2.31) for $i = 3$ by the same argument as in [18], i.e. we observe that

$$\text{var}(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \leq E(\hat{\xi}_v^2 \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L) \leq c E(\hat{\xi}_v \mathbf{1}[\hat{\xi}_v \leq c] \mid \mathcal{G}_L),$$

and, thus, (2.31) for $i = 3$ follows from (2.58) by calculations similar to those in (2.59).

Appendix A. Proof of Lemma 1.1

Recall that Ω_L is the σ -field generated by $\{n(v), d(v) \leq L\}$. First, (1.4) gives

$$E\left(\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \mid \Omega_L\right) = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i}. \tag{A.1}$$

Note that, conditioned on Ω_L , the summands $\Upsilon(T_i)$, $i \in \{1, \dots, b^L\}$, are independent. By applying the Cauchy–Schwarz inequality, and using the facts that $E(N^2) = \mathcal{O}(n^2)$ (see, e.g. [17, Theorem 2.1]) and $E(D_k^2) = \mathcal{O}(\ln^2 n)$ for all k (see e.g. [17, Theorem 2.3]), we obtain

$$\begin{aligned} \text{var}\left(\sum_{i=1}^{b^L} \Upsilon(T_i) \mid \Omega_L\right) &= \sum_{i=1}^{b^L} \text{var}(\Upsilon(T_i) \mid \Omega_L) \\ &\leq \sum_{i=1}^{b^L} E(\Upsilon(T_i)^2 \mid \Omega_L) \\ &= \sum_{i=1}^{b^L} \mathcal{O}(n_i^2 \ln^2 n_i). \end{aligned} \tag{A.2}$$

For a large enough constant β , by taking expectations in (A.2) and applying (2.17), we obtain

$$E\left(\text{var}\left(\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \mid \Omega_L\right)\right) = o\left(\frac{n^2}{\ln^4 n}\right). \tag{A.3}$$

Using (A.1), (A.2), and applying (A.3), the Chebyshev inequality results in, conditioned on Ω_L ,

$$\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i} + o_p\left(\frac{n}{\ln^2 n}\right). \tag{A.4}$$

Let $k > 0$ be a fixed constant, and assume that n_i is at least n/b^{kL} ; by Taylor’s expansion we obtain

$$\frac{1}{\ln^2 n_i} = \frac{1}{\ln^2 n} + \mathcal{O}\left(\frac{\ln \ln n}{\ln^3 n}\right). \tag{A.5}$$

By applying (2.11) and using (A.5), we deduce that

$$\sum_{i=1}^{b^L} \frac{o(n_i)}{\ln^2 n_i} = o\left(\sum_{i=1}^{b^L} \frac{n_i}{\ln^2 n_i}\right) + o\left(\frac{n}{\ln^2 n}\right) = o\left(\frac{n}{\ln^2 n}\right). \tag{A.6}$$

By applying [4, Corollary 5.1], which states that $r(n)$ converges to a constant ζ , it follows from (2.11), (A.5), and (A.6) that

$$\sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{\zeta n_i}{\mu^{-2} \ln^2 n_i} + \sum_{i=1}^{b^L} \frac{o(n_i)}{\ln^2 n_i} = \frac{\zeta n}{\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \quad (\text{A.7})$$

Thus, the lemma follows from (A.4) and (A.7).

Acknowledgement

I gratefully acknowledge the help and support of Professor Svante Janson, for introducing me to this problem area and for helpful discussions and guidance.

References

- [1] ASMUSSEN, S. (1987). *Applied Probability and Queues*. John Wiley, Chichester.
- [2] BELJAEV, JU. K. AND MAKSIMOV, V. M. (1963). Analytical properties of a generating function for the number of renewals. *Theoret. Prob. Appl.* **8**, 108–112.
- [3] BOURDON, J. (2001). Size and path length of Patricia tries: dynamical sources context. *Random Structures Algorithms* **19**, 289–315.
- [4] BROUTIN, N. AND HOLMGREN, C. (2011). The total path length of split trees. Submitted.
- [5] CALIEBE, A. (2003). Symmetric fixed points of a smoothing transformation. *Adv. Appl. Prob.* **35**, 377–394.
- [6] CALIEBE, A. AND RÖSLER, U. (2003). Fixed points with finite variance of a smoothing transformation. *Stoch. Process. Appl.* **107**, 105–129.
- [7] DEVROYE, L. (1998). Universal limit laws for depths in random trees. *SIAM J. Comput.* **28**, 409–432.
- [8] DEVROYE, L. (2005). Applications of Stein's method in the analysis of random binary search trees. In *Stein's Method and Applications* (Lecture Notes Ser. Inst. Math. Sci. Natl. Univ. Sing. **5**), Singapore University Press, pp. 247–297.
- [9] DRMOTA, M., IKSANOV, A., MÖHLE, M. AND RÖSLER, U. (2009). A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures Algorithms* **34**, 319–336.
- [10] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. John Wiley, New York.
- [11] FILL, J. A. AND JANSON, S. (2002). Quicksort asymptotics. *J. Algorithms* **44**, 4–28.
- [12] FILL, J. A. AND KAPUR, N. (2004). Limiting distributions for additive functionals on Catalan trees. *Theoret. Comput. Sci.* **326**, 69–102.
- [13] FLAJOLET, P., ROUX, M. AND VALLÉE, B. (2010). Digital trees and memoryless sources: from arithmetics to analysis. To appear in *Discrete Math. Theoret. Comput. Sci.*
- [14] GUT, A. (1988). *Stopped Random Walks*. Springer, New York.
- [15] GUT, A. (2005). *Probability: A Graduate Course*. Springer, New York.
- [16] HOLMGREN, C. (2010). A weakly 1-stable limiting distribution for the number of random records and cuttings in split trees. Preprint. Available at <http://arxiv.org/abs/1005.4590v1>.
- [17] HOLMGREN, C. (2010). Novel characteristics of split trees by use of renewal theory. Preprint. Available at <http://arxiv.org/abs/1005.4594v1>.
- [18] HOLMGREN, C. (2010). Random records and cuttings in binary search trees. *Combinatorics Prob. Comput.* **19**, 391–424.
- [19] IKSANOV, A. AND MEINERS, M. (2010). Exponential rate of almost-sure convergence of intrinsic martingales in supercritical branching random walks. *J. Appl. Prob.* **47**, 513–525.
- [20] IKSANOV, A. AND MÖHLE, M. (2007). A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Prob.* **12**, 28–35.
- [21] JANSON, S. (2004). Random records and cuttings in complete binary trees. In *Mathematics and Computer Science. III*, Birkhäuser, Basel, pp. 241–253.
- [22] JANSON, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29**, 139–179.
- [23] JANSON, S., ŁUCZAK, T. AND RUCINSKI, A. (2000). *Random Graphs*. Wiley-Interscience, New York.
- [24] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
- [25] MAHMOUD, H. (1986). On the average internal path length of m -ary search trees. *Acta Informatica* **23**, 111–117.

- [26] MAHMOUD, H. AND PITTEL, B. (1989). Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10**, 52–75.
- [27] MEIR, A. AND MOON, J. W. (1970). Cutting down random trees. *J. Austral. Math. Soc.* **11**, 313–324.
- [28] NEININGER, R. AND RÜSCHENDORF, L. (1999). On the internal path length of d -dimensional quad trees. *Random Structures Algorithms* **15**, 25–41.
- [29] PANHOLZER, A. (2003). Non-crossing trees revisited: cutting down and spanning subtrees. In *Discrete Random Walks* (Paris, 2003), Association of Discrete Mathematics and Theoretical Computer Science, Nancy, pp. 265–276.
- [30] PANHOLZER, A. (2006). Cutting down very simple trees. *Quaest. Math.* **29**, 211–227.
- [31] RÖSLER, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica* **29**, 238–261.