A Weakly 1-Stable Limiting Distribution for the Number of Random Records and Cuttings in Split Trees

Cecilia Holmgren

Abstract

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 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 [10], which is one specific case of split trees. Other important examples of split trees include m-ary search trees, quadtrees, medians of $(2^k + 1)$-trees, simplex trees, tries and digital search trees.

1 Introduction

1.1 Preliminaries

We study the number of records in random split trees which were introduced by Devroye [3]. As shown by Janson [14], 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 $\lambda_v$ attached to it, and assume that these values are i.i.d. with a continuous distribution. We say that the value $\lambda_v$ 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 one 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 $\lambda_v$’s are important, so the distribution of $\lambda_v$ does not matter, i.e., one can choose any continuous distribution for $\lambda_v$.

The same random variables appear when we consider cuttings of the tree $T$ as introduced by Meir and Moon [19] with the following definition. Make
a random cut by choosing one vertex [respectively edge] at random. Delete this vertex [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 [respectively only the root is left]. Then the total (random) number of cuts made is $X_v(T)$ [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 [14, 13].

In [14] the asymptotic distributions for the number of cuts (or the number of records) are found for random trees that can be constructed as conditioned Galton–Watson trees, e.g., labelled trees and random binary trees. There the proof relies on the fact that the method of moments can be used.

For the deterministic (non random) complete binary tree it is, however, not possible to use the method of moments. To deal with this Janson [13] introduced another strategy, which is to approximate $X_v(T)$ by a sum of independent random variables derived from $\lambda_v$, and then apply a classical limit theorem for triangular arrays, see e.g., [16, Theorem 15.28]. We recently showed that Janson’s approach could also be applied to the random binary search tree [10].

In this paper we consider all types of (random) split trees defined by Devroye [3]; the binary search tree that we consider in [10] is one example of such trees. Some other important examples of split trees are $m$-ary search trees, quadtrees, median of $(2^k + 1)$-trees, simplex trees, tries and digital search trees. The split trees belong to the family of so-called log $n$ trees, that are trees with height (maximal depth) $a.a.s. \mathcal{O}(\log n)$. (For the notation $a.a.s.$ see [15].) These have similar properties to the deterministic complete binary tree with height $\lfloor \log_2 n \rfloor$ considered in [13]. In the complete binary tree (with high probability) most vertices are close to $\lfloor \log_2 n \rfloor$ (the height of the tree). In split trees on the other hand (with high probability) most vertices are close to depth $\sim c \ln n$, where $c$ is a constant (it is natural to use the $e$-logarithm); for the binary search tree that we investigated in [10] this depth is $\sim 2 \ln n$ (e.g., [4]). Here by the use of renewal theory we extend the methods used in [10] for the specific case of the binary search tree to show that also for split trees in general it is possible to apply a limit theorem, see e.g., [16, Theorem 15.28], for convergence of sums of triangular arrays to infinitely divisible distributions to determine the asymptotic distribution of $X_v(T)$.

The split tree generating algorithm:

The formal, comprehensive “split tree generating algorithm” is as follows with the following introductory notation, see [3] and [11]. 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 that are numbered $1, 2, \ldots, b$). The split tree
All internal vertices have $s_0 = 1$ balls
Note that $s_1 = 0$.
All leaves have between $1$ and $s = 3$ balls.

Figure 1: This figure illustrates a split tree with parameters $b = 4$, $s = 3$, $s_0 = 1$ and $s_1 = 0$.

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. There is also a so-called vertex capacity, $s > 0$, which means that each node can hold at most $s$ balls. Each vertex $v$ of $S_b$ is given an independent copy of the so-called random split vector $V = (V_1, V_2, \ldots, V_b)$ of probabilities, where $\sum_i V_i = 1$, $V_i \geq 0$. There are also two other parameters: $s_0, s_1$ (related to the parameter $s$) that occur in the algorithm below; see Figure 1 and Figure 2 where two examples of split trees are illustrated. Let $n_v$ denote the total number of balls that the vertices in the subtree rooted at vertex $v$ hold together, and $C_v$ be the number of balls that are held by $v$ itself. We say that a vertex $v$ is a leaf in a split tree if the node itself holds at least one ball but no descendants of $v$ hold any balls. An equivalent definition of a leaf is to say that $v$ is 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.

Below there is a description of the algorithm which determines how the \( n \) balls are distributed over the vertices. Initially there are no balls, i.e., \( C_v = 0 \) for each vertex \( v \). Choose an independent copy \( \mathcal{V}_v \) of \( \mathcal{V} \) for every vertex \( v \in S_b \). Add balls one by one to the root by the following recursive procedure for adding a ball to the subtree rooted at \( v \).

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.

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.

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 \( \mathcal{V}_v = (V_1, V_2, \ldots, V_b) \), and then using the algorithm described in steps 1, 2 and 3 applied to the subtree rooted at the selected child.

From 3 it follows that the integers \( s_0 \) and \( s_1 \) have to satisfy the inequality

\[
0 \leq s_0 \leq s, \quad 0 \leq bs_1 \leq s + 1 - s_0.
\]

We can assume that the components \( V_i \) of the split vector \( \mathcal{V} \) are identically distributed. If this were not the case they can anyway be made identically distributed by using a random permutation as explained in 3. Let \( V \) be a random variable with this distribution. This gives (because \( \sum_i V_i = 1 \)) that \( \mathbb{E}(V) = \frac{1}{b} \). We use the notation \( T^n \) to denote a split tree with \( n \) balls. However, note that even given the fact that the split tree has \( n \) balls, the number of nodes \( N \), is still a random number. The only parameters that are important in this work (and in general these parameters are the important ones for most results concerning split trees) are the cardinality \( n \), the branch factor \( b \) and the split vector \( \mathcal{V} \); this is illustrated in Section 1.4.1. In a binary search tree \( b = 2 \), the split vector \( \mathcal{V} = (V_1, V_2) \) is distributed as \( (U, 1 - U) \) where \( U \) is a uniform \( U(0, 1) \) random variable. For the binary search tree the number of balls \( n \) is the same number as the number of vertices \( N \); this is not true for split trees in general.
1.2 Some Important Facts and Results for Split Trees

1.2.1 Results Concerning Depth Analysis

In [3, Theorem 1] Devroye presents a strong law and a central limit law for the depth $D_n$ of the last inserted ball in a split tree with $n$ balls and split vector $V$. Recall that $V$ is distributed as the identically distributed components in the split vector. Let

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

If $P(V = 1) = 0$ and $P(V = 0) < 1$, then

$$
\frac{D_n}{\ln n} \xrightarrow{d} \mu^{-1},
$$

and

$$
\frac{\mathbb{E}(D_n)}{\ln n} \rightarrow \mu^{-1}.
$$

Furthermore, if $\sigma > 0$, then

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

where $N(0, 1)$ denotes the standard Normal distribution and $\xrightarrow{d}$ denotes convergence in distribution. Assuming that $\sigma > 0$ is equivalent to assuming that $V$ is not monoatomic, i.e., it is not the case that $V = \frac{1}{b}$.

Let $D_k$ be the depth of the $k$-th ball. In [11, Theorem 2.3] by using the same assumptions for $V$ as Devroye uses for proving (2) and (4) we also show results concerning the variance of depths, i.e., for all $\frac{n}{\ln n} \leq k \leq n$,

$$
\frac{\text{Var}(D_k)}{\ln n} \rightarrow \sigma^2 \mu^{-3}.
$$

1.2.2 Results Concerning the Number of Nodes

(A1). In this work we assume as in Section 1.2.1 that $P(V = 1) = 0$, and as in [11] for simplicity we also assume that $P(V = 0) = 0$ and that $-\ln V$ is non-lattice.
The non-lattice assumption we do since we use renewal theory for sums depending on the distribution of \(-\ln V\), and in renewal theory it often becomes necessary to distinguish between lattice and non-lattice distributions. Tries and digital search trees are special forms of split trees with a random permutation of deterministic components \((p_1, p_2, \ldots, p_b)\) and therefore 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 \(p_1 = p_2 = \cdots = p_b = \frac{1}{b}\)) does \(-\ln V\) have a lattice distribution. By assuming that (A1) holds we show in [11, Theorem 2.1] that there is a constant \(\alpha\) depending on the type of split tree such that for the random number of nodes \(N\) we have that
\[
E(N) = \alpha n + o(n); \quad (5)
\]
and
\[
\text{Var}(N) = o(n^2). \quad (6)
\]
Let \(d(v)\) denote the depth of a node. In [11, Theorem 2.2] we show that the expected number of nodes in a tree with \(n\) balls, where \(d(v) \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n\) or \(d(v) \geq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n\), for some arbitrary \(\epsilon > 0\), is \(O\left(\frac{n}{\ln^3 n}\right)\), for any constant \(k\). In this paper we use in particular that this is \(O\left(\frac{n}{\ln^3 n}\right)\). In [11, Remark 4.3] we also note 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 \(O\left(\frac{1}{n^r}\right)\), hence, we can bound the number of vertices with ”large” depths with very small error terms.

### 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 \(T\) (distances to the root). 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 probably the sum of all depths of balls in \(T\), which we denote by \(\Psi(T)\).

From (3) it follows that
\[
E\left(\Psi(T^n)\right) = \mu^{-1} n \ln n + nq(n), \quad (7)
\]
where \(q(n) = o(\ln n)\) is a function that depends on the type of split tree. By using (3) and (5) we easily show in [11] that
\[
E\left(\Upsilon(T^n)\right) = \mu^{-1} \alpha n \ln n + nr(n), \quad (8)
\]
where \( \alpha \) is the constant that occurs in (5) and \( r(n) = o(\ln n) \) is a function that depends on the type of split tree.

**(A2).** Assume that the functions \( q(n) \) in (7) converge to some constant \( \varsigma \).

In [20] there is an analogous assumption. Examples of split trees where it is shown that \( q(n) \) converges to a constant are binary search trees (e.g. [7]), random \( m \)-ary search trees [17], quad trees [20] and the random median of a \((2k+1)\)-tree [21], tries and Patricia tries [2].

**(A3).** Assume that the result in (5) can be improved such that

\[
\mathbb{E}(N) = \alpha n + f(n),
\]

where \( f(n) = \mathcal{O}\left(\frac{n}{\ln^{1+\epsilon} n}\right) \).

Stronger second order terms of the size have previously been shown to hold e.g., for \( m \)-ary search trees [18], for these \( f(n) \) in assumption (A3) is \( o(\sqrt{n}) \) when \( m \leq 26 \) and is \( \mathcal{O}\left(n^{1-\epsilon}\right) \) when \( m \geq 27 \). Further, as described in Section 1.2.2 tries are special cases of split trees which are not as random as other types of split trees. Flajolet and Vallée (personal communication) have recently shown that also for most tries (as long as \( -\ln V \) is not too close to being lattice) assumption (A3) holds.

In [11, Theorem 5.1] by assuming (A2) and (A3) we show that \( r(n) \) in (8) converges to some constant \( \zeta \). In [11, Theorem 5.2] by applying [11, Theorem 5.1] we show the following result, which we will apply in the proof of the main theorem below: Let \( L = \lfloor \beta \log_b \ln n \rfloor \) for some large constant \( \beta \), then

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

where \( \zeta \) is the constant that \( r(n) \) converges to.

### 1.3 The Main Theorem

The main theorem of this study is presented below:

**Theorem 1.1.** Let \( n \to \infty \), and suppose that assumptions (A1)-(A3) hold. Then

\[
\left( X_v(T^n) - C_n \right) \sqrt{\frac{\alpha n}{\mu^{-2 \ln^2 n}}} \xrightarrow{d} -W,
\]

where \( \alpha \) is the constant that occurs in (5) and \( r(n) = o(\ln n) \) is a function that depends on the type of split tree.
where
\[ C_n := \frac{\alpha n}{\mu^{-1} \ln n} + \frac{\alpha n \ln n}{\mu^{-1} \ln^2 n} - \frac{\zeta n}{\mu^{-1} \ln^2 n}, \] (11)
for the constant \( \zeta \) in (9), and where \( W \) has an infinitely divisible distribution. More precisely \( W \) has a weakly 1-stable distribution, with characteristic function
\[ E(e^{itW}) = \exp \left( -\frac{\mu^{-1}}{2} \pi |t| + it \left( C - \mu^{-1} \ln |t| \right) \right), \] (12)
where \( \mu \) is the constant in (1) and \( \alpha \) is the constant in (5) and \( C \) is a constant which is defined in (15) below. The same result holds for \( X_{en(T^n)} \).

Remark 1.1. Even if we only have \( E(N) = \alpha n + o(n) \) as in (5) (i.e., ignoring the assumptions (A2)–(A3) the normalized \( X_{en(T^n)} \) (or \( X_{e(T^n)} \)) ought to still converge to a weakly 1-stable distribution with characteristic function as in (12) for some constant \( C \). However, in this case \( C_n \) in (10) ought to be
\[ C_n := 2 \frac{E(N)}{\mu^{-1} \ln n} - 2E \left( \sum_{d(v)=L} \frac{E(N_v | n_v) \ln(n_v)}{\mu^{-1} \ln^2 n} \right) - \frac{E(N)L}{\mu^{-1} \ln^2 n} + \frac{\alpha n \ln n}{\mu^{-1} \ln^2 n} - E \left( \sum_{i=1}^{b_L} \frac{\gamma(T_i)}{\mu^{-2} \ln^2 n_i} |n_i| \right), \]
where \( \gamma(T_i) \) is the total path length for the nodes of the subtrees \( T_i \) rooted at depth \( L \).

The class of \( \alpha \)-stable distributions are included in the larger class of infinitely divisible distributions. The general formula for the characteristic function of an infinitely divisible distribution is
\[ \exp \left( itb - \frac{a^2 t^2}{2} + \int_{-\infty}^{\infty} (e^{itx} - 1 - itxI[|x| < 1])d\nu(x) \right), \] (13)
for constants \( a \geq 0, b \in \mathbb{R} \) and \( \nu \) is the so called Lévy measure. The characteristic function in (13) of a 1-stable distribution (i.e., \( \alpha = 1 \)) can be simplified to
\[ \exp \left( i\beta t - c|t| \left( 1 + i\beta \frac{2}{\pi} \text{sign}(t) \ln |t| \right) \right), \]
for constants \( c > 0, \beta \in [-1, 1] \) and \( d \in \mathbb{R} \). If the Lévy measure \( \nu \) in (13) satisfies \( \frac{\nu}{dx} = \frac{c_+}{|x|^{\alpha+\tau}} \) on \( \mathbb{R}_\pm \), for \( \alpha \in (0, 2) \) and constants \( c_\pm \) the corresponding
infinitely divisible distribution is weakly $\alpha$-stable. The most well-known 1-stable distribution is the Cauchy distribution. However, in contrast to the distribution of $W$ in Theorem 1.1 (which is weakly 1-stable), the Cauchy distribution is strictly 1-stable and symmetric. The random variable $W$ in Theorem 1.1 has support on $(-\infty, \infty)$, and has a heavy tailed distribution. As for other random variables with $\alpha$-stable distributions where $\alpha < 2$ the variance of $W$ is infinite. Also since $\alpha \leq 1$ the expected value of $W$ is not defined. For further information about stable distributions, see e.g., [6, Section XVII.3].

Remark 1.2. In the proof of Theorem 1.1 we get

$$E\left(e^{itW}\right) = \exp \left(it \left( C + \mu^{-1}(\gamma - 1) \right) + \int_0^\infty \left(e^{itx} - 1 - itx1[x < 1]\right)d\nu(x)\right),$$

(14)

where $C$ is the constant in (12), $\gamma$ is the Euler constant and the Lévy measure $\nu$ is supported on $(0, \infty)$ and has density

$$\frac{d\nu}{dx} = \frac{\mu^{-1}}{x^2}.$$

Thus, we see that $W$ has a weakly 1-stable distribution. The constant $C$ can be expressed as

$$C = -\mu^{-1} \ln \mu^{-1} + 2\mu^{-1} - \mu^{-2} - \mu^{-1} \gamma - \frac{\sigma^2 - \mu^2}{2\mu^2},$$

(15)

where $\mu$ and $\sigma^2$ are the constants in (1). We can simplify the expression in (14) to get (12) above.

Remark 1.3. We note in analogy with [13] and [10] that most records occur close to the depth where most vertices are, i.e., $\sim \mu^{-1} \ln n$ for split trees. Also in analogy with [13] and [10], 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. For random trees [10], $E(X_v(T)) = E\left(\sum_{v\neq \sigma} \frac{1}{d(v)}\right)$ (where $\sigma$ is the root) and $E(X_v(T)) = E\left(\sum_{v} \frac{1}{d(v)+1}\right)$. Thus, as we noted for the specific case of the binary search tree [10, Remark 1.3] also for all other split trees

$$E(X_v(T^n)) - E(X_v(T^n)) = E\left(\sum_{v\neq \sigma} \frac{1}{d(v)(d(v)+1)}\right) - 1 \sim C_1 \frac{an}{\log^2 n}.$$
for some constant $C_1 > 0$, while there is no similar difference in the limit distribution, see Theorem 1.1 above. As in [10], this behaviour suggests that it is impossible to use the method of moments to find the record distribution for split trees as one could do for the conditioned Galton-Watson trees in [13]. In [10] we instead used methods similar to those that Janson used for the complete binary tree in [13]. In this paper we generalize the proofs in [10] to consider general split trees.

**Remark 1.5.** Most likely the method that is used here should 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 [5, Theorem 1.1] and [12, Theorem 1.1]. However, the methods used for the recursive tree in [5, 12] differ completely from our methods. The advantage with studying split trees compared to the whole class of log $n$ trees is that there is a common definition that describe 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 $(s_1, \ldots, s_1)$ plus a multinomial vector $(n - s_0 - bs_1, V_1, \ldots, V_b)$. Thus, conditioning on the random $V$-vector that belongs to the root, the subtrees rooted at the children have cardinalities close to $nV_1, \ldots, nV_b$. This is often used in applications of random binary search trees. In particular we used this frequently in [10].

Conditioning on the split vectors, $n_v$ at depth $d$, is in the stochastic sense bounded from above by

$$n_v \leq \text{Binomial}(n, \prod_{r=1}^{d} W_{r,v}) + s_1d, \quad (16)$$

and bounded from below by

$$n_v > \text{Binomial}(n, \prod_{r=1}^{d} W_{r,v}) - sd, \quad (17)$$
where $W_{r,v}, r \in \{1, \ldots, d\}$, 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 [3] and [11]. This means in particular that $W_{r,v} \overset{d}{=} V$. An application of the Chebyshev inequality gives that $n_v$ for $v$ at depth $d$ is close to

$$M^n_v := nW_{1,v}W_{2,v} \ldots W_{d,v},$$

(18)

see [11]. Since the $n_v$’s (conditioned on the split vectors) for all $v$ at the same depth are identically distributed, we sometimes skip the vertex index of $W_{r,v}$ in (16) and just write $W_r$.

### 1.4.2 Results Obtained by Using Renewal Theory

In [11] we introduce renewal theory in the context of split trees, and in this study we use this theory frequently for the proof of the Main Theorem, i.e., Theorem 1.1 below.

For each vertex $v$, where $W_{r,v} \overset{d}{=} V$ 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 skip the vertex index and just write $Y_k$, since for vertices $v$ on the same level $k$ the $Y_{k,v}$’s are identically distributed. This is the corresponding notation, as the one we use in [10] for the specific case of the binary search tree, where we define $Y_k := -\sum_{r=1}^{k} \ln U_r$, where $U_r$ are uniform $U(0,1)$ random variables. Recall from (18) in Section 1.4.1 that the subtree size $n_v$ for a vertex $v$ at depth $k$ is close to $M^n_v$ and note that

$$M^n_v := nW_{1,v}W_{2,v} \ldots W_{k,v} = ne^{-Y_k}.$$

Recall that in a binary search tree, the split vector is distributed as $(U, 1-U)$ where $U$ is uniform $U(0,1)$ random variable. For the binary search tree, the sum $\sum_{r=1}^{k} \ln U_r$ is distributed as a $-\Gamma(k, 1)$ random variable. For general split trees we do not know the common distribution function of $Y_k$, instead we use renewal theory. (For an introduction to renewal theory, see e.g., [3] Chapter II or [1].) We define the renewal function

$$U(t) = \sum_{k=1}^{\infty} b^k P(Y_k \leq t) = \sum_{k=1}^{\infty} F_k,$$

(19)

and also denote $F(t) := F_1(t) = bP(-\ln W_{r,v} \leq t)$, which in contrast to standard renewal theory is not a probability measure. For $U(t)$ we obtain the following renewal equation

$$U(t) = F(t) + \sum_{k=1}^{\infty} (F_k \ast F)(t) = F(t) + (U \ast F)(t).$$
Recall the definitions of the constants $\mu$ and $\sigma$ in [1]. In [11, Lemma 3.1] we show the following result which is fundamental for the proof of Theorem 1.1. Let $t \to \infty$, then the renewal function $U(t)$ in (19) has the solution

$$U(t) = (\mu^{-1} + o(1))e^t.$$  

(20)

In [11] we also define

$$W(x) := \int_0^x e^{-t}(U(t) - \mu^{-1}e^t)dt,$$

and in [11, Corollary 3.2], we show that

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

(21)

2 Proofs

2.1 Notation

Most of our notation are similar to the ones that we use in [10], 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. Let $\{x\} = x - \lfloor x \rfloor$ be the fractional part of a real number $x$. We treat the case $X_v(T_n)$ in Theorem 1.1 in detail and then indicate why the same result holds for $X_e(T_n)$ too.

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 write $d$ for equality in distribution.

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 $n \to \infty$.

We say that, $Y_n = O_L(a_n)$ if $a_n$ is a positive number and $Y_n$ is a random variable such that $(\mathbb{E}(Y_n^p))^{\frac{1}{p}} \leq C a_n$ for some constant $C$.

We sometimes use the notation $m = \mu^{-1}\ln n$. For simplicity in the proofs below we write $\ln n$ when we mean $\max\{1, \ln n\}$.

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 $\theta$, i.e., the density function $f(x) = e^{-x}/\theta$. We can without loss of generality assume that the labels $\lambda_v$ 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., distance to the root.

Recall that $V$ is a random variable distributed as the identically distributed components in the split vector $\mathcal{V} = (V_1, \ldots, V_n)$. Also recall that for each vertex $v$ we let $Y_{k,v} := -\sum_{r=1}^{k} \ln W_{r,v}$, where $W_{r,v} \overset{d}{=} V$ 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 (or depth), we sometimes skip the vertex index and just write $Y_k$. Recall from (19) that we define the renewal function $U(t) := \sum_{k=1}^{\infty} b^k \mathbf{P}(Y_k \leq t)$.

Let $\Lambda_{v_i}$ be the minimum of $\lambda_v$ along the path $P(v_i) = \sigma_i, \ldots, v_i$, from the root $\sigma$ of $T$ to $v_i$, $1 \leq i \leq b^L$, where $v_i$ are the vertices at depth $L = \lfloor \beta \log_b \ln n \rfloor$ for some constant $\beta$. Thus, the definition of $\Lambda_{v_i}$ and the assumption $\lambda_v \overset{d}{=} \exp(1)$ give $\Lambda_{v_i} \overset{d}{=} \exp(1 + \frac{i}{T^L})$.

For simplicity we write $T_i := T_{v_i}$, $n_i := n_{v_i}$, $\Lambda_i := \Lambda_{v_i}$, and $\Lambda_i := \Lambda_{v_i}$. We also let $T_{iv}$ denote a subtree of $T_i$ rooted at $v$ (note that $T_{iv}$ is $T_v$ for $v \in T_i$). Let $n_{iv}$ denote the number of balls in $T_{iv}$.

We write $d_i(v) := d(v) - L$ (i.e., the depth in the subtree $T_i$, $i \in \{1, \ldots, b^L\}$, 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 otherwise it is bad. In particular a vertex $v \in T_i$ is "good" if $$\mu^{-1} \ln n_i - \ln 0.6 n_i \leq d_i(v) \leq \mu^{-1} \ln n_i + \ln 0.6 n_i,$$
and otherwise it is bad.

We define $\varphi(T_i, \Lambda_i) := \mathbf{E}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$ (the conditional expected value of $X(T_i)_{\Lambda_i}$ given the tree $T_i$ and $\Lambda_i$). (We can think of $X(T_i)_{\Lambda_i}$ as $X(T_i) - 1$ conditioned on the root label $\lambda_{v_i} = \lambda_i$.) Similarly we let $\psi(T_i, \Lambda_i) := \mathbf{Var}(X(T_i)_{\Lambda_i} \mid T_i, \Lambda_i)$ for vertices $v_i$, $1 \leq i \leq b^L$ (the conditional variance of $X(T_i)_{\Lambda_i}$ given the tree $T_i$ and $\Lambda_i$).

The conditional expected value of a random variable $Z$ given the subtree size $n_i$ of $T_i$ is denoted by $\mathbf{E}_{n_i}(Z) := \mathbf{E}(Z \mid n_i)$.

We write $\xi_v := \frac{n_v \mu^{-1} \ln n}{n} \cdot e^{-\lambda_v \mu^{-1} \ln n}$, which is used in the later part of the proof when we consider triangular arrays.

We use the notation $\Omega_L$ for the $\sigma$-field generated by $\{n_v, d(v) \leq L\}$. Finally, we write $\mathcal{G}_j$ as the $\sigma$-field generated by the $\mathcal{V}$ vectors for all vertices $v$ with $d(v) \leq j$. Equivalently, this is the $\sigma$-field generated by $\{W_{r,v}, r \in \{1, 2, \ldots, j\}\}$, for all vertices $v$ with $d(v) = j$. In particular we use that the subtree sizes $\{n_v, d(v) \leq L\}$ up to small errors are determined by the $\sigma$-field $\mathcal{G}_L$; this follows because of the representation of subtree sizes in Section 1.4.1.
2.2 Expressing the normalized number of records as a sum of triangular arrays

Recall from Section 2.1 that we define $\varphi(T_i, \Lambda_i) := E(X(T_i) | T_i, \Lambda_i)$, where $T_i$ is the subtree rooted at $v_i$ at depth $L$ and $\Lambda_i$ is the minimum of $\lambda_v$ in the path from $v_i$ to the root $\sigma$ of $T$.

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

$$\varphi(T_i, \Lambda_i) = \frac{N_i}{\mu^{-1} \ln n_i} (1 - e^{-(\mu^{-1} \ln n_i)\Lambda_i}) - \frac{\Upsilon(T_i) - \mu^{-1} N_i \ln n_i}{\mu^{-2} \ln^2 n_i} + \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} + O_L \left( \frac{n_i}{\ln^2 n_i} \right),$$

(23)

where $\Upsilon(T_i)$ is the total path length of the tree $T_i$, and the good vertices $v \in T_i$ are those with $d_i(v)$ satisfying (22).

Proof. Let for each vertex $v \in T_i$, $I_v$ be the indicator that $\lambda_v$ is the minimum value given $T_i$ and $\Lambda_i$. We get

$$\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} E(I_v).$$

If $d_i(v) = j$ in $T_i$, let $v_i, v_i1, ..., v_ij = v$ be the vertices in the path from the root $v_i$ to $v$. Then, $I_v = 1$, if and only if, $\lambda_{v_i} < \Lambda_i$ and $\lambda_{v_k} > \lambda_{v_i}$ for $k \in \{1, ..., j - 1\}$. Since the $\lambda_v$’s (for all vertices $v$ in $T_i$) are independent $\text{Exp}(1)$ random variables

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

(24)

Thus,

$$\varphi(T_i, \Lambda_i) = \sum_{v \neq v_i} \frac{1 - e^{-d_i(v)\Lambda_i}}{d_i(v)}.$$

Expanding $\frac{1}{d_i(v)}$ for arbitrary good $v \in T_i$ gives

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

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 (22), is $O_L \left( \frac{n_i}{\ln n_i} \right)$ and can thus be ignored. Thus,
summing over all nodes \( v \in T_i \) gives

\[
\sum_{v \neq v_i} \frac{1}{d_i(v)} = \sum_{\text{good } v \neq v_i} \frac{1}{d_i(v)} + O_L(\frac{n_i}{\mu n_i}) = \frac{N_i}{\mu^{-1} \ln n}
\]

\[
\sum_{v \in T_i} \mu^{-1} N_i \ln n_i + \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-2} \ln^2 n_i} + O_L(\frac{n_i}{\ln^2 n_i}) = \sum_{v \in T_i} \mu^{-1} N_i \ln n_i + O_L(\frac{n_i}{\ln^2 n_i}).
\]

(25)

Now we prove that

\[
\sum_{v \neq v_i} \frac{1}{d_i(v)} (e^{-d_i(v)} - e^{-(\mu^{-1} \ln n_i)} \Lambda_i) = O_L(\frac{n_i}{\ln^2 n_i}),
\]

(26)

which obviously implies,

\[
\varphi(T_i, \Lambda_i) = (1 - e^{-(\mu^{-1} \ln n_i)} \Lambda_i) + O_L(\frac{n_i}{\ln^2 n_i}).
\]

(27)

For simpler calculations we show the bound in (26) by considering, \( e^{-\Lambda_i [\mu^{-1} \ln n_i]} \) instead of \( e^{-(\mu^{-1} \ln n_i)} \Lambda_i \). That one can do this is because multiplying the Taylor estimate in (25) by \( e^{-\Lambda_i [\mu^{-1} \ln n_i]} \), gives the same expression up to the error term \( O_L(\frac{n_i}{\ln^2 n_i}) \) as multiplying by \( e^{-(\mu^{-1} \ln n_i)} \Lambda_i \). For \( j > 0 \),

\[
e^{-|\mu^{-1} \ln n_i| + j} \Lambda_i = e^{-\Lambda_i [\mu^{-1} \ln n_i]} + e^{-(\mu^{-1} \ln n_i) + j} \Lambda_i (1 - e^{-j \Lambda_i})
\]

and

\[
e^{-|\mu^{-1} \ln n_i| - j} \Lambda_i = e^{-\Lambda_i [\mu^{-1} \ln n_i]} + e^{-(\mu^{-1} \ln n_i) + j} \Lambda_i (e^{-2j \Lambda_i} - e^{-j \Lambda_i}).
\]

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

\[
Q_1 + Q_2 = O_L(\frac{n_i}{\ln^2 n_i}),
\]

(28)

where

\[
Q_1 := \sum_{j=1}^{\lfloor \ln^{0.6} n_i \rfloor} \sum_{d_i(v)=j} e^{-(\mu^{-1} \ln n_i) + j} \Lambda_i \cdot (1 - e^{-j \Lambda_i}) \cdot \frac{1}{[\mu^{-1} \ln n_i] - j},
\]

\[
Q_2 := \sum_{j=1}^{\lfloor \ln^{0.6} n_i \rfloor} \sum_{d_i(v)=j} e^{-(\mu^{-1} \ln n_i) + j} \Lambda_i \cdot (e^{-2j \Lambda_i} - e^{-j \Lambda_i}) \cdot \frac{1}{[\mu^{-1} \ln n_i] + j}.
\]
We have
\[ Q_1 \leq N_i e^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i) \Lambda_i} \cdot (1 - e^{-\ln^0.6 n_i \Lambda_i}) \cdot \frac{1}{[\mu^{-1} \ln n_i] - \ln^0.6 n_i} = N_i \mathcal{O} \left( \frac{\ln^0.6 n_i \Lambda_i}{\mu^{-1} \ln n_i} \right) e^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i) \Lambda_i}. \] (29)
and similarly
\[ Q_2 = N_i \mathcal{O} \left( \frac{\ln^0.6 n_i \Lambda_i}{\mu^{-1} \ln n_i} \right) e^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i) \Lambda_i}. \]
Since \( \Lambda_i \) is \( \text{Exp} \left( \frac{1}{L + 1} \right) \) random variable, we get that
\[
\mathbb{E}(\Lambda_i e^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i) \Lambda_i}) = \int_0^\infty (L + 1) ye^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i)y} e^{-y(L+1)} dy = \frac{|(L+1) ye^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i y)}|}{|[\mu^{-1} \ln n_i] + \ln^0.6 n_i - L - 1|} \int_0^\infty (L + 1) e^{(-[\mu^{-1} \ln n_i] + \ln^0.6 n_i (L+1)y)} dy = \frac{(L + 1)}{([\mu^{-1} \ln n_i] - \ln^0.6 n_i + L + 1)^2}.
\]
Thus, (28) holds and it follows that (27) is satisfied.

Now we show that (27) implies (23) 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,
\[
e^{-(\mu^{-1} \ln n_i) \Lambda_i} \cdot \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} = \mathcal{O}_{L^1} \left( \frac{n_i}{\ln n_i} \right).
\]
Recall from Section 1.2.2 that the number of bad nodes in \( T_i \) is \( \mathcal{O}_{L^1} \left( \frac{n_i}{\ln^3 n_i} \right) \) 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 \( \mathcal{O}_{L^1} \left( \frac{1}{n^r} \right) \). By using these facts we get an obvious upper bound of the total path length, i.e.,
\[
|Y(T_i) - \mu^{-1} N_i \ln n_i| \leq N_i \ln^0.6 n_i + \mathcal{O}_{L^1} \left( \frac{n_i}{\ln n_i} \right).
\]
Hence,
\[
\frac{|Y(T_i) - \mu^{-1} N_i \ln n_i| e^{-(\mu^{-1} \ln n_i) \Lambda_i}}{\mu^{-2} \ln^2 n_i} = \mathcal{O}_{L^1} \left( \frac{n_i}{\ln^{2.2} n_i} \right),
\]
and Lemma 2.1 follows. \( \square \)
Recall from Section 2.1 that we define \( \psi(T_i, \Lambda_i) := \text{Var}(X(T_i) | T_i, \Lambda_i) \), and that we write \( E_{n_i}(.,:) := E(.) | n_i \) for the conditional expected value given \( n_i \).
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. For all vertices \( v \in T_i \), let \( I_v \) be the same indicator as in the proof of Lemma 2.1 above. Suppose that \( v \) and \( w \) are two vertices in \( T_i \) at depth \( d(i) = j, d(w) = k \) with last common ancestor at depth \( d(v) = d \). Suppose first that \( d < j, d < k \). Let \( \{v_i, u_1, \ldots, u_d = u\} \) be the vertices in the path from \( v \) to \( u \) and let \( Z = \min\{\lambda_{u_i} : 1 \leq s \leq d\} \). Conditioned on \( Z \), \( I_v \) and \( I_w \) are independent. Let \( Z \wedge \Lambda_i \) denote the minimum of \( Z \) and \( \Lambda_i \). Since \( v \) has depth \( j - d \) above \( u \), (24) yields
\[
E(I_v \mid Z) = \frac{1 - e^{-(j-d)(Z \wedge \Lambda_i)}}{j - d},
\]
and similarly for \( I_w \). (Compare this with [10, Lemma 2.2].) As in [10, equation (18)],
\[
E(I_v I_w) = \frac{1}{j - d} \left( 1 - e^{-d\Lambda_i} - \frac{d}{j} (1 - e^{-j\Lambda_i}) - \frac{d}{k} (1 - e^{-k\Lambda_i}) + \frac{d}{j + k - d} \left( 1 - e^{-(j+k-d)\Lambda_i} \right) + e^{-d\Lambda_i} - e^{-j\Lambda_i} - e^{-k\Lambda_i} + e^{-(j+k-d)\Lambda_i} \right). \tag{30}
\]
The covariance of \( I_v \) and \( I_w \) is
\[
\text{Cov}(I_v, I_w) = E(I_v I_w) - E(I_v)E(I_w).
\]
We say that a pair \((v, w)\) 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 otherwise it is ”bad”. From [13, equation (7)] by (24) and (30) above, for a good pair
\[
\text{Cov}(I_v, I_w) = \frac{1}{jk} e^{-(j+k-d)\Lambda_i} (1 - e^{-d\Lambda_i}) + \mathcal{O}\left( \frac{d}{\ln^3 n_i} \right) = \mathcal{O}_L\left( \frac{d}{\ln^3 n_i} \right). \tag{31}
\]
(Compare this with [10, equation (19)].) Since the number of bad vertices is \( \mathcal{O}_L\left( \frac{n_i}{\ln^{0.6} n_i} \right) \) it follows that the number of bad pairs, is \( \mathcal{O}_L\left( \frac{n_i^2}{\ln^{0.6} n_i} \right) \). Hence, because of the obvious upper bound that \( \text{Cov}(I_v, I_w) \) is at most 1, the sum of covariances for the bad pairs is \( \mathcal{O}\left( \frac{n_i^2}{\ln^3 n_i} \right) \). Thus,
\[
E_{n_i}(\psi(T_i, \Lambda_i)) = E_{n_i} \left( \sum_{\text{good } (v, w) \in T_i} \text{Cov}(I_v, I_w) \right) + \mathcal{O}\left( \frac{n_i^2}{\ln^3 n_i} \right). \tag{32}
\]
Recall that $\mathcal{G}_j$ is the $\sigma$-field generated by the split vectors for all vertices $v$ with $d(v) \leq j$. Recall the representation of subtree sizes in split trees described in (16) in Section 1.4.1. Recall that $n_{iv}$ denotes the number of balls in the subtrees rooted at $v$ for $v \in T_i$. From (16) we get that for $v$, where $d_i(v) = d$,

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

Thus,

$$E_{n_i}(n_{iv}) \leq n_i \prod_{r=1}^{d} E(W_r) + ds_1 = \frac{n_i}{b^d} + ds_1.$$ 

Again by using (16) we get that

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

Thus,

$$E_{n_i}(n_{iv}^2) \leq n_i^2 \prod_{r=1}^{d} E(W_r^2) + O\left(\frac{n_i d}{b^d}\right) + O(d^2). \quad (33)$$

Note that $E(W_r^2) < E(W_r) = \frac{1}{b}$ since $W_r \in [0,1]$. Hence, there is an $\epsilon > 0$ such that the right hand-side in (33) is bounded by

$$\frac{n_i^2}{(b+\epsilon)^d} + O\left(\frac{n_i d}{b^d}\right) + O(d^2). \quad (34)$$

From (32) by using (31), (33) and (34),

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

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 $\beta > \frac{1}{-\log_b E(V^2)-1}$. Then

$$X(T^n) = \sum_{i=1}^{b^L} \varphi(T_i, \Lambda_i) + o_p\left(\frac{n}{\ln^3 n}\right).$$

---

18
Proof. We write the number of records as \( \{ P^* + P_1 + \ldots + P_{bL} \} \), 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 \( \sigma \)-field generated by \( \{ \lambda_v : d(v) \leq L \} \) and \( \mathcal{F}_L^* \) the \( \sigma \)-field generated by \( T^* \) and \( \mathcal{F}_L \). We also note that \( \mathbb{E}(P_i | \mathcal{F}_L^*) = \varphi(T_i, \Lambda_i) \). By the same calculation as in [10, equation (22)],

\[
\mathbb{E}\left( X(T^*) - P^* - \sum_{i=1}^{bL} \varphi(T_i, \Lambda_i) \right)^2 \bigg| \mathcal{F}_L^* \bigg) = \sum_{i=1}^{bL} \psi(T_i, \Lambda_i). \quad (35)
\]

Taking the expectation of the conditional expected value in (35) yields

\[
\mathbb{E}\left( X(T^*) - P^* - \sum_{i=1}^{bL} \varphi(T_i, \Lambda_i) \right)^2 \bigg| \mathcal{F}_L \bigg) = \sum_{i=1}^{bL} \mathbb{E}\psi(T_i, \Lambda_i). \quad (36)
\]

We observe the obvious fact that the sum of those \( n_i, i \in \{1, \ldots, bL\} \), that are less than \( n_i kL \) for \( k \) large enough, is bounded by

\[
bL \cdot \frac{n}{b^kL} = \mathcal{O}\left( \frac{n}{\ln^3 n} \right). \quad (37)
\]

(Note that by choosing \( k \) large enough in (37) the power of the logarithm can be taken arbitrarily large.) Lemma 2.2 and (37) give that

\[
\sum_{i=1}^{bL} \mathbb{E}_n(\psi(T_i, \Lambda_i)) = \mathcal{O}\left( \sum_{i=1}^{bL} \frac{n^2}{\ln^3 n} \right). \quad (38)
\]

(Compare this with [10, equation (25)].) The expected value of the sum in (38) is equal to the expected value of the left hand-side in (36). From the calculations in (33) above for \( i \in \{1, \ldots, bL\} \),

\[
\mathbb{E}(n_i^2) \leq n^2(\mathbb{E}(V^2))^L \quad (39)
\]

Hence, choosing \( \beta > \frac{1}{-\log_b \mathbb{E}(V^2) - 1} \) one gets from (39) that

\[
\sum_{i=1}^{bL} \mathbb{E}(n_i^2) = o\left( \frac{n^2}{\ln n} \right), \quad (40)
\]

and thus the left hand-side in (36) is \( o\left( \frac{n^2}{\ln^2 n} \right) \). Thus, Lemma 2.3 follows from the well-known Markov inequality. \( \square \)
Applying Lemma 2.1 and Lemma 2.3 yields for $\beta > -\log_b \frac{1}{E(V^2) - 1}$ that

$$X(T^n) = \sum_{i=1}^{b^L} \left( \frac{2N_i}{\mu^{-1} \ln n_i} - \frac{N_i e^{-(-1) \ln n_i} \Lambda_i}{\mu^{-1} \ln n_i} \right) + \sum_{\text{good } v \in T_i} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i}$$

$$- \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} + o_p\left( \frac{n}{\ln^2 n} \right),$$

(41)

where we used that the Markov inequality gives $O_L\left( \frac{n}{\ln^{2+\varepsilon} n} \right) = o_p\left( \frac{n}{\ln^2 n} \right)$.

In [11, Corollary 2.2] we prove that

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

(42)

We get for $n_i \geq \frac{n}{b^2 \varepsilon}$,

$$E_{n_i} \left( \frac{N_i e^{-(1) \ln n_i} \Lambda_i - N_i e^{-(1) \ln n} \Lambda_i}{\mu^{-1} \ln n_i} \right)$$

$$= \frac{L + 1}{L + 1 + \mu^{-1} \ln n_i} - \frac{L + 1}{L + 1 + \mu^{-1} \ln n}$$

$$= O\left( \frac{L^2}{\ln^2 n} \right),$$

and it follows that

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

(43)

Again we use the bound in (37) for those $n_i < \frac{n}{b^2 \varepsilon}$ (for large enough $k$) so that we can ignore them in the sums in (41). Thus, by (42) and (43) with another application of the Markov inequality, the approximation in (41) can be simplified to

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2N_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i}$$

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

(44)

(Compare this with [10, equation(27)].)
By choosing $\beta$ large enough we can sharpen the error term in (40), i.e.,

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

for arbitrary large $k$. Applying (45), the variance result in (6), and assuming (A3), Chebyshev’s inequality results in

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

(46)

The third sum in (44) is treated similarly. For simplicity (in the calculations below) we change the notation $N_i$, $1 \leq i \leq b^L$, to $N_{iv}$, $d(v) = L$, and similarly for $n_i$, $1 \leq i \leq b^L$. Hence, from (44), for $\beta$ large enough, we get

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2\alpha n_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i}$$

$$- \frac{1}{\mu^{-1} \ln n} \sum_{i=1}^{b^L} \alpha n_i e^{-\left(\mu^{-1} \ln n\right) \lambda_i} + \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right).$$

(47)

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

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

Thus, choosing $\beta > \frac{1}{-\log_b E(V^{-1})}$ from (47),

$$X(T^n) = \sum_{i=1}^{b^L} \frac{2\alpha n_i}{\mu^{-1} \ln n_i} - \sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i}$$

$$- \frac{1}{\mu^{-1} \ln n} \sum_{d(v) \leq L} \alpha n_v e^{-\left(\mu^{-1} \ln n\right) \lambda_v} + \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right).$$

Proof. Recall that we write $m := \mu^{-1} \ln n$, and $\Lambda_i$ for the minimum of the $L + 1$ i.i.d. random variables $\lambda_v$, $v \in P(v_i) = \{\sigma, \ldots, v_i\}$, where $P(v_i)$ is the path from the root $\sigma$ to $v_i$. Thus, $e^{-m \Lambda_i}$ is the maximum. Now we define $\Lambda_j^l$ as the $j$-th smallest value in $\{\lambda_v, v \in P(v_i)\}$, so that $e^{-m \Lambda_j^l}$ is the $j$-th maximum. Note in particular that $\Lambda_1^1 = \Lambda_i$. Choosing $a = \frac{2\ln m}{m}$ gives that
for some \(i\), the probability that at least \(\lceil \beta \rceil + 1\) of the \(\lambda_v\)'s, \(v \in P(v_i)\), are less than \(a\) is

\[
\mathcal{O}\left( b^L L^{\lceil \beta \rceil + 1} a^{|\beta| + 1} \right) = \mathcal{O}\left( \frac{b^L \ln^2(\lceil \beta \rceil + 1) m}{m^{|\beta| + 1}} \right) = o(1).
\]

Thus, with probability tending to 1, there are at most \(\lceil \beta \rceil\) values \(\lambda_v\) less than \(a\) in each \(P(v_i)\), giving for each \(i\),

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

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

\[
\sum_{i=1}^{b^L} n_i \sum_{j=1}^{\lfloor \beta \rfloor} e^{-m \Lambda^j} = \sum_{i=1}^{b^L} n_i \sum_{v \in P(v_i)} e^{-m \lambda_v} + o_p\left( \frac{n_{\ln n}}{\ln n} \right)
\]

\[
= \sum_{d(v) \leq L} e^{-m \lambda_v} \sum_{i:v \in P(v_i)} n_v + o_p\left( \frac{n_{\ln n}}{\ln n} \right)
\]

\[
= \sum_{d(v) \leq L} n_v e^{-m \lambda_v} + o_p\left( \frac{n_{\ln n}}{\ln n} \right).
\]

Observing that the second smallest value \(\Lambda^2\) in \(i : v \in P(v_i)\), is at most \(x\) if at least two \(\lambda_v\) are at most \(x\), and using that the \(\lambda_v\)'s are i.i.d. we calculate the distribution function of \(\Lambda^2\) as

\[
P\left( \Lambda^2_i \leq x \right) = 1 - P(\lambda_v > x)^L - LP(\lambda_v > x)^{L-1} P(\lambda_v \leq x)
\]

\[
= 1 - e^{-Lx} - Le^{-(L-1)x}(1 - e^{-x}).
\]

Hence,

\[
E\left( e^{-m \Lambda^2_i} \right) = \int_0^\infty e^{-mx} \left( (L - L^2)e^{-Lx} + L(L-1)e^{-(L-1)x} \right) dx
\]

\[
= \frac{L - L^2}{m + L} + \frac{L^2 - L}{m + L - 1} = \mathcal{O}\left( \frac{L^2}{m^2} \right),
\]

implying

\[
\sum_{i=1}^{b^L} n_i \sum_{j=2}^{\lfloor \beta \rfloor} e^{-m \Lambda^j} = \mathcal{O}_1\left( \frac{nL^2}{m^2} \right).
\]
Thus, the Markov inequality gives

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

Thus, from Lemma 2.4 (where \( \beta \) is chosen large enough), by applying (46) and the total path length result in (9) we get

\[ X(T^n) = \sum_{d(v) = L} \frac{\alpha n_v}{\mu^{-1} \ln n_v} - \frac{1}{\mu^{-1} \ln n} \sum_{d(v) \leq L} \alpha n_v e^{-(\mu-1 \ln n) \lambda_v} - \frac{\zeta n}{\mu^{-2} \ln^2 n} \]

\[ + \frac{\alpha n \sigma^2}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \quad (48) \]

As in [13] and [10] the proof of Theorem 1.1, i.e., the main theorem, will be completed by a classical theorem for convergence of triangular arrays to infinitely divisible distributions, see e.g., [16, Theorem 15.28]. First we recall the definition of

\[ \xi_v := \frac{mn_v}{n} e^{-m \lambda_v} \quad (49) \]

in Section 2.1. Normalizing \( X(T^n) \) gives by using (48),

\[ \frac{\mu^{-2} \ln^2 n}{\alpha n} \left( X(T^n) - \frac{\alpha n}{\mu^{-1} \ln n} - \frac{\alpha n \ln \ln n}{\mu^{-2} \ln^2 n} + \frac{\zeta n}{\mu^{-2} \ln^2 n} \right) \]

\[ = - \sum_{d(v) \leq L} \xi_v + \frac{\mu^{-2} \ln^2 n}{n} \sum_{d(v) = L} \frac{n_v}{\mu^{-1} \ln n_v} - \mu^{-1} \ln n \]

\[ - \mu^{-1} \ln n + \mu^{-2} \sigma^2 + o_p(1). \quad (50) \]

Let

\[ D := \frac{\mu^{-2} \ln^2 n}{n} \sum_{d(v) = L} \frac{n_v}{\mu^{-1} \ln n_v} - \mu^{-1} \ln n - \mu^{-1} \ln n + \mu^{-2} \sigma^2. \quad (51) \]

and \( \xi'_i = \frac{-D}{n} \). Thus,

\[ \frac{\mu^{-2} \ln^2 n}{\alpha n} \left( X(T^n) - \frac{\alpha n}{\mu^{-1} \ln n} - \frac{\alpha n \ln \ln n}{\mu^{-2} \ln^2 n} + \frac{\zeta n}{\mu^{-2} \ln^2 n} \right) \]

\[ = - \sum_{d(v) \leq L} \xi_v - \sum_{i=1}^{n} \xi'_i + o_p(1). \quad (52) \]

23
As in [10] since the \( n_v \)'s in the sums in (50) are not independent (although they are less dependent for vertices that are far from each other), \( \{\xi_v\} \cup \{\xi'_i\} \) is not a triangular array. Recall the definition of \( \Omega_L \) as the \( \sigma \)-field generated by \( \{n_v, \ d(v) \leq L\} \). Hence, conditioned on \( \Omega_L \), \( \{\xi_v\} \cup \{\xi'_i\} \) is a triangular array with \( \xi'_i \) conditioned on \( \Omega_L \) 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 [13] and [10], 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., [16, Theorem 15.28]. For the sake of independence we intend to condition on the \( n_v \)'s in the sums in (52). We show that conditioned on the \( n_v \)'s we get convergence in distribution for the normalized \( X(T_n) \) to a random variable \( W \) with an infinitely divisible distribution, which is not depending on the \( n_v \)'s we conditioned on. Then it follows in the same way as in [10] that also unconditioned the normalized \( X(T_n) \) converges in distribution to \( W \). The main Theorem 1.1 is proven by Theorem 2.1 below.

**Theorem 2.1.** Choose any constant \( c > 0 \) and let \( n \to \infty \). Conditioning on the \( \sigma \)-field \( \Omega_L \), where \( L = \lceil \beta \log_b \ln n \rceil \), if the constant \( \beta \) is chosen large enough the following hold:

1. \( \sup_v P(\xi_v > x | \Omega_L) \to 0 \) for every \( x > 0 \),
2. \( \Delta_1 := \sum_{d(v) \leq L} P(\xi_v > x | \Omega_L) \overset{p}{\to} \nu(x, \infty) = \frac{\mu^{-1}}{x} \) for every \( x > 0 \),
3. \( \Delta_2 := \sum_{d(v) \leq L} E(\xi_v 1[\xi_v \leq c] | \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 
\overset{p}{\to} -\mu^{-1} \ln \mu^{-1} + \mu^{-1} - \mu^{-2} \sigma^2 - \frac{\sigma^2 - \mu^2}{2\mu^2} + \mu^{-1} \ln c, 
4. \( \Delta_3 := \sum_{d(v) \leq L} \text{Var}(\xi_v 1[\xi_v \leq c] | \Omega_L) \overset{p}{\to} \mu^{-1} c. 

Before proving Theorem 2.1 we will show how it proves Theorem 1.1
Recall from (51) that
\[ D = \frac{\mu - 2 \ln^2 n}{n} \sum_{d(v) \leq L} \frac{n_v}{\mu - 1 \ln n_v} - \mu^{-1} \ln n - \mu^{-1} \ln n + \mu^{-2} \sigma^2. \]

We apply [16, Theorem 15.28] with
\[ a = 0, \quad b = -\mu^{-1} \ln \mu^{-1} + \mu^{-1} - \mu^{-2} \sigma^2 - \frac{\sigma^2 - \mu^2}{2 \mu^2} \] (53)
to \( \sum_{d(v) \leq L} \xi_v + \sum_{i=1}^n \xi_i' \) conditioned on \( \Omega_L \) with \( \xi_i' = \frac{-D}{n} \) deterministic. The constants \( a \) and \( b \) are the constants that occur in the general formula of the characteristic function for infinitely divisible distributions in \( \{13\} \). Note that \( \frac{D}{n} \to 0 \), thus because of (i), conditioned on \( \Omega_L \), \( \{\xi_v\} \cup \{\xi_i'\} \) is a null array.

We define \( S(n) := \sum_{d(v) \leq L} \xi_v + \sum_{i=1}^n \xi_i' \). From (ii) we have that \( \frac{d\nu}{dx} = \frac{\mu^{-1}}{x^2} \), hence
\[ \int_c^0 x^2 d\nu(x) = \int_c^0 \mu^{-1} dx = \mu^{-1} c \text{ and } \int_c^1 x d\nu(x) = \int_c^1 \mu^{-1} \frac{1}{x} dx = -\mu^{-1} \ln c. \]

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 in (53). The convergence in Theorem 2.1 is in the probabilistic sense, while [16, Theorem 15.28] requires usual convergence, i.e., standard point-wise convergence of sequences with no probability involved. However, if the convergence instead were a.s. in Theorem 2.1 then it would have been easy to see from this theorem that conditionally on \( \Omega_L \) the conditions of [16, Theorem 15.28] are fulfilled for \( S(n) \). Thus, assuming a.s. convergence in Theorem 2.1 [16, Theorem 15.28] implies that conditioned on \( \Omega_L \),
\[ S(n) \overset{d}{\to} W, \quad \text{as } n \to \infty, \] (54)
where \( W \) has an infinitely divisible distribution (in particular a weakly 1-stable distribution in this case) with characteristic function
\[ \mathbb{E}(e^{itW}) = \exp \left( itb + \int_0^\infty (e^{itx} - 1 - itx + 1) d\nu(x) \right); \]
this is (14) in Remark 1.2 (since \( b = C + \mu^{-1}(\gamma - 1) \)) which can be simplified to (12) in Theorem 1.1.

It follows from (54) that conditioning on \( \Omega_L \) has no influence on the distributional convergence of \( S(n) \) (unconditioned), since for any continuous bounded function \( g : \mathbb{R} \to \mathbb{R} \),
\[ \mathbb{E}(g(S(n)) \mid \Omega_L) = \int g dF(S(n) \mid \Omega_L) \overset{n \to \infty}{\to} \mathbb{E}(g(W)). \]
Thus, taking expectation by dominated convergence
\[ E\left(g(S(n))\right) \xrightarrow{n \to \infty} E\left(g(W)\right). \]

This shows that also unconditioned \( S(n) \xrightarrow{d} W \). Thus, unconditioned the normalized \( X(T^n) \) in (50) converges in distribution to \(-W\).

It remains to show that convergence in probability (which is the type of convergence in Theorem 2.1) actually is sufficient for \( S(n) \xrightarrow{d} W \) to hold. In [10] we proved this fact for the binary search tree in two ways, in one by using subsequences and in the other one by using Skorohod’s coupling theorem, see e.g., [16, Theorem 3.30]. By analogy these proofs also work for general split trees. Thus, the proof of Theorem 1.1 for \( X_v(T) \) is completed.

Now it follows easily, by the same type of argument as for the binary search tree [10] that the result holds for \( X_e(T) \) too. One way to see this is to consider \( \hat{T} \) as the tree \( T \) with the root deleted. Then there is a natural 1-1 correspondence between edges of \( T \) and vertices of \( \hat{T} \), and this correspondence also preserves the record (and cutting) operations. Since it is very unlikely that the root value would decide if values at high levels are records or not, it follows that asymptotically \( X_e(T) \) and \( X_v(T) \) have the same distribution.

Thus, the proof of Theorem 1.1 is completed.

The idea of the proof of Theorem 2.1 is as for the binary search tree [10, Theorem 2.1] to use Chebyshev’s inequality to prove (ii), (iii) and (iv) of Theorem 2.1 ((i) is very easy to prove). For the binary search tree we frequently used in [10, Theorem 2.1] that the sum \( \sum_{r=1}^{k} \ln U_r \), where \( U_r \) are uniform \( U(0,1) \) random variables, is distributed as a \(-\Gamma(k,1)\) random variable. For general split trees, the solution of the renewal function \( U(t) \) in (20) is fundamental for the proof of Theorem 2.1.

### 2.3.2 Lemmas for the Proof of Theorem 2.1

Recall that we write \( \Omega_j \) for the \( \sigma \)-field generated by \( \{n_v, d(v) \leq j\} \) and \( \mathcal{G}_j \) for the \( \sigma \)-field generated by \( \{W_{r,v}, r \in \{1,2,\ldots,j\}\} \), for all vertices \( v \) with \( d(v) = j \). Also recall that we write \( L = [\beta \log_b \ln n]\). We also write

\[ \hat{n}_v := n \prod_{r=1}^{k} W_{r,v}, \quad \text{and} \quad \hat{\xi}_v := \frac{m \hat{n}_v}{n} e^{-m \lambda_v}, \quad (55) \]

where \( m := \mu^{-1} \ln n \). Note that \( \mathcal{G}_j \) is thus equivalently the \( \sigma \)-field generated by \( \{\hat{n}_v : d(v) \leq j\} \).

We present below four crucial lemmas by which we can then easily prove Theorem 2.1.
Lemma 2.5. Suppose that \( n \to \infty \) and choose any constant \( c > 0 \). Then for \( L = \lfloor \beta \log_b \ln n \rfloor \) and \( \beta \) large enough, the following hold

\[
\sum_{d(v) \leq L} P(\xi_v > x | \Omega_L) = \sum_{d(v) \leq L} P(\hat{\xi}_v > x | \mathcal{G}_L) + o_p(1),
\]
\[
\sum_{d(v) \leq L} E(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} E(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o_p(1),
\]
\[
\sum_{d(v) = L} \frac{n_v}{\mu^{-1} \ln n_v} = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \frac{\hat{n}_v \ln (\hat{n}_v)}{n} + o_p\left(\frac{n}{\ln^2 n}\right),
\]
\[
\sum_{d(v) \leq L} \text{Var}(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o_p(1).
\]

For simplicity we sometimes use a short notation for the following sums, i.e.,

\[
\Phi_v := \left( \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \frac{\hat{n}_v \ln (\hat{n}_v)}{n} \right),
\]
\[
R_1 := \sum_{d(v) \leq L} P(\hat{\xi}_v > x | \mathcal{G}_L), \tag{56}
\]
\[
R_2 := \sum_{d(v) \leq L} E(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) - \frac{\mu^{-2} \ln^2 n}{n} \cdot \Phi_v, \tag{57}
\]
\[
R_3 := \sum_{d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L). \tag{58}
\]

Lemma 2.6. Suppose that \( n \to \infty \) and choose any constant \( c > 0 \). Then for \( L = \lfloor \beta \log_b \ln n \rfloor \) and \( \beta \) large enough, the following hold

\[
E\left(R_1\right) = \frac{\mu^{-1}}{x} + o(1) = \nu(x, \infty) + o(1), \quad \text{for every } x > 0,
\]
\[
E\left(R_2\right) = -\mu^{-1} \ln n - \mu^{-1} \ln \ln n + \mu^{-1} - \mu^{-1} \ln \mu^{-1} + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1),
\]
\[
E\left(R_3\right) = \mu^{-1} c + o(1).
\]
Let \( l := \left\lfloor \frac{\log_b \ln n}{2} \right\rfloor \) and for short write
\[
S_1 : = \sum_{l \leq d(v) \leq L} P(\xi_v > x | \mathcal{F}_L),
\]
\[
S_2 : = \sum_{l \leq d(v) \leq L} E(\xi_v 1[\xi_v \leq c] | \mathcal{F}_L) - \frac{\mu^2 \ln^2 n}{n} \cdot \Phi_v,
\]
\[
S_3 : = \sum_{l \leq d(v) \leq L} \text{Var}(\xi_v 1[\xi_v \leq c] | \mathcal{F}_L).
\]

**Lemma 2.7.** Suppose that \( n \to \infty \). Then for \( L = \lfloor \beta \log_b \ln n \rfloor \) (where \( \beta \) is large enough) and \( l = \lfloor \frac{\log_b \ln n}{2} \rfloor \), the following limits hold
\[
\text{Var} \left( E(S_1 | \mathcal{F}_L) \right) \to 0,
\]
\[
\text{Var} \left( E(S_2 | \mathcal{F}_L) \right) \to 0,
\]
\[
\text{Var} \left( E(S_3 | \mathcal{F}_L) \right) \to 0.
\]

**Lemma 2.8.** Suppose that \( n \to \infty \). Then for \( L = \lfloor \beta \log_b \ln n \rfloor \) (where \( \beta \) is large enough) and \( l = \lfloor \frac{\log_b \ln n}{2} \rfloor \) the following limits hold
\[
E \left( \text{Var}(S_1 | \mathcal{F}_L) \right) \to 0,
\]
\[
E \left( \text{Var}(S_2 | \mathcal{F}_L) \right) \to 0,
\]
\[
E \left( \text{Var}(S_3 | \mathcal{F}_L) \right) \to 0.
\]

Before proving these lemmas we show how their use leads to the proof of Theorem 2.1.

### 2.3.3 Proof of Theorem 2.1

Recall that \( m = \mu^{-1} \ln n \). For any \( x > 0 \), and \( v \) with \( d(v) \leq L \), we have
\[
P(\xi_v > x | \Omega_L) = P(e^{-m\lambda_v} > \frac{nx}{mn_v} | \Omega_L) = P(\lambda_v < \frac{1}{m} \ln \frac{mn_v}{nx} | \Omega_L)
\]
\[
= 1 - \exp \left( -\frac{1}{m} \ln \frac{mn_v}{nx} \right).
\]
Thus, for every \( x > 0 \),
\[
P(\xi_v > x | \Omega_L) \leq \frac{1}{m} \ln_+ \frac{mn_v}{nx} \leq \frac{1}{m} \ln_+ \frac{m}{x} \to 0,
\]

28
which proves (i).

Recall the definitions of $R_1$, $R_2$ and $R_3$ in (56), (57) and (58). Note that Lemma 2.5 shows that in Theorem 2.1 the left hand-sides of (ii), (iii) and (iv), i.e., $\Delta_1$, $\Delta_2$ and $\Delta_3$, respectively, are equal to

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

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

We complete the proof of Theorem 2.1 by showing that

\[\text{Var} (R_1) \to 0 \text{ for every } x > 0, \quad \text{Var} (R_2) \to 0, \quad \text{and } \text{Var} (R_3) \to 0. \ (70)\]

Then by Chebyshev’s inequality (ii), (iii) and (iv) of Theorem 2.1 follow. Thus, it remains to show how (70) follows from Lemma 2.6 and Lemma 2.7.

By using (69), one easily obtains

\[
\sum_{d(v) \leq L} \mathbb{P} (\hat{\xi}_v > x | \mathcal{G}_L) = \sum_{l \leq d(v) \leq L} \mathbb{P} (\hat{\xi}_v > x | \mathcal{G}_L) + o(1), \quad (71)
\]
\[
\sum_{d(v) \leq L} \mathbb{E} (\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) = \sum_{l \leq d(v) \leq L} \mathbb{E} (\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o(1), \quad (72)
\]
\[
\sum_{d(v) \leq L} \mathbb{V}ar (\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) = \sum_{l \leq d(v) \leq L} \mathbb{V}ar (\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o(1). \quad (73)
\]

Hence,

\[
R_1 = S_1 + o(1), \quad R_2 = S_2 + o(1), \quad R_3 = S_3 + o(1). \quad (74)
\]

To show (70) we use a variance formula that is easy to establish, see e.g., [9, exercise 10.17-2],

\[
\text{Var}(X) = \mathbb{E}(\text{Var}(X | \mathcal{G})) + \text{Var}(\mathbb{E}(X | \mathcal{G})), \quad (75)
\]

where $X$ is a random variable and $\mathcal{G}$ is a sub $\sigma$-field.

Recall that $\mathcal{G}_j$ is the $\sigma$-field generated by $\{W_{r,v}, r \in \{1,2,\ldots,j\}\}$, for all vertices with $d(v) = j$. Consequently, by applying the variance formula in (75), from Lemma 2.7 and Lemma 2.8 we get as $n \to \infty$

\[
\text{Var}(S_1) = \mathbb{E}(\text{Var}(S_1 | \mathcal{G}_j)) + \text{Var}(\mathbb{E}(S_1 | \mathcal{G}_j)) \to 0,
\]
\[
\text{Var}(S_2) = \mathbb{E}(\text{Var}(S_2 | \mathcal{G}_j)) + \text{Var}(\mathbb{E}(S_2 | \mathcal{G}_j)) \to 0,
\]
\[
\text{Var}(S_3) = \mathbb{E}(\text{Var}(S_3 | \mathcal{G}_j)) + \text{Var}(\mathbb{E}(S_3 | \mathcal{G}_j)) \to 0,
\]

29
and thus (70) follows from (74).

We have proved Theorem 2.1 by the use of the lemmas, and thus also Theorem 1.1.

2.3.4 Proofs of the Lemmas of Theorem 2.1

Finally we present the proofs of Lemma 2.5, Lemma 2.6, Lemma 2.7 and Lemma 2.8.

Proof of Lemma 2.5.

From (16) and (17) in Section 1.4.1 we get in particular that given \( G \)

\[
V \leq \text{Binomial}(n, \prod_{r=1}^{k} W_{r,v}) + s_1 L,
\]

\[
V > \text{Binomial}(n, \prod_{r=1}^{k} W_{r,v}) - s L.
\]

Since a Binomial \((k, p)\) random variable has expected value \(kp\) and variance \(kp(1 - p)\), the Chebyshev inequality results in

\[
P( | V - n \prod_{r=1}^{k} W_{r,v} | > n^{0.6} | \Omega_L ) \leq \frac{1}{n^{0.19}}.
\] (76)

This motivates the notation of \( \hat{n}_v := n \prod_{r=1}^{k} W_{r,v} \) in (55). Also recall that we write \( \hat{\xi}_v := \frac{m_\nu}{\alpha} e^{-m_\lambda v} \) for \( m := \mu^{-1}\ln n \), and that \( \mathcal{G}_j \) is the \( \sigma \)-field generated by \( \{ \hat{n}_v : d(v) \leq j \} \). By using (68) and (69) we get (compare with [10, equation (55)]),

\[
\sum_{d(v) \leq L} P(\xi_v > x | \Omega_L ) = \sum_{k=1}^{L} \sum_{d(v)=k} \frac{1}{m} \ln_+ \frac{mn_v}{nx} \left( 1 + O\left( \frac{\ln m}{m} \right) \right)
\] (77)

and similarly

\[
\sum_{d(v) \leq L} P(\hat{\xi}_v > x | \mathcal{G}_L ) = \sum_{k=1}^{L} \sum_{d(v)=k} \frac{1}{m} \ln_+ \frac{m\hat{n}_v}{nx} \left( 1 + O\left( \frac{\ln m}{m} \right) \right).
\] (78)

By using (77), (78) and (76) we get

\[
\sum_{d(v) \leq L} P(\xi_v > x | \Omega_L ) = \sum_{d(v) \leq L} P(\hat{\xi}_v > x | \mathcal{G}_L ) + o_p(1).
\] (79)
One easily gets (compare with [13, p.251] and [10, equation (61)–(62)]) that
\[
\sum_{d(v) \leq L} \mathbb{E}(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} \frac{mn_v}{n(m+1)} e^{- \frac{m+1}{m} \ln_+ \left( \frac{mn_v}{m} \right)} \tag{80}
\]
and similarly
\[
\sum_{d(v) \leq L} \mathbb{E}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) = \sum_{d(v) \leq L} \frac{m\hat{n}_v}{n(m+1)} e^{- \frac{m+1}{m} \ln_+ \left( \frac{m\hat{n}_v}{m} \right)}. \tag{81}
\]
Thus, (76) implies that
\[
\sum_{d(v) \leq L} \mathbb{E}(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} \mathbb{E}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o_p(1). \tag{82}
\]
Using the bound in (37) for the sum of the subtree sizes with \( n_v \) less than \( n_{kL} \) (for \( k \) large enough) we get the expansion
\[
\sum_{d(v) = L} \frac{n_v}{\mu^{-1} \ln n_v} = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \frac{n_v \ln \left( \frac{n_v}{n} \right)}{\mu^{-1} \ln^2 n} + o_p \left( \frac{n}{\ln^2 n} \right). \tag{83}
\]
By again using (76) (compare with [10, equation (68)]) we get
\[
\sum_{d(v) = L} \frac{n_v}{\mu^{-1} \ln n_v} = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \frac{\hat{n}_v \ln \left( \frac{\hat{n}_v}{n} \right)}{\mu^{-1} \ln^2 n} + o_p \left( \frac{n}{\ln^2 n} \right). \tag{84}
\]
By using the calculations in [13, p.251-252] (compare with [10, equation (70)]) we get
\[
\sum_{d(v) \leq L} \text{Var}(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} \frac{m^2 n_v^2}{2mn^2} e^{- \frac{2m+1}{m} \ln_+ \left( \frac{mn_v}{m} \right)} + o(1), \tag{83}
\]
and similarly
\[
\sum_{d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) = \sum_{d(v) \leq L} \frac{m^2 \hat{n}_v^2}{2mn^2} e^{- \frac{2m+1}{m} \ln_+ \left( \frac{m\hat{n}_v}{m} \right)} + o(1) \tag{84}
\]
Thus, using (76) we obtain
\[
\sum_{d(v) \leq L} \text{Var}(\xi_v 1[\xi_v \leq c] | \Omega_L) = \sum_{d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) + o_p(1). \tag{85}
\]
Proof of Lemma 2.6. Recall that we write
\[ Y_k = -\sum_{r=1}^{k} \ln W_r \] (86)
and that we write
\[ R_1 = \sum_{d(v) \leq L} P(\xi_v > x | \mathcal{G}_L). \]

As in the calculations in [10, equations (56)] from (78) one gets
\[ E(R_1) = (1 + o(1)) \sum_{k=1}^{L} b_k E\left( \frac{(\ln m - \ln x - Y_k)}{m} I\{Y_k \leq \ln m - \ln x\} \right). \] (87)

By using integration by parts we get that the sum in (87) is equal to
\[ \sum_{k=1}^{L} b_k \frac{1}{m} \int_{\ln m - \ln x}^{\ln m - \ln x} P(Y_k \leq t) dt = \frac{1}{m} \int_{\ln m - \ln x}^{\ln m - \ln x} \sum_{k=1}^{L} b_k P(Y_k \leq t) dt. \] (88)

Recall the definition of the renewal function \( U(t) := \sum_{k=1}^{\infty} b_k P(Y_k \leq t) \) in (19) above. We want to show that
\[ \frac{1}{m} \int_{0}^{\ln m - \ln x} \sum_{k=L+1}^{\infty} b_k P(Y_k \leq t) dt = o(1). \] (89)

To show this we use large deviations. Choose an arbitrary \( s > 0 \), by applying the Markov inequality and using that the \( W_{r,v}, r \in \{1, \ldots, k\} \), are i.i.d. we get
\[ P(Y_k \leq t) = P(-Y_k \geq -t) = P(e^{-stY_k} \geq e^{-st}) \leq \left( \mathbb{E}(V^s) \right)^{k} e^{st}. \] (90)

Choosing \( s > 1 \), we get
\[ \mathbb{E}(V^s) < \mathbb{E}(V) = \frac{1}{b}. \]

Thus, we can find \( \delta > 0 \) such that
\[ \mathbb{E}(V^s) \leq \frac{1}{b^{1+\delta}}. \] (91)
In the definition of $L = \lfloor \beta \log_b \ln n \rfloor$ the constant $\beta$ can be chosen arbitrarily large. It is enough to show that $\sum_{k=L+1}^{\infty} b^k P(Y_k \leq \ln m - \ln x)$ is $o(1)$ for proving (89). By applying (90) and (91) we get that

$$ \sum_{k=L+1}^{\infty} b^k P(Y_k \leq \ln m - \ln x) \leq \sum_{k=L+1}^{\infty} \frac{b^k m^s}{x^s} = \sum_{k=L+1}^{\infty} b^{-\delta k} m^s = O \left( m^{-\delta} m^s \right). \quad (92) $$

Thus, choosing $\beta > \frac{s-1}{\delta}$ in $L$ gives (89). Now the solution of $U(t)$ in (20) gives that the quantity in (88) is equal to

$$ \int_0^{\ln m - \ln x} U(t) dt + o(1) = \mu^{-1} + o(1) \int_0^{\ln m - \ln x} e^t dt + o(1) $$

$$ = \mu^{-1} + o(1) = \nu(x, \infty) + o(1). \quad (93) $$

Hence, $E(R_1) = \nu(x, \infty) + o(1)$.

In analogy with [10, equation (63)] by using (81) and (86) we deduce that

$$ \mathbb{E} \left( \sum \mathbb{E}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) \right) := E_1 + E_2, \quad (94) $$

where

$$ E_1 = \mathbb{E} \sum_{d(v) \leq L} \frac{m}{m+1} e^{-Y_k} e^{-\frac{m+1}{m} (\ln m - \ln c - Y_k)} I\{Y_k \leq \ln m - \ln c\}, $$

$$ E_2 = \mathbb{E} \sum_{d(v) \leq L} \frac{m}{m+1} e^{-Y_k} I\{Y_k > \ln m - \ln c\}. \quad (95) $$

By using integration by parts, applying the solution of $U(t)$ in (20) and using (92) we obtain that

$$ E_1 = e^{-\frac{m+1}{m} (\ln m - \ln c)} \frac{m}{m+1} \sum_{k=1}^{L} b^k \int_0^{\ln m - \ln c} e^{t} dP(Y_k \leq t) $$

$$ = e^{-\frac{m+1}{m} (\ln m - \ln c)} \frac{m}{m+1} \left( \left[ \sum_{k=1}^{L} b^k e^{t} P(Y_k \leq t) \right]_{0}^{\ln m - \ln c} \right. $$

$$ - \left. \int_0^{\ln m - \ln c} \sum_{k=1}^{L} b^k e^{t} P(Y_k \leq t) dt \right) $$

$$ = \mu^{-1} + o(1). \quad (96) $$
By similar calculations as in (96),

$$E_2 = \frac{m}{m+1}L - \frac{m}{m+1} \int_0^{\ln m - \ln c} \sum_{k=1}^L b_k e^{-t} \mathbb{P}(Y_k \leq t) dt$$

$$= \frac{m}{m+1}L - \frac{m}{m+1} \left( \left| \sum_{k=1}^L b_k e^{-t} \mathbb{P}(Y_k \leq t) \right|^{\ln m - \ln c} \right)_0^{\ln m - \ln c}$$

$$+ \int_0^{\ln m - \ln c} \sum_{k=1}^L b_k e^{-t} \mathbb{P}(Y_k \leq t) dt$$

$$= L - \mu^{-1} - \frac{m}{m+1} \int_0^{\ln m - \ln c} \sum_{k=1}^L b_k e^{-t} \mathbb{P}(Y_k \leq t) dt + o(1). \quad (97)$$

From (92) it follows that

$$\int_0^{\ln m - \ln c} \sum_{k=1}^L b_k e^{-t} \mathbb{P}(Y_k \leq t) dt = \int_0^{\ln m - \ln c} e^{-t} U(t) dt + o(1)$$

$$= \int_0^{\ln m - \ln c} e^{-t} (U(t) - \mu^{-1} e^t) dt + \mu^{-1} (\ln m - \ln c) + o(1).$$

Applying the solution of $W(x) := \int_0^x e^{-t} (U(t) - \mu^{-1} e^t) dt$ in (21), from (97) we get that

$$E_2 = L - \mu^{-1} \ln m + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1). \quad (98)$$

Recalling (94) and applying the approximations of $E_1$ in (96) and $E_2$ in (98) we deduce that

$$\mathbb{E} \left( \sum_{d(v) \leq L} \mathbb{E} (\xi_v \mathbb{1}[\xi_v \leq c] | \mathcal{G}_L) \right) = L + \mu^{-1} - \mu^{-1} \ln m + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1)$$

which is equal to

$$K = L + \mu^{-1} - \mu^{-1} \ln n - \mu^{-1} \ln \mu^{-1} + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1). \quad (99)$$

By the definition of $\widehat{n}_v$ in (55),

$$\Phi_v := \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \widehat{n}_v \ln \left( \frac{\widehat{m}_v}{\widehat{n}} \right)$$

$$= \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} \frac{n \prod_{r=1}^L W_{r,v} \sum_{r=1}^L \ln W_{r,v}}{\mu^{-1} \ln^2 n}, \quad (100)$$
Hence, by using the definition of $\mu$ in (1) we get that
\[
E \left( \Phi_v \right) = \frac{n}{\mu^{-1} \ln n} + \frac{nL}{\mu^{-2} \ln^2 n} .
\] (101)

Thus, recalling the definition of $R_2$ in (57) we get $E \left( R_2 \right) = K - \mu^{-1} \ln n - L,$
where $K$ is defined in (99).

Recall that
\[
R_3 = \sum_{d(v) \leq L} \text{Var} \left( \xi_v 1 [\xi_v \leq c] | G_L \right).
\]

By using (84) we get that
\[
E \left( R_3 \right) = \sum_{k=1}^{L} \frac{b^k m}{2} E \left( \prod_{r=1}^{k} W_{r,v}^2 e^{-2m+1/(\ln m - \ln c)} I \{ Y_k \leq \ln m - \ln c \} \right) + o(1)
\]
\[
= V_1 + V_2 + o(1),
\] (102)

where
\[
V_1 := e^{-2m+1/(\ln m - \ln c)} \frac{m}{2} \int_0^{\ln m - \ln c} \sum_{k=1}^{L} \frac{b^k c^k}{k} dP(Y_b \leq t),
\]
\[
V_2 := E \left( \sum_{k=1}^{L} \frac{b^k m}{2} \prod_{r=1}^{k} W_{r,v}^2 I \{ Y_k > \ln m - \ln c \} \right).
\] (103)

By applying the solution of $U(t)$ in (20), integration by parts results in
\[
V_2 = \int_{\ln m - \ln c}^{\infty} \sum_{k=1}^{L} \frac{b^k m}{2} e^{-2t} dP(Y_k \leq t)
\]
\[
= \frac{m}{2} e^{-2U(t)} \bigg|_{\ln m - \ln c}^{\infty} + m \int_{\ln m - \ln c}^{\infty} e^{-2t} U(t) dt + o(1) = \frac{\mu^{-1} c}{2} + o(1),
\] (104)

where we used (90) (choosing $1 < s < 2$) and then similar calculations as in (92) to show that if we sum over all $k$ instead of $k \leq L$ the error term is just $o(1)$.

Similarly, by using (92), integration by parts gives
\[
V_1 = \frac{\mu^{-1} c}{2} + o(1).
\] (105)

\[\square\]
Hence, \( E(R_3) = \mu^{-1}c + o(1) \).

**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

\[
Z_{j-l,v} := Y_{j,v} - Y_{l,v_i} = -\sum_{r=l+1}^{j} \ln W_{r,v}.
\] (106)

Recall from (59) that

\[
S_1 = \sum_{l \leq d(v) \leq L} P(\hat{\xi}_v > x | \mathcal{G}_L).
\]

Using (71) and the solution of the renewal equation \( U(t) \) in (20) we get by similar calculations as in (87)–(93),

\[
E(S_1 | \mathcal{G}_L) = \sum_{i=1}^{b^l} \frac{1}{m} \int_0^{\ln m - \ln x - Y_{l,v_i}} \sum_{j=l+1}^{L} b^{j-l} P(Z_{j-l,v} \leq t) dt + o(1)
\]

\[
= \sum_{i=1}^{b^l} \frac{1}{m} \int_0^{\ln m - \ln x - Y_{l,v_i}} \mu^{-1} e^t dt + o(1)
\]

\[
= \sum_{i=1}^{b^l} \prod_{r=1}^{l} W_{r,v_i} \mu^{-1} e^{-t} + o(1) = \mu^{-1} x + o(1).
\] (107)

Thus, \( \text{Var}(E(S_1 | \mathcal{G}_L)) \) is \( o(1) \), which shows (62).

We show that (63) is true by similar calculations as for showing (62). Recall from (60) that

\[
S_2 = \sum_{l \leq d(v) \leq L} E(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathcal{G}_L) - \frac{\mu^{-2} \ln^2 n}{n} \cdot \Phi_v,
\]

where

\[
\Phi_v = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v) = L} n \prod_{r=1}^{L} W_{r,v_i} \sum_{r=1}^{L} \ln W_{r,v_i}.
\]

First, as before we let \( v_i \) with \( d(v_i) = l \) be a given vertex so that there are at most \( b^{j-l} \) choices of \( v \) at depth \( j \) with ancestor \( v_i \). Recall the notation
of \( Z_{j-l,v} \) in (106), i.e., \( Y_{j,v} = Y_{l,v} + Z_{j-l,v} \). By similar calculations as in (107) (glancing at the calculations in (94)) we obtain

\[
E \left( \sum_{l \leq d(v) \leq L} E(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | \mathscr{F}_L) \right) = F_1 + F_2,
\]

where

\[
F_1 := E \left( \sum_{l \leq d(v) \leq L} \frac{m}{m+1} e^{-Y_{l,v} - Z_{j-l,v}} e^{-\frac{m+1}{m}(\ln m - \ln c - Y_{l,v} - Z_{j-l,v})} I\{Y_{l,v} + Z_{j-l,v} \leq \ln m - \ln c\} \big| \mathscr{F}_l \right),
\]

\[
F_2 := E \left( \sum_{l \leq d(v) \leq L} \frac{m}{m+1} e^{-Y_{l,v} - Z_{j-l,v}} I\{Y_{l,v} + Z_{j-l,v} > \ln m - \ln c\} \big| \mathscr{F}_l \right).
\]

Then by similar calculations as in (96),

\[
F_1 = e^{-\frac{m+1}{m}\ln m - \ln c} \frac{m}{m+1} \sum_{i=1}^{b^l} \int_0^{\ln m - \ln c - Y_{l,v}} \sum_{j=l+1}^L b^{j-l} e^{\frac{j}{m}} dP(Z_{j-l,v} \leq t)
\]

\[
= \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^l W_{r,v_i} + o(1) = \mu^{-1} + o(1).
\] (108)

By similar calculations as in (96)–(98), we obtain

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

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

\[
= L - l - \mu^{-1} \ln m + \mu^{-1} \ln c - \frac{\sigma^2 - \mu^2}{2\mu^2} - \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^l W_{r,v_i} \ln W_{r,v_i} + o(1).
\] (109)
Thus, by applying the approximations of $F_1$ in (108) and $F_2$ in (109) we get
\[
E\left( \sum_{l \leq d(v) \leq L} E(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | G_L) \right) = \mu^{-1} + L - \mu^{-1} \ln m + \mu^{-1} \ln c
- \frac{\sigma^2 - \mu^2}{2\mu^2} - \sum_{i=1}^{b^l} \mu^{-1} \prod_{r=1}^{l} W_{r,v_i} \sum_{r=1}^{l} \ln W_{r,v_i} + o(1). \tag{110}
\]

Let $v_i$ be a vertex at depth $l$ and let $v$ be a vertex at depth $L$. Similarly as in (100) and (101) (compare with [10, equations (78)–(79)]), we get that
\[
E(\Phi_v | G_l) = n \mu^{-1} \ln n + n(L - l) - \sum_{i=1}^{b^l} n \prod_{r=1}^{l} W_{r,v_i} \sum_{r=1}^{l} \ln W_{r,v_i} + o\left(\frac{n}{\ln^2 n}\right). \tag{111}
\]

From (110) and (111) we obtain that $\text{Var}(E(S_2 | G_l))$ is $o(1)$, which shows (63).

For (64) we proceed with the same method as for showing (62) and (63). Recall from (61) that
\[
S_3 = \sum_{l \leq d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | G_L).
\]

By similar calculations as in (102) and (103) we get
\[
E\left( \sum_{l \leq d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c] | G_L) \right) = I_1 + I_2 + o(1),
\]

where
\[
I_1 := e^{-2m+1\ln m - \ln c} \sum_{i=1}^{b^l} \sum_{j=l+1}^{L} \frac{b^j m}{2^i} \mathbb{P}(Z_{j-l,v} \leq t),
\]
\[
I_2 := \sum_{i=1}^{b^l} E\left( \sum_{j=l+1}^{L} \frac{b^j m}{2^i} \prod_{r=1}^{l} W_{r,v_i} \prod_{r=l+1}^{j} W_{r,v_i} I\{Y_{l,v_i} + Z_{j-l,v} > \ln m - \ln c\} | G_l \right).
\]

Using integration by parts we calculate (similarly as in (104) and (105)),
\[
I_1 + o(1) = I_2 + o(1) = \sum_{i=1}^{b^l} \prod_{r=1}^{l} W_{r,v_i} \frac{\mu^{-1}}{2} c + o(1) = \frac{\mu^{-1}}{2} c + o(1).
\]

Thus, $\text{Var}(E(S_3 | G_l))$ is $o(1)$, which shows (64).
Proof of Lemma 2.8. Recall from (59) that
\[ S_1 = \sum_{l \leq d(v) \leq L} P(\hat{\xi}_v > x | G_L). \]

For showing (65) we first note that
\[
\text{Var}\left( \sum_{l \leq d(v) \leq L} P(\hat{\xi}_v > x | G_L) \right) = \sum_{l \leq d(v) \leq L, l \leq d(w) \leq L} \text{Cov}(P(\hat{\xi}_v > x | G_L), P(\hat{\xi}_w > x | G_L)).
\]

To estimate these conditional covariances we can suppose that the closest ancestor \( u \) for \( v \) with \( d(v) \leq L \), and \( w \) with \( d(w) \leq L \) is at depth \( d \geq l \), since the other terms are just 0 because of independence. For \( d \geq l \), we use
\[
\text{Cov}(P(\hat{\xi}_v > x | G_L), P(\hat{\xi}_w > x | G_L)) \leq E(P(\hat{\xi}_v > x | G_L)P(\hat{\xi}_w > x | G_L)).
\]

Denote by \((v_u, w_u)\) a general pair of vertices with closest ancestor \( u \). Then (112) implies that
\[
E\left( \text{Var}(S_1 | G_l) \right) \leq \sum_{d=l}^{L} \sum_{u : d(u) = d(v_u, w_u)} E\left( P(\hat{\xi}_{v_u} > x | G_L)P(\hat{\xi}_{w_u} > x | G_L) \right).
\]

Recall that \( G_{d+1} \) is the \( \sigma \)-field generated by \( \{ \hat{\nu}_v : d(v) \leq d + 1 \} \). For the pair \((v_u, w_u)\) with \( d(u) = d \), conditioned on \( G_{d+1} \), \( E\left( P(\hat{\xi}_{v_u} > x | G_L) \right. \left| G_{d+1} \right) \) and \( E\left( P(\hat{\xi}_{w_u} > x | G_L) \right. \left| G_{d+1} \right) \) are independent. Thus,
\[
E\left( P(\hat{\xi}_{v_u} > x | G_L)P(\hat{\xi}_{w_u} > x | G_L) \right| G_{d+1}) = E\left( P(\hat{\xi}_{v_u} > x | G_L) | G_{d+1} \right) \cdot E\left( P(\hat{\xi}_{w_u} > x | G_L) | G_{d+1} \right).
\]
Using (68) and (69), by similar calculations as in (87) for \( v, w \) and \( u \) with 
\[ d(v) = \hat{j}, \quad d(w) = k \quad \text{and} \quad d(u) = d \]
respectively (where \( u \) is the closest ancestor to \( v \) and \( w \)), we get

\[
\sum_{d=1}^{L} \sum_{w:d(u)=d} \sum_{v,w\in T_u;} \sum_{v\wedge w=u} E\left(P\left(\xi_v > x|\mathcal{G}_L\right)P\left(\xi_w > x|\mathcal{G}_L\right)|\mathcal{G}_{d+1}\right)\hspace{1cm} (115)
\]

\[
\leq \sum_{d=1}^{L} \sum_{w:d(u)=d} \sum_{j=d}^{L} \sum_{v,w\in T_u; d(v)=j} E\left(\frac{1}{m} \ln + \frac{m \prod_{r=1}^{k} W_{r,v}}{x} \bigg| \mathcal{G}_{d+1}\right).
\]

\[
\sum_{d=1}^{L} \sum_{w:d(u)=d} \sum_{k=d}^{L} \sum_{v,w\in T_u; d(w)=k} E\left(\frac{1}{m} \ln + \frac{m \prod_{r=1}^{k} W_{r,w}}{x} \bigg| \mathcal{G}_{d+1}\right),
\]

where the inequality follows by applying (114) and using analogous calculations as in (111) equations (83)–(84). Note that the expected value of the left hand-side of the inequality in (115) is equal to the right hand-side of the inequality in (113). Let \( u \) be the closest ancestor vertex of \( v \) and \( w \). Let \( u_v \) be the child of \( u \) that is an ancestor of \( v \), respectively \( u_w \) be the child of \( u \) that is an ancestor of \( w \). Let \( \widehat{W}_{u,v} \) be the component in the split vector of vertex \( u \) that corresponds to the child \( u_v \) of \( u \), and use the analogous notation for \( \widehat{W}_{u,w} \). For a triple \( (v, w, u) \) with \( d(v) = j \), \( d(w) = k \) and \( d(u) = d \) we have

\[
\tilde{n}_v = n \prod_{r=1}^{j} W_{r,v} = n \widehat{W}_{u,v} \prod_{r=1}^{d} W_{r,u} \prod_{r=d+2}^{j} W_{r,v},
\]

\[
\tilde{n}_w = n \prod_{r=1}^{k} W_{r,w} = n \widehat{W}_{u,w} \prod_{r=1}^{d} W_{r,u} \prod_{r=d+2}^{k} W_{r,w}.
\]

(117)

For given \( d(u) = d \geq l \), \( d(v) = j \) and \( d(w) = k \), there are at most \( b^d \) choices of \( u \), and then at most \( b^{j-d} \) choices of \( v \) and \( b^{k-d} \) choices of \( w \). (We can assume that \( j > d + 1 \) and \( k > d + 1 \), since it is easy to see that the other terms are few and the sum of them is small.) For the child \( u_v \) of \( u \), \( d(u_v) = d + 1 \) and we have \( Y_{d+1,u_v} = -\sum_{r=1}^{d} \ln W_{r,u} - \ln \widehat{W}_{u,v} \) (and for the child \( u_w \) of \( u \), \( Y_{d+1,u_w} \) is defined in analogy). Recall the definition of \( Z_{j-l,v} := Y_{j,v} - Y_{d+1,u_v} \) in (106). For the vertex \( v \) with \( d(v) = j \) we have that \( Z_{j-d-1,v} := -\sum_{r=d+2}^{j} \ln W_{r,v} \) (and the analogous notation for \( Z_{k-d-1,v} \)). (For simplicity we skip the vertex index in the calculations below.) Thus, by
similar calculations as in (88) and (93) the sum in (115) is equal to

\[
\frac{1}{m^2} O\left( \sum_{d=1}^{L} \sum_{d(u)=d}^{L} \left( \sum_{j=d+2}^{L} \int_0^{\ln m - \ln x - Y_{d+1,w}} b_j^{-1} \mathbb{P}(Z_{j-d} \leq t) \right) \right).
\]

\[
\left( \sum_{k=d+2}^{L} \int_0^{\ln m - \ln x - Y_{d+1,w}} b_k^{-1} \mathbb{P}(Z_{k-d} \leq t) \right) = O\left( \sum_{d=1}^{L} \sum_{d(u)=d}^{L} \frac{1}{x^2} \prod_{j=1}^{d} W_j^2 \right).
\]

Since \( E(W_j^2) < \frac{1}{b} \) the expected value of this is \( o(1) \), and thus the right hand-side of the inequality in (113) is \( o(1) \). Hence, \( E(\text{Var}(S_1|G_l)) \) is \( o(1) \), which shows (65). We proceed by showing (66). Recall from (60) that

\[
S_2 = \sum_{l \leq d(v) \leq L} E(\xi_v 1[\xi_v \leq c] | G_L) - \frac{\mu^{-2} \ln^2 n}{n} \cdot \Phi_v,
\]

where

\[
\Phi_v = \frac{n}{\mu^{-1} \ln n} - \sum_{d(v)=L} \frac{n \prod_{r=1}^{L} W_{r,v} \sum_{r=1}^{L} \ln W_{r,v}}{\mu^{-1} \ln^2 n}.
\]

First we consider

\[
\text{Var}\left( \sum_{l \leq d(v) \leq L} E(\xi_v 1[\xi_v \leq c] | G_L) \right)
\]

\[
= \sum_{l \leq d(v) \leq L} \text{Cov}(E(\xi_v 1[\xi_v \leq c] | G_L), E(\xi_w 1[\xi_w \leq c] | G_L) | G_l).
\]  

(118)

As we argued for showing (65), we can suppose that the closest ancestor \( u \) for \( v \) and \( w \) is at depth \( d \geq l \). Similar to (112),

\[
E(\text{Cov}(E(\xi_v 1[\xi_v \leq c] | G_L)E(\xi_w 1[\xi_w \leq c] | G_L) | G_l))
\]

\[
\leq E(E(\xi_v 1[\xi_v \leq c] | G_L)E(\xi_w 1[\xi_w \leq c] | G_L)).
\]

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

\[
E(\xi_v 1[\xi_v \leq c] | G_L) = \frac{m \hat{n}_v}{n(m+1)} e^{-\frac{m+1}{m} \ln + \left( \frac{m \hat{n}_v}{n} \right)} \leq \frac{m \hat{n}_v}{n} = \prod_{r=1}^{j} W_{r,v}.
\]

41
Denote by \((v_u, w_u)\) a pair of vertices with closest ancestor \(u\) as in (113). Consider one such pair \((v_u, w_u)\), and let \(d(u) = d\), \(d(v) = j\) and \(d(w) = k\). Since \(E(W_{r,v}^2) < \frac{1}{b^{\delta \tau}}\) for some \(\delta > 0\) it follows that

\[
\begin{align*}
E\left(E\left(\xi_{v_u}{1}_{\xi_{w_u} \leq c} \mid \mathcal{G}_L\right) E\left(\xi_{w_u}{1}_{\xi_{w_u} \leq c} \mid \mathcal{G}_L\right)\right) & \leq C_1 E \prod_{r=1}^{d} W_{r,u}^2 b^{-j - (j - d) - (k - d)} \\
& \leq C_1 \left(\frac{1}{b^{\delta \tau}}\right)^d b^{-(j - d) - (k - d)}, \\
(119)
\end{align*}
\]

where \(C_1\) is a constant depending on \(E(W_{u,v} \bar{W}_{u,w})\), where \(\bar{W}_{u,v}\) and \(\bar{W}_{u,w}\) are the random variables that we introduced for (117). Thus, by using (118)–(119), and as in (115) letting \(v, w : v \wedge w = u\) denote that the vertices \(v, w\) have closest ancestor \(u\) we get

\[
\begin{align*}
E\left(\text{Var}\left( \sum_{d = l}^{L} E\left(\xi_v \mid \mathcal{G}_L\right) \right) \mid \mathcal{G}_L\right) & \leq \sum_{d = l}^{L} \sum_{v : d(v) = d} \sum_{v, w \in T_u : v \wedge w = u} E\left(E\left(\xi_v \mid \mathcal{G}_L\right) \right) \right) \\
& \leq C_1 \sum_{d = l}^{L} b^{-\delta d} \sum_{j = d}^{L} b^{j - d - (j - d)} \sum_{k = d}^{L} b^{k - d - (k - d)} \leq C_2 L^2 b^{-\delta l} \to 0, \\
(120)
\end{align*}
\]

where \(C_2\) is a constant. (Compare with the calculations in [10, equation (87)].) We now show that

\[
E\left(\text{Var}\left( \frac{\mu^2 \ln^2 n}{n} \cdot \Phi \right) \mid \mathcal{G}_L\right) \to 0. \\
(121)
\]

To show this, it is enough to show that

\[
E\left(\text{Var}\left( \sum_{v : d(v) = L} \prod_{r=1}^{L} W_{r,v} \sum_{r=1}^{L} \ln W_{r,v} \mid \mathcal{G}_L\right) \right) \to 0.
\]

Using (120), we obtain for each \(s \leq L\),

\[
E\left(\text{Var}\left( \sum_{v : d(v) = L} \prod_{r=1}^{L} W_{r,v} \ln W_{s,v} \mid \mathcal{G}_L\right) \right) = O\left(L^2 b^{-\delta l}\right).
\]
Thus, the conditional Hölder inequality, see e.g., [9, p. 476], yields (121). From (120) and (121) and again applying the conditional Hölder inequality we deduce that $E\left( \text{Var}(S_2|\mathcal{G}_1) \right)$ is o(1), which shows (66).

Recall from (61) that

$$S_3 = \sum_{l \leq d(v) \leq L} \text{Var}(\hat{\xi}_v 1[\hat{\xi}_v \leq c]|\mathcal{G}_L).$$

It remains to show that $E\left( \text{Var}(S_3|\mathcal{G}_1) \right)$ is o(1). To show this we observe that

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

and thus (67) follows from (119) by similar calculations as in (120).

\[ \square \]

**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] S. Asmussen, *Applied Probability and Queues*. John Wiley Sons, Chichester, 1987.

[2] J. Bourdon, Size and path length of Patricia tries: dynamical sources context. *Random Structures Algorithms* 19 (2001), no. 3-4, 289–315.

[3] L. Devroye, Universal limit laws for depths in random trees. *Siam J. Comput.* 28 (1998), no 2, 409–432.

[4] L. Devroye, Applications of Stein’s method in the analysis of random binary search trees. *Stein’s Method and Applications*, 47–297 (ed. Chen, Barbour) Inst. for Math. Sci. Lect. Notes Ser. 5, World Scientific Press, Singapore, 2005.

[5] M. Drmota, A. Iksanov, M. Moehle, U. Roesler, A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Struct. Alg.* 34 (2009), 319–336.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II*. 2nd ed., Wiley, New York, 1971.
[7] J. A. Fill, S. Janson, Quicksort asymptotics. *J. Algorithms* **44** (2002), 4–28.

[8] A. Gut, *Stopped Random Walks*. Springer Verlag, New York, Berlin, Heidelberg, 1988.

[9] A. Gut, *Probability: A Graduate Course*, Springer, New York, 2005.

[10] C. Holmgren, Random records and cuttings in binary search trees. Accepted in *Combinat. Probab. Comput.* (2009).

[11] C. Holmgren, Novel characteristics of split trees by use of renewal theory. *Submitted for publication*.

[12] A. Iksanov, M. Möhle, 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** (2007), 28–35.

[13] S. Janson, Random records and cuttings in complete binary trees. *Mathematics and Computer Science III* Birkhäuser, Basel (2004), 241–253.

[14] S. Janson, Random cuttings and records in deterministic and random trees. *Random Struct. Alg.* **29** (2006), 139–179.

[15] S. Janson, T. Łuczak, A. Rucinski, *Random Graphs*, Wiley, New York, 2000.

[16] O. Kallenberg, *Foundations of Modern Probability*. 2nd ed., Springer Verlag, Reading, Mass., 2002.

[17] H. Mahmoud, On the average internal path length of $m$-ary search trees. *Acta Inform.* **23** (1986), 111–117.

[18] H. Mahmoud, B. Pittel, Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms** 10** (1989), no. 1, 52–75.

[19] A. Meir, J. W. Moon, Cutting down random trees. *J. Australian Math. Soc.* **11** (1970), 313–324.

[20] R. Neininger and L. Rüschendorf, On the internal pathlength of $d$-dimensional quad trees. *Random Struct. Alg.* **15** (1999), no. 1, 25–41.

[21] U. Roesler, On the analysis of stochastic divide and conquer algorithms. Average-case analysis of algorithms (Princeton, NJ, 1998), *Algorithmica* **29** (2001), no. 1-2, 238–261.