Electronic Journal of Probability

Electron. J. Probab. **17** (2012), no. 5, 1–27.
ISSN: 1083-6489 DOI: 10.1214/EJP.v17-1723

Novel characteristics of split trees by use of renewal theory

Cecilia Holmgren*

Abstract

We investigate characteristics of random split trees introduced by Devroye [SIAM J Comput **28**, 409–432, 1998]; split trees include e.g., binary search trees, m-ary search trees, quadtrees, median of $(2k+1)$-trees, simplex trees, tries and digital search trees. More precisely: We use renewal theory in the studies of split trees, and use this theory to prove several results about split trees. A split tree of cardinality n is constructed by distributing n balls (which often represent data) to a subset of nodes of an infinite tree. One of our main results is a relation between the deterministic number of balls n and the random number of nodes N. In [5] there is a central limit law for the depth of the last inserted ball so that most nodes are close to depth $\frac{\ln n}{\mu} + \mathcal{O}(\sqrt{\ln n})$, where μ is some constant depending on the type of split tree; we sharpen this result by finding an upper bound for the expected number of nodes with depths $\geq \frac{\ln n}{\mu} - \ln^{\frac{1}{2}+\epsilon} n$ or depths $\leq \frac{\ln n}{\mu} + \ln^{\frac{1}{2}+\epsilon} n$ for any choice of $\epsilon > 0$. We also find the first asymptotic of the variances of the depths of the balls in the tree.

Keywords: Random Trees; Split Trees; Renewal Theory.
AMS MSC 2010: Primary 05C05; 05C80; 68W40; 68P10, Secondary 68R10; 60C05; 68P05.
Submitted to EJP on June 15, 2010, final version accepted on August 9, 2011.
Supersedes arXiv:1005.4594v1.

1 Introduction

In this paper we use renewal theory as a powerful tool to gain results regarding (random) split trees (introduced by Devroye [5]). The split trees constitute a large class of random trees of logarithmic height, i.e., there exists a constant C such that $\mathbf{P}(\frac{H_n}{\log n} > C) \to 0$, where H_n is the height (maximal depth) of the tree. Some important examples of split trees are binary search trees [14], m-ary search trees [17], quadtrees [10], median of $(2k+1)$-trees [2], simplex trees, tries [11] and digital search trees [4].

1.1 Preliminaries

In this subsection we introduce the split tree model as defined by Devroye. We also give some background and state a proposition concerning the depth of balls.

*Department of Mathematics, Uppsala University, Sweden. Present address: DPMMS, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK. E-mail: C.Holmgren@dpmms.cam.ac.uk

1.1.1 The Split Tree Model

The formal definition of split trees is given in the "split tree generating algorithm" below. To facilitate the penetration of this rather complex algorithm we first provide a brief heuristic description. A skeleton tree S_b of branch factor b is an infinite rooted tree in which each node has exactly b children. A split tree is a finite subtree of a skeleton tree S_b. The split tree is constructed iteratively by distributing balls one at a time to a subset of nodes of S_b. We say that the tree has cardinality n if n balls are distributed. Since many of the common split trees come from algorithms in Computer Science the balls often represent some "keys" or other data symbols. There is also a so-called node capacity, $s > 0$, which means that each node can hold at most s balls. We say that a node v is a leaf if v itself holds at least one ball but no descendants of v hold any balls. The split tree consists of the leaves and all the ancestors of the leaves. See Figure 1 and Figure 2, which illustrate two split trees (the parameters s_0 and s_1 in the figures are introduced in the formal algorithm).

The first ball is placed in the root of S_b. Each new ball is added by starting at the root, and then letting the ball fall down in the tree until it reaches a leaf. Each node v of S_b is given an independent copy of the so-called random split vector $\mathcal{V} = (V_1, V_2 \ldots, V_b)$ of probabilities, where $\sum_i V_i = 1$ and $V_i \geq 0$. The split vectors control the path that the ball takes until it reaches a leaf; when the ball falls from node v to one of its children, it chooses the ith child of v with probability V_i, i.e., the ith component of the split vector associated to v. When a full leaf gets a new ball it splits; hence, some of the $s+1$ balls are given to its children, leading to new leaves. When all the n balls have been added we get a split tree with a finite number of nodes which we denote by the parameter N.

The split tree generating algorithm: The (random) split tree has the parameters b, n, s and $\mathcal{V}$ as we described above; there are also two other parameters: s_0, s_1 (related to the parameter s) that occur in the algorithm. Let n_v denote the total number of balls that the nodes in the subtree rooted at node v hold together, and C_v be the number of balls that are held by v itself. Note that v is a leaf if and only if $C_v = n_v > 0$ and that a node $v \in S_b$ is included in the split tree if, and only if, $n_v > 0$.

Initially there are no balls, i.e., $C_v = 0$ for each node v. Choose an independent copy $\mathcal{V}_v$ of $\mathcal{V}$ for every node $v \in S_b$. Add balls one by one to the root by the following iterative 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 $C_v = n_v < s$, add the ball to v and stop. Thus, C_v and n_v increase by 1.
3. If v is a leaf and $C_v = n_v = s$, there is no space for the new ball at v. 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.

Once the original n balls are distributed the algorithm stops.

From step 3, it follows that s_0 and s_1 have to satisfy the inequality $0 \leq bs_1 \leq s+1-s_0$. Note that if $s_0 > 0$ or $s_1 > 0$, step 3 does not need to be repeated in this iteration of the procedure since no child could reach the capacity s, whereas if $s_0 = s_1 = 0$ step 3 may have to be repeated several times. Note that every nonleaf has $C_v = s_0$ and every leaf has $0 < C_v \leq s$. The algorithm gives a recursive construction of the subtree sizes

$n_v,\ v \in S_b$. The tree consists of n items and the root σ has $n_\sigma = n$. Given the cardinality n_v and the split vector $\mathcal{V}_v = (V_1, \ldots, V_b)$ the cardinalities $(n_{v_1}, \ldots, n_{v_b})$ of the b subtrees rooted at $v_1, \ldots, v_b$ are distributed as

$$\mathrm{Mult}(n_v - s_0 - bs_1, V_1, V_2, \ldots, V_b) + (s_1, s_1, \ldots, s_1). \tag{1.1}$$

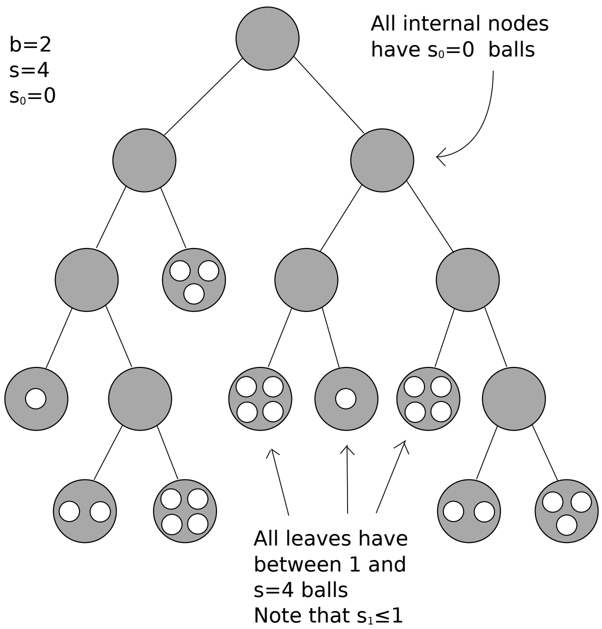

Figure 1: A split tree with $b = 2,\ s = 4,\ s_0 = 0$ and $s_1 \leq 1$.

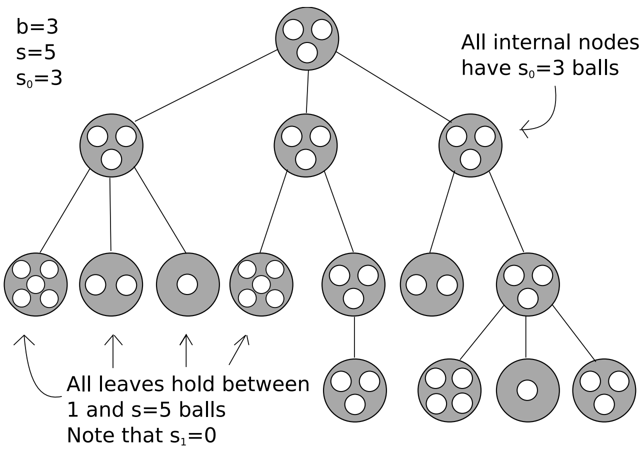

Figure 2: A split tree with $b = 3,\ s = 5,\ s_0 = 3$ and $s_1 = 0$.

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, see [5]. Let V be a random variable with this distribution; hence $\mathbf{E}(V) = \frac{1}{b}$. We use the notation T^n to denote a split tree with n balls. Note that even conditioned on the fact that the split tree has n balls, the number of nodes N, is still usually random.

Example 1: Binary Search Tree The binary search tree is the graph of one of the most used sorting algorithm Quicksort: Draw a (uniformly) random key κ_σ (a number) from the set $\{1, 2 \ldots, n\}$, and associate it to the root σ. Then sort the other keys into two sets, where the keys that are smaller than κ_σ are sent to the left child and the keys that are larger are sent to the right child. The sizes of the two subtrees of the root are $\kappa_\sigma - 1$ and $n - \kappa_\sigma$. Since κ_σ is equally likely to be $\{1, 2, \ldots, n\}$, one has

$$(\kappa_\sigma - 1, n - \kappa_\sigma) \stackrel{d}{=} \mathrm{Mult}(n-1; U, 1-U),$$

where U is a uniform $U(0,1)$ random variable and $\stackrel{d}{=}$ denotes distributional equality . Thus, a binary search tree is a split tree with $b = 2$, $s_0 = 1$, $s = 1$, $s_1 = 0$ and $V \stackrel{d}{=} U$.
Example 2: Tries Let $X_1, \ldots, X_n$ be n infinite strings on the alphabet $\{1, \ldots, b\}$. The strings are drawn independently, and the symbols of each string are also independent with distribution on $\{1, \ldots, b\}$ given by $p_1, \ldots, p_b$. Each string naturally corresponds to an infinite path in S_b: symbol $i \in \{1, \ldots, b\}$ is associated to the ith child. The trie is then defined as the minimal subtree so that the paths corresponding to the infinite strings are distinct. The internal nodes store no data, each leaf stores a unique string. For tries, n_v corresponds to the number of strings that have the first d symbols up to node v at depth d in common; for all internal nodes $n_v > 1$ and for the leaves $n_v = 1$. One clearly has for the b children of the root

$$(n_1, \ldots, n_b) \stackrel{d}{=} \mathrm{Mult}(n; p_1, \ldots, p_b).$$

The trie is thus a split tree with $s = 1, s_0 = s_1 = 0$ and $\mathcal{V}$ is a random permutation of $(p_1, p_2, \ldots, p_b)$.

1.1.2 A weak law and a central limit law for the depth

Recall that V is a random variable with the distribution of the identically distributed components $V_i,\ i \in \{1, \ldots, b\}$ in the split vector $\mathcal{V} = (V_1, \ldots, V_b)$. Let $\Delta = V_S$ be the size biased distribution of $(V_1, \ldots, V_b)$, i.e., given $(V_1, \ldots, V_b)$, let $\Delta = V_j$ with probability V_j, see [5]. Let

$$\mu := \mathbf{E}(-\ln\Delta) = b\mathbf{E}(-V\ln V), \quad \sigma^2 := \mathbf{Var}(\ln\Delta) = b\mathbf{E}(V\ln^2 V) - \mu^2. \tag{1.2}$$

Note that the second equalities of μ and σ imply that they are bounded. Similarly all moments of $-\ln\Delta$ are bounded.

In [5] Devroye presented a weak law of large numbers and a central limit law for D_n (depth of the last inserted ball). Devroye [5, Theorem 1] showed that if $\mathbf{P}(V = 1) = 0$, then $\frac{D_n}{\ln n} \xrightarrow{p} \mu^{-1}$ and

$$\frac{\mathbf{E}(D_n)}{\ln n} \to \mu^{-1}. \tag{1.3}$$

Let $D_{k,n}$ be the depth of the kth inserted ball when $n \geq k$ balls have been added; in particular $D_n = D_{n,n}$. Let D_n^* be the average depth in a tree with n balls, i.e., $D_n^* = \frac{1}{n}\sum_{k=1}^{n} D_{k,n}$. Note that $D_k \leq D_{k,n},\ n \geq k$, since ball k can move during the splitting process when new balls are added to the tree. From the following Proposition it simply follows that (1.3) also holds for D_n^*.

Proposition 1.1. *For $i \leq j$, we have that $D_{i,n} \leq D_{j,n}$ in the stochastic sense.*

This proposition is shown in Section 3.3.

Corollary 1.1. *For the average depth D_n^*, we have*

$$\frac{\mathbf{E}(D_n^*)}{\ln n} \to \mu^{-1}, \quad \text{as } n \to \infty. \tag{1.4}$$

Proof. Proposition 1.1 implies that for all $k \le n$,

$$\mathbf{E}(D_k) \le \mathbf{E}(D_{k,n}) \le \mathbf{E}(D_n). \tag{1.5}$$

By applying (1.3) and (1.5) we get $\frac{1}{n}\sum_{k=1}^{\lfloor \frac{n}{\ln^2 n} \rfloor} \mathbf{E}(D_{k,n}) = o(1)$, and for $k \ge \frac{n}{\ln^2 n}$ we have $\mathbf{E}(D_{k,n}) \sim \mu^{-1}\ln n$. Hence, $\frac{\mathbf{E}(D_n^*)}{\ln n} \to \mu^{-1}$. □

For $\sigma > 0$, and assuming that $\mathbf{P}(V = 1) = 0$ and that V is not monoatomic, i.e., $\mathcal{V} \neq (\frac{1}{b}, \dots, \frac{1}{b})$ Devroye showed [5, Theorem 1]

$$\frac{D_n - \mu^{-1}\ln n}{\sqrt{\sigma^2\mu^{-3}\ln n}} \xrightarrow{d} N(0,1), \tag{1.6}$$

where $N(0,1)$ is the normal distribution. Tries, see Example 2 above, are split trees with a random permutation of deterministic components $(p_1, p_2, \dots, p_b)$ and therefore not as random as many other examples. Digital search trees are closely related to tries; they also have split vector $(p_1, p_2, \dots, p_b)$ however their internal nodes are not empty. Of all the most common examples of split trees only the symmetric tries and symmetric digital search trees (i.e., with $p_1 = p_2, \dots = p_b$) have a monoatomic distribution of V. From (1.6) it follows that "most" nodes lie at depth $\mu^{-1}\ln n + \mathcal{O}(\sqrt{\ln n})$.

1.2 Main Results

In this section we present the main theorems of this work. Since we use renewal theory in our proofs it is necessary to distinguish between lattice and non-lattice distributions. This is the reason for the non-lattice assumption (A1) below.

(A1). We assume as in Section 1.1.2 that $\mathbf{P}(V = 1) = 0$, and for simplicity we also assume that $-\ln V$ has a non-lattice distribution.

Note that the assumption that V is not monoatomic in Section 1.1.2 is included in assumption (A1). Again of the common split trees only for some special cases of tries and digital search trees does $-\ln V$ have a lattice distribution.

Our first most important result is on the relation between the number of nodes N (recall that this is a random variable) and the number of balls n.

Theorem 1.1. *There is a constant α, depending on the type of split tree, such that*

$$\mathbf{E}(N) = \alpha n + o(n), \tag{1.7}$$

and

$$\mathbf{Var}(N) = o(n^2). \tag{1.8}$$

For specific cases of split trees the constant α in (1.7) can be calculated explicitly, see e.g.,[16] for m-ary search trees and [3, 18] for non-lattice cases of tries.

Recall that there is a central limit law for D_n (the depth of the last ball) in (1.6) so that most nodes are close to depth $\frac{\ln n}{\mu} + \mathcal{O}(\sqrt{\ln n})$ (where μ is the constant in (1.2)); our next result sharpens this result. Given a constant $\epsilon > 0$, we say that a node v in T^n is *good* if

$$\frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon} \le d(v) \le \frac{\ln n}{\mu} + (\ln n)^{\frac{1}{2}+\epsilon},$$

and *bad* otherwise.

Theorem 1.2. *For any choice of* $\epsilon > 0$ *and for any constant* k *the expected number of bad nodes in* T^n *is* $\mathcal{O}\left(\frac{n}{\ln^k n}\right)$.

In the third main result we sharpen the limit laws in (1.3) and (1.4) for the expected values of D_n and D_n^* (the average depth). We also find the first asymptotic of the variances of the kth ball $D_{k,n}$ for all k, $\frac{n}{\ln n} \leq k \leq n$.

Theorem 1.3. *Let* μ *and* σ^2 *be the constants in (1.2). For the expected value of the depth of the last ball we have*

$$\frac{\mathbf{E}(D_n) - \mu^{-1} \ln n}{\sqrt{\ln n}} \to 0, \quad \text{as } n \to \infty \tag{1.9}$$

and the same result holds for the average depth D_n^*, *i.e.,*

$$\frac{\mathbf{E}(D_n^*) - \mu^{-1} \ln n}{\sqrt{\ln n}} \to 0, \quad \text{as } n \to \infty. \tag{1.10}$$

Furthermore, for all $\frac{n}{\ln n} \leq k \leq n$ *the variance of the depth of the* k*th ball satisfies*

$$\frac{\mathbf{Var}(D_{k,n})}{\ln n} \to \sigma^2 \mu^{-3}, \quad \text{as } n \to \infty. \tag{1.11}$$

1.3 Notation

In this section some of the notation that we use in the present study is collected.

Let T^n denote a split tree with n balls; for simplicity we sometimes write T. Let N be the number of nodes in T^n. Let $d(v)$ denote the depth of a node v, sometimes we just write d for the depth of v. Recall that D_n is the depth of the last ball and that D_n^* is the average depth, and that we write $D_{k,n}$ for the depth of the kth inserted ball when $n \geq k$ balls have been added. Let T_v be a subtree rooted at v. We write n_v for the number of balls in T_v and we write N_v for the number of nodes.

We use the standard notation, $N(\mu, \sigma^2)$ for a normal distribution with expected value μ and variance σ^2, and $\text{Bin}(m, p)$ for a binomial distribution with parameters m and p. We also use the notation mixed binomial distribution (X, Y) or for short $\text{mBin}(X, Y)$ for a binomial distribution where at least one of the parameters X and Y is a random variable (the other one could be deterministic). We write $\stackrel{d}{=}$ for equality in distribution and $\leq_{st}$ respectively $\geq_{st}$ for inequality in the stochastic sense. We write $|S|$ for the number of elements in a set S.

We use the standard notation $g(k) = \mathcal{O}(f(k))$ to denote that there exist constants C and k_0 such that $|g(k)| \leq Cf(k)$ for $k \geq k_0$; in fact the qualifier $k \geq k_0$ is not necessary as long as $f(k) > c > 0$ for a given constant $c > 0$ and $g(k),\ k \in [0, k_0]$ is bounded, since C can be replaced by $C' = \max(C, \sup_{k \leq k_0} |g(k)/f(k)|) < \infty$. Furthermore, to simplify the discussion, when we use the notation $\mathcal{O}(f(n, \epsilon))$ as n tends to infinity, we ensure that the hidden constant in $\mathcal{O}$ does not depend on ϵ. More precisely, when we write $g(n, \epsilon) = \mathcal{O}(f(n, \epsilon))$ we mean that

$$\exists C,\ \forall \epsilon > 0,\ \ \exists n_\epsilon,\ \ n \geq n_\epsilon \ \ :\Rightarrow\ |g(n, \epsilon)| \leq Cf(n, \epsilon).$$

We define Ω_d as the σ-field generated by $\{n_v,\ d(v) \leq d\}$. Finally, we write $\mathscr{G}_d$ for the σ-field generated by the $\mathcal{V}$-vectors for all v, $d(v) \leq d$.

1.4 Applying Renewal Theory to Split Trees

Let v be a node at depth d, conditioning on $\mathscr{G}_d$ (i.e., the σ-field generated by the $\mathcal{V}$ vectors for all nodes v with $d(v) \leq d$) and applying the fact that a $\text{mBin}(X, p_1)$ in which

X is $\mathrm{Bin}(m, p_2)$ is distributed as a $\mathrm{Bin}(m, p_1p_2)$, we get from (1.1) that

$$n_v \leq_{st} \mathrm{Bin}(n, \prod_{j=1}^{d} V_{j,v}) + \mathrm{Bin}(s_1, \prod_{j=2}^{d} V_{j,v}) + \cdots + \mathrm{Bin}(s_1, V_{d,v}) + s_1, \tag{1.12}$$

where $V_{j,v}, j \in \{1, \ldots, d\}$ are i.i.d. random variables $V_{j,v} \stackrel{d}{=} V$, given by the split vectors associated with the nodes in the unique path from v to the root. (Note that equivalently, $\mathscr{G}_d$ is the σ-field generated by $V_{j,v}, j \in \{1, \ldots, d\}$ for all v with $d(v) = d$.) Note that the terms in (1.12) are not independent. Similarly, we have a lower bound for n_v, i.e., for v at depth d, conditioning on $\mathscr{G}_d$,

$$n_v \geq_{st} \mathrm{Bin}(n, \prod_{j=1}^{d} V_{j,v}) - \mathrm{Bin}(s, \prod_{j=2}^{d} V_{j,v}) - \cdots - \mathrm{Bin}(s, V_{d,v}); \tag{1.13}$$

we can replace the term s by $s_0 + bs_1 \leq s$ for a sharper bound.

Recall that for a $\mathrm{Bin}(m, p)$ distribution, the expected value is mp and the variance is $mp(1-p)$. Thus, Chebyshev's inequality applied to the dominating term $\mathrm{Bin}(n, \prod_{j=1}^{d} V_{j,v})$ in (1.12) gives that n_v for v at depth d is close to

$$M_v^n := nV_{1,v}V_{2,v}\ldots V_{d,v}; \tag{1.14}$$

since the n_v's (conditioned on the split vectors) for all v at the same depth are identically distributed, we sometimes skip the node index of $V_{j,v}$ and just write V_j. We now state a more precise relation between n_v and M_v^n.

Lemma 1.1. *For any node v, we have for all n large enough that*

$$\mathbf{P}\big(\mid n_v - M_v^n \mid > n^{0.6}\big) \leq \frac{1}{n^{0.1}}. \tag{1.15}$$

Proof. By using (1.12) and (1.13), the Chebyshev and Markov inequalities give for v with $d(v) = d$, that for large n,

$$\begin{aligned}\mathbf{P}\big(\mid n_v - M_v^n \mid > n^{0.6}\big) &\leq 4\frac{\mathbf{E}\big(\mathbf{Var}\big(\mathrm{Bin}(n, \prod_{j=1}^{d} V_j)\big|\mathscr{G}_d\big)\big)}{n^{1.2}} \\ +4\mathbf{E}\big(\mathbf{E}\big(\mathrm{Bin}(s, \prod_{j=2}^{d} V_j) &+ \mathrm{Bin}(s, \prod_{j=3}^{d} V_j) + \cdots + s\big|\mathscr{G}_d\big)\big)/n^{0.6} \\ &\leq \frac{4nb^{-d}}{n^{1.2}} + \frac{\sum_{k=1}^{\infty} 4sb^{-k}}{n^{0.6}} \leq \frac{1}{n^{0.1}}.\end{aligned}$$

□

Renewal theory is a widely used branch of probability theory that generalizes Poisson processes to arbitrary holding times. A classic in this field is Feller [7] on recurrent events. First we recall some standard notation. Let $X_0 = 0$ a.s.. Let $X_k,\ k \geq 1$, be i.i.d. nonnegative random variables distributed as X and let $S_m,\ m \geq 1$, be the partial sums. We write F for the distribution function of X and F_m for the distribution function of $S_m,\ m \geq 0$. Thus, for $x \geq 0$,

$$F_0(x) = 1, \quad F_1(x) = F(x), \quad F_m(x) = F^{m^*}(x),$$

i.e., F_m equals the m-fold convolution of F. The *renewal counting process* $\{\mathcal{N}(t),\ t \geq 0\}$ is defined by $\mathcal{N}(t) := \max\{m : S_m \leq t\}$. In the specific case when $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$, then

$\{\mathcal{N}(t),\ t \geq 0\}$ is a Poisson process. An important well studied function is the *standard renewal function*

$$V(t) := \sum_{m=0}^{\infty} F_m(t) = \mathbf{E}(\mathcal{N}(t)). \tag{1.16}$$

The renewal function $V(t)$ satisfies the so-called *renewal equation*

$$V(t) = 1 + (V * dF)(t), \quad t \geq 0.$$

For a broader introduction to renewal theory, see e.g. [1], [8], [9] and [12].

One of the main purposes of this study is to use renewal theory to study split trees. Renewal theory has also been used in [18] to study tries and similar split tree structures for which the parameters $s_0 = s_1 = 0$. Recall that n_v is close to the product M_v^n in (1.14). Now let $Y_k := -\sum_{j=1}^{k} \ln V_j$. Note that $M_v^n = ne^{-Y_k}$. For the specific case of the binary search tree the sum Y_k, (where $V_j,\ j \in \{1, \dots, k\}$, in this case are i.i.d. uniform $U(0,1)$ random variables) is distributed as a $\Gamma(k,1)$ random variable; this fact was used by, e.g., Devroye [6] to determine the height of the tree. For general split trees, for which we don't know the common distribution function of Y_k, renewal theory can be used instead. Let $\nu_k(t) := b^k \mathbf{P}(Y_k \leq t)$. We define the renewal function

$$U(t) := \sum_{k=1}^{\infty} \nu_k(t). \tag{1.17}$$

Let $\nu(t) := \nu_1(t) = b\mathbf{P}(-\ln V_j \leq t)$. For $U(t)$ we obtain the following renewal equation

$$U(t) = \nu(t) + \sum_{k=1}^{\infty} (\nu_k * d\nu)(t) = \nu(t) + (U * d\nu)(t). \tag{1.18}$$

2 Some Fundamental Renewal Theory Results

The main goal of this section is to present a renewal theory lemma and a corollary of this lemma, which are both frequently used in this study. In contrast to standard renewal theory the distribution function $\nu(t)$ in (1.18) is not a probability measure. However, to solve (1.18) we can apply [1, Theorem VI.5.1] which deals with non probability measures to deduce the following result.

Lemma 2.1. *The renewal function $U(t)$ in (1.17) satisfies*

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

Proof. Since the distribution function $\nu(t)$ is not a probability measure, we define another ("conjugate" or "tilted") measure ω on $[0, \infty)$ by

$$d\omega(t) = e^{-t} d\nu(t).$$

Recall from Section 1.1.2 that $\Delta = V_S$ is the size biased distribution of $(V_1, \dots, V_b)$. Note that $\omega(x)$ is the distribution function (and therefore a probability measure) of $-\ln \Delta$ since

$$\mathbf{P}(-\ln \Delta \leq x) = \mathbf{E}(\mathbf{E}(I\{-\ln V_S \leq x\} | (V_1, \dots, V_b))) = \mathbf{E}(\sum_{i=1}^{b} I\{-\ln V_i \leq x\} V_i) = \omega(x).$$

Hence, (recalling $\mu = \mathbf{E}(-\ln \Delta)$ and $\sigma^2 = \mathbf{Var}(-\ln \Delta)$) we have

$$\mathbf{E}(\omega) = \mu, \quad \text{and } \mathbf{Var}(\omega) = \sigma^2. \tag{2.1}$$

Define $\widehat{U}(t) := e^{-t}U(t)$ and $\widehat{\nu}(t) := e^{-t}\nu(t)$. We shall apply [1, Theorem VI.5.1], but first we need to show that the condition that $\widehat{\nu}(t)$ is "directly Riemann integrable" (d.R.i.) is satisfied. Note that $\widehat{\nu}(t) \leq be^{-t}$, and thus since $\widehat{\nu}(t)$ is also continuous almost everywhere, by [1, Proposition IV.4.1.(iv)] it follows that $\widehat{\nu}(t)$ is d.R.i. if be^{-t} is d.R.i.. That be^{-t} is d.R.i. follows by applying [1, Proposition IV.4.1.(v)], since be^{-t} is a nonincreasing and Lebesgue integrable function. Then by applying [1, Theorem VI.5.1] and (2.1) we get

$$\widehat{U}(t) = \widehat{\nu}(t) + (\widehat{U} * d\omega)(t), \tag{2.2}$$

where $\omega(t)$ is a probability measure, and

$$\widehat{U}(t) \to \mu^{-1} \int_0^\infty \widehat{\nu}(x)dx = \mu^{-1} \int_0^\infty \nu(x)e^{-x}dx =: \kappa. \tag{2.3}$$

Integration by parts now gives

$$\kappa = \mu^{-1}\big(b| - e^{-t}\mathbf{P}(-\ln V \leq t)\big|_0^\infty - \int_0^\infty -e^{-t}d\nu(t)\big) = \mu^{-1}b\mathbf{E}(e^{\ln V}) = \mu^{-1}, \tag{2.4}$$

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

The following result is a useful corollary of Lemma 2.1. Recall that we write

$$M_v^n = n \prod_{j=1}^{d(v)} V_j.$$

Corollary 2.1. *Let $K_n,\ n \geq 1$ be a sequence such that as $n \to \infty$ we have that $\frac{n}{K_n} \to \infty$. Then for the expected number of nodes with $M_v^n \geq K_n$ we have*

$$\mathbf{E}(\left|v \in T^n;\ M_v^n \geq K_n\right|) = U(\ln n - \ln K_n) + 1 = (\mu^{-1} + o(1))\frac{n}{K_n}, \quad \text{as } n \to \infty.$$

Proof. Lemma 2.1 gives

$$\mathbf{E}(\left|v \in T^n;\ M_v^n \geq K_n\right|) = \sum_{d=0}^{\infty} b^d \mathbf{P}\big(M_v^n \geq K_n\big)$$
$$= \sum_{d=0}^{\infty} b^d \mathbf{P}\big(Y_d \leq \ln n - \ln K_n\big) = (\mu^{-1} + o(1))\frac{n}{K_n}.$$

□

We complete this section with a more general result in renewal theory, and a corollary of a more specific result that is valid for the renewal function $U(t)$ in (1.17).

Theorem 2.1. *Let F be a non-lattice probability measure and suppose that we have $0 < \mu = \mathbf{E}(X) = \int_0^\infty x dF(x) < \infty$ and $\mathbf{E}(X^2) = \sigma^2 + \mu^2 < \infty$. Let*

$$Z(t) = z(t) + \int_0^t Z(t-u)dF(u), \quad t \geq 0,$$

where $z(t)$ is a nonnegative function, such that $a := \int_0^\infty z(u)du < \infty$. Define

$$G(x) = \int_0^x \big(Z(t) - \frac{a}{\mu}\big)dt.$$

Then

$$\lim_{x \to \infty} G(x) = -\frac{1}{\mu}\int_0^\infty uz(u)du + a\frac{\sigma^2 + \mu^2}{2\mu^2}. \tag{2.5}$$

Proof. Let $V(t)$ be the standard renewal function in (1.16), where

$$F_m(t) = \mathbf{P}\big(\sum_{k=0}^{m} X_k \leq t\big).$$

By applying [1, Theorem IV.2.4],

$$Z(t) = \int_0^t z(t-u)dV(u) = \int_0^\infty z(u)dV(t-u), \tag{2.6}$$

where the last equality follows because $V(t) = 0$ for $t \leq 0$. By applying (2.6) and Fubini's Theorem we get

$$G(x) = \int_0^\infty z(u) \int_0^x dV(t-u)du - \frac{ax}{\mu} = \int_0^\infty z(u)V(x-u)du - \frac{ax}{\mu}.$$

Hence,

$$\begin{aligned} G(x) &= \int_0^\infty z(u)\big(V(x-u) - \frac{x}{\mu}\big)du \\ &= -\frac{1}{\mu}\int_0^x z(u)udu - \frac{1}{\mu}\int_x^\infty z(u)xdu + \int_0^x z(u)\big(V(x-u) - \frac{x-u}{\mu}\big)du. \end{aligned} \tag{2.7}$$

From [1, Proposition VI.4.1] we have $V(t) - \frac{t}{\mu} \to \frac{\sigma^2+\mu^2}{2\mu^2}$ and by [1, Proposition VI.4.2], $0 \leq V(t) - \frac{t}{\mu} \leq \frac{\sigma^2+\mu^2}{\mu^2}$. Hence, the Lebesgue dominated convergence theorem applied to the last integral in (2.7) gives

$$\lim_{x\to\infty} \int_0^\infty z(u)\big(V(x-u) - \frac{x-u}{\mu}\big)I\{u \leq x\}du = \int_0^\infty z(u)\frac{\sigma^2+\mu^2}{2\mu^2}du.$$

Note that for all x, $\int_x^\infty z(u)(u-x)du \geq 0$. Thus, if $\int_0^\infty z(u)udu$ is integrable, then

$$\lim_{x\to\infty} \int_x^\infty z(u)xdu = 0,$$

and the convergence result in (2.5) obviously follows. If $\int_0^\infty z(u)udu$ is not integrable then we have a special case of (2.5), i.e., $\lim_{x\to\infty} G(x) = -\infty$. □

We define the function

$$W(x) = \int_0^x e^{-t}(U(t) - \mu^{-1}e^t)dt. \tag{2.8}$$

The next result is a corollary of Theorem 2.1 which we apply in [15].

Corollary 2.2. *The function $W(x)$ in (2.8) satisfies*

$$W(x) = \frac{\sigma^2 - \mu^2}{2\mu^2} - \mu^{-1} + o(1), \quad as\ x \to \infty. \tag{2.9}$$

Proof. We apply Theorem 2.1 to $Z(t) = \widehat{U}(t) = e^{-t}U(t)$ defined in the proof of Lemma 2.1 (recall that $\widehat{U}(t)$ satisfies the renewal equation in (2.2)). Now, the constant a as defined in Theorem 2.1, satisfies $a = \int_0^\infty \widehat{\nu}(u)du$, thus (2.3) and (2.4) give $a = 1$. From (2.1) and (2.3)–(2.4) we get

$$\int_0^\infty \widehat{\nu}(u)udu = \int_0^\infty e^{-u}\nu(u)udu = \int_0^\infty e^{-u}\nu(u)du + \int_0^\infty ue^{-u}d\nu(u) = 1 + \mu.$$

□

3 Proofs of the Main Results

Below we will use $\mathcal{O}(\cdot)$ notation to simplify the discussion as was described in Section 1.3. It is understood that the hidden constants in $\mathcal{O}(\cdot)$ do not depend on n, ϵ, $\gamma := \epsilon^2$ or $B := \epsilon^{-30}$.

3.1 Proof of Theorem 1.1

In this section we first present some crucial lemmas by which we can then prove Theorem 1.1. The proof of Theorem 1.1 consists of two parts one concerning (1.7) and one concerning (1.8). The proofs of these lemmas are given in Section 3.1.4.

3.1.1 Lemmas of Theorem 1.1

The first lemma is fundamental for the proof.

Lemma 3.1. *For the first moment of the number of nodes N we have*

$$\mathbf{E}(N) = \mathcal{O}(n), \tag{3.1}$$

and for the second moment of N we have

$$\mathbf{E}(N^2) = \mathcal{O}(n^2). \tag{3.2}$$

Lemma 3.2. *Adding K balls to a split tree with n balls will only affect the expected number of nodes by $\mathcal{O}(K)$, i.e., for all natural integers K and n,*

$$0 \leq \mathbf{E}(|T^{n+K}|) - \mathbf{E}(|T^n|) = \mathcal{O}(K).$$

Let R be the set of nodes such that given the split vectors, $r \in R$, if

$$M_r^n = n \prod_{j=1}^{d(r)} V_j < B$$

but for all strict ancestors v of r we have $M_v^n = n \prod_{j=1}^{d(v)} V_j \geq B$. We choose $B = \epsilon^{-30}$, we can assume that ϵ is close to 0 making B fairly large. To show (1.7) we consider all subtrees $T_r,\ r \in R$. Let n_r be the number of balls and let N_r be the number of nodes in the $T_r,\ r \in R$, subtree. Recall from Lemma 1.1 that with "large" probability the cardinality n_r is "close" to M_r^n. Corollary 2.1 implies that most nodes are in the $T_r,\ r \in R$, subtrees, i.e.,

$$\mathbf{E}(N) = \mathbf{E}\big(\sum_{r\in R} N_r\big) + \mathcal{O}\big(\frac{n}{B}\big). \tag{3.3}$$

Since the variance of a Bin(m,p) distribution is $m(p-p^2)$, the Chebyshev and Markov inequalities give similarly as in Lemma (1.1) that for large B,

$$\mathbf{P}\big(|n_r - M_r^n| \geq B^{0.6}\big) \leq 4\frac{\mathbf{E}\big(M_r^n\big)}{B^{1.2}} + \frac{\sum_{k=1}^{\infty} 4sb^{-k}}{B^{0.6}} \leq \frac{1}{B^{0.1}}. \tag{3.4}$$

From (3.3) we have

$$\mathbf{E}(N) = \mathbf{E}\big(\sum_{r\in R} N_r I\{|n_r - M_r^n| \geq B^{0.6}\}\big) + \mathbf{E}\big(\sum_{r\in R} N_r I\{|n_r - M_r^n| \leq B^{0.6}\}\big) + \mathcal{O}\big(\frac{n}{B}\big). \tag{3.5}$$

The next lemma shows that the expected number of nodes in the $T_r,\ r \in R$, subtrees with subtree sizes n_r that differ significantly from M_r^n is bounded by a "small" error term for large B.

Lemma 3.3. *The expected number of nodes in the $T_r,\ r \in R$, subtrees with subtree size n_r that differs from M_r^n with at least $B^{0.6}$ balls, is*

$$\mathbf{E}(\sum_{r\in R} N_r I\{|n_r - M_r^n| \geq B^{0.6}\}) = \mathcal{O}\big(\frac{n}{B^{0.1}}\big),$$

hence, from (3.5)

$$\mathbf{E}(N) = \mathbf{E}\big(\sum_{r\in R} N_r I\{|n_r - M_r^n| \leq B^{0.6}\}\big) + \mathcal{O}\big(\frac{n}{B^{0.1}}\big). \tag{3.6}$$

We also sub-divide the $T_r,\ r \in R$, subtrees into smaller classes, wherein the M_r^n's in each class are close to each-other. Choose $\gamma := \epsilon^2$ and let

$$Z := \{B, B - \gamma B, B - 2\gamma B, \dots, \epsilon B\},$$

where $\epsilon = \frac{1}{k}$ for some positive integer k. We write $R_z \subseteq R,\ z \in Z$, for the set of nodes $r \in R$, such that $M_r^n \in [z - \gamma B, z)$ and $M_v^n \geq B$ for all strict ancestors v of r. (Note that the intervals are of length γB and that the set Z contains at most $\frac{1}{\gamma}$ elements.) We write $|R_z|$ for the number of nodes in R_z. The next lemma is deduced from renewal theory applied to the renewal function $U(t)$ in (1.17).

Lemma 3.4. *Let $\epsilon = \frac{1}{k}$ for some positive integer k. Define $S := \{1, 1-\gamma, 1-2\gamma, \dots, \epsilon\}$, where $\gamma = \epsilon^2$, and let $B = \epsilon^{-30}$. Choose $\zeta \in S$, then*

$$\frac{\mathbf{E}(|R_{\zeta B}|)}{\frac{n}{B}} = c_\zeta + o(1), \quad \text{as } n \to \infty. \tag{3.7}$$

for a constant c_ζ (only depending on ζ), and also $\sum_{\zeta\in S} c_\zeta \leq \frac{b}{\mu}$.

Before proving these lemmas we show how their use leads to the proof of Theorem 1.1.

3.1.2 Proof of part one of Theorem 1.1

Proof of Theorem 1.7. To show (1.7) it is enough to prove that there exists a constant C such that for all $\epsilon > 0$ there exists n_ϵ such that for two arbitrary values of the cardinality n and $\widehat{n}$, where $\widehat{n} \geq n \geq n_\epsilon$, we have

$$\big|\frac{\mathbf{E}(N)}{n} - \frac{\mathbf{E}(\widehat{N})}{\widehat{n}}\big| \leq C\epsilon. \tag{3.8}$$

Since (3.8) implies that $\frac{\mathbf{E}(N)}{n}$ is Cauchy it follows that $\frac{\mathbf{E}(N)}{n}$ converges to some constant $\alpha > 0$ as n tends to infinity; hence, we deduce (1.7).

We will now prove (3.8). Recall from Section 3.1.1 that we will consider the subtrees $T_r,\ r \in R$. Let $R' \subseteq R$ be the set of nodes such that $r \in R'$ if

$$|n_r - M_r^n| \leq B^{0.6}. \tag{3.9}$$

Lemma 3.3 shows that we only need to consider the nodes in R'.

Let $R'' \subseteq R'$ be the set of nodes such that $r \in R''$ if $r \in R'$ and

$$\epsilon B < M_r^n < B.$$

We will now explain that it is enough to consider the nodes in R''. Corollary 2.1 for $K_n = B$ gives that the expected number of nodes such that $M_v^n \geq B$ is $\mathcal{O}\big(\frac{n}{B}\big)$; thus, since the branch factor is bounded, also the expectation of $|R|$ is $\mathcal{O}\big(\frac{n}{B}\big)$. Hence, for $r \in R'$ by using (3.1) in Lemma 3.1, we get that the expected number of nodes in the T_r, $r \in R'$, subtrees with $M_r^n \leq \epsilon B$ is $\mathcal{O}(\epsilon n)$. From (3.6) in Lemma 3.3, we get

$$\mathbf{E}(N) = \mathbf{E}\big(\sum_{r \in R''} N_r\big) + \mathcal{O}(\epsilon n) + \mathcal{O}\big(\frac{n}{B^{0.1}}\big). \tag{3.10}$$

Recall that $R_z \subseteq R,\ z \in Z$, for the set of nodes $r \in R$ such that $M_r^n \in [z - \gamma B, z)$. Hence, (3.10) gives

$$\mathbf{E}(N) = \mathbf{E}\big(\sum_{z \in Z} \sum_{r \in R' \cap R_z} N_r\big) + \mathcal{O}(\epsilon n) + \mathcal{O}\big(\frac{n}{B^{0.1}}\big). \tag{3.11}$$

We will now apply Lemma 3.2 to calculate the expected value in (3.11). Let r_z be an arbitrarily chosen node in $R' \cap R_z$, where $z \in Z$. By using (3.9) and Lemma 3.2, for any node $r_z \in R' \cap R_z$, we get that the expected number of nodes in a tree with the number of balls in an interval $[z - \gamma B, z)$ is equal to $\mathbf{E}(N_{r_z}) + \mathcal{O}(\gamma B)$. By using (3.11) this implies that

$$\mathbf{E}(N) = \sum_{z \in Z} \mathbf{E}(|R' \cap R_z|)\big(\mathbf{E}(N_{r_z}) + \mathcal{O}(\gamma B)\big) + \mathcal{O}(\epsilon n) + \mathcal{O}\big(\frac{n}{B^{0.1}}\big). \tag{3.12}$$

From Lemma 3.4 we can deduce the asymptotics for $\mathbf{E}(|R' \cap R_z|)$, $z \in Z$. Recall that $R' = \{r \in R : |n_r - M_r^n| \leq B^{0.6}\}$. Clearly, $\mathbf{E}(|R' \cap R_{\zeta B}|) \leq \mathbf{E}(|R_{\zeta B}|)$. Furthermore, by applying (3.4) we get

$$\begin{aligned}
\mathbf{E}(|R' \cap R_{\zeta B}|) &= \sum_{r \in R} \mathbf{P}(|n_r - M_r^n| \leq B^{0.6}, (\zeta - \gamma)B \leq M_r^n < \zeta B) \\
&= \sum_{r \in R} \mathbf{P}((\zeta - \gamma)B \leq M_r^n < \zeta B)\mathbf{P}(|n_r - M_r^n| \leq B^{0.6} \mid (\zeta - \gamma)B \leq M_r^n < \zeta B) \\
&\geq \mathbf{E}(|R_{\zeta B}|)(1 - \mathcal{O}(B^{-0.1})).
\end{aligned} \tag{3.13}$$

By using (3.7) in Lemma 3.4 and applying (3.13), we deduce that for each choice of $\gamma = \epsilon^2$ and $\zeta \in S = \{1, 1 - \gamma, 1 - 2\gamma, \ldots, \epsilon\}$, there is a K_γ such that for a constant c_ζ (depending on ζ),

$$\Big|\frac{\mathbf{E}(|R' \cap R_{\zeta B}|)}{\frac{n}{B}} - c_\zeta\Big| \leq \gamma^2 + \mathcal{O}\big(\frac{1}{B^{0.1}}\big), \tag{3.14}$$

whenever $\frac{n}{B} \geq K_\gamma$ (recall that $B = \epsilon^{-30}$).

Note that since $\sum_{\zeta \in S} c_\zeta \leq \frac{b}{\mu}$, we have that $\sum_{\zeta \in S} c_\zeta \frac{\mathcal{O}(B\gamma)}{B} = \mathcal{O}(\gamma)$. Define $a(x)$ as the quotient of the expected number of nodes in a tree with cardinality $\lfloor x \rfloor$ divided by $\lfloor x \rfloor$. Note from Lemma 3.1 that $a(x)$ is bounded by some uniform constant. Thus, for a constant c_ζ (depending on ζ) and $a(\zeta B)$ bounded by some uniform constant, we get from (3.12) and (3.14) that

$$\begin{aligned}
\mathbf{E}(N) &= n \sum_{\zeta \in S} c_\zeta \frac{1}{B}\Big(\zeta B a(\zeta B) + \mathcal{O}(B\gamma)\Big) + n \sum_{\zeta \in S} \mathcal{O}(\gamma^2) + \mathcal{O}(\epsilon n) \\
&= n \sum_{\zeta \in S} \zeta c_\zeta a(\zeta B) + \mathcal{O}(\epsilon n).
\end{aligned}$$

In analogy also for $\widehat{n} \geq n$,

$$\mathbf{E}(\widehat{N}) = \widehat{n} \sum_{\zeta \in S} \zeta c_\zeta a(\zeta B) + \mathcal{O}(\epsilon \widehat{n}).$$

Thus, (3.8) follows, which shows (1.7).

□

3.1.3 Proof of part two of Theorem 1.1

Proof of (1.8) in Theorem 1.1. First note that (3.2) in Lemma 3.1 implies that

$$\mathbf{Var}(N) = \mathcal{O}(n^2).$$

We intend to use the variance formula

$$\mathbf{Var}(Y) = \mathbf{E}(\mathbf{Var}(Y|\mathscr{G})) + \mathbf{Var}(\mathbf{E}(Y|\mathscr{G})), \tag{3.15}$$

where Y is a random variable and $\mathscr{G}$ is a sub-σ-field, see e.g.[13, exercise 10.17-2]. For some arbitrary small $\epsilon > 0$ there is a constant $c > 0$ such that the number of nodes Z_D between depth $D = \lfloor c \ln n \rfloor$ and the root is bounded by n^ϵ. Choose the constant c corresponding to $\epsilon = \frac{1}{2}$ and consider the subtrees $T_i,\ 1 \leq i \leq b^D$ at depth $D = \lfloor c \ln n \rfloor$. Let n_i be the number of balls and N_i the number of nodes in T_i. By applying (1.7) in Theorem 1.1 gives

$$\mathbf{E}(N|\Omega_D) = \sum_{i=1}^{b^D} \big(\alpha n_i + o(n_i)\big) + \mathbf{E}(Z_D|\Omega_D). \tag{3.16}$$

By applying (3.16) and the fact that $Z_D \leq \sqrt{n}$ we get

$$\mathbf{Var}(\mathbf{E}(N|\Omega_D)) = \mathbf{Var}(\alpha n + o(n)) = o(n^2). \tag{3.17}$$

Recall that Ω_D is the σ-field generated by $\{n_v,\ d(v) \leq D\}$. Conditioned on Ω_D, the sizes $N_i,\ 1 \leq i \leq b^D$, are independent, hence

$$\mathbf{Var}(N|\Omega_D) = \mathbf{Var}(\sum_{i=1}^{b^D} N_i + Z_D|\Omega_D) = \sum_{i=1}^{b^D} \mathbf{Var}(N_i|\Omega_D) = \sum_{i=1}^{b^D} \mathcal{O}(n_i^2). \tag{3.18}$$

Taking expectation in (3.18) gives

$$\mathbf{E}(\mathbf{Var}(N|\Omega_D)) = \sum_{i=1}^{b^D} \mathcal{O}(\mathbf{E}(n_i^2)). \tag{3.19}$$

Lemma 3.5. *Fix a constant $c > 0$ and let $D = \lfloor c \ln n \rfloor$, then there is a $\delta > 0$ such that*

$$\mathbf{E}(\sum_{i=1}^{b^D} n_i^2) = \mathcal{O}(n^{2-\delta}).$$

Proof. From (1.12) we get that conditioning on $\mathscr{G}_D$, for i at depth D,

$$n_i \leq_{st} \mathrm{Bin}(n, \prod_{j=1}^{D} V_j) + s_1 D. \tag{3.20}$$

The fact that the second moment of a $\mathrm{Bin}(m,p)$ is $m^2p^2 + mp - mp^2$ and the bound of n_i in (3.20) give

$$\mathbf{E}({n_i}^2|\mathscr{G}_D) \leq n^2 \prod_{j=1}^{D} V_j^2 + \mathcal{O}(nD \prod_{j=1}^{D} V_j) + \mathcal{O}(D^2).$$

Note that $\mathbf{E}(V^2) < \mathbf{E}(V) = \frac{1}{b}$, since $V \in (0,1)$. Hence, there is an $\epsilon > 0$ such that

$$\begin{aligned}\mathbf{E}({n_i}^2) &\leq n^2 \prod_{j=1}^{D} \mathbf{E}(V^2) + \mathcal{O}(\frac{nD}{b^D}) + \mathcal{O}(D^2) \\ &\leq \frac{n^2}{(b+\epsilon)^D} + \mathcal{O}(\frac{nD}{b^D}) + \mathcal{O}(D^2),\end{aligned}$$

and thus there is a $\delta > 0$ such that $\mathbf{E}(\sum_{i=1}^{b^D} n_i^2) = \mathcal{O}(n^{2-\delta})$. □

Thus, (3.19) and Lemma 3.5 (where we choose the constant c such that $Z_D \leq \sqrt{n}$) give

$$\mathbf{E}(\mathbf{Var}(N|\Omega_D)) = \mathcal{O}(n^{2-\delta}). \tag{3.21}$$

By applying the variance formula in (3.15) we get from (3.21) and (3.17) that

$$\mathbf{Var}(N) = o(n^2).$$

□

Remark 3.1. The proof shows that if we can improve the result in (1.7) such that $\mathbf{E}(N) = \alpha n + \mathcal{O}(n^{1-c_1})$ for some constant $c_1 > 0$, we also get a sharper variance result, i.e., $\mathbf{Var}(N) = o(n^{2-c_2})$ for some constant $c_2 > 0$.

Remark 3.2. There is no uniform bound for the variance of N on the form $o(n^\alpha)$, that holds for all split trees, which is sharper than $o(n^2)$. E.g., for the case of m-ary search trees [16], for each $\epsilon > 0$ there exists m_0 such that for m-ary search trees where $m \geq m_0$ we have $\mathbf{Var}(N) \geq n^{2-\epsilon}$ as n tends to infinity.

3.1.4 Proofs of the Lemmas of Theorem 1.1

Proof of Lemma 3.1. Note that if $s_0 > 0$ it is always true that $N \leq n$ and if $s_1 > 0$ also $N \leq 2n$ holds. For $s_0 = s_1 = 0$ we can argue as follows: When a new ball is added to the tree the expected number of additional nodes is bounded by the expected number of nodes one gets from a splitting node. Let Z be the number of nodes that one gets when a node of $s+1$ balls splits, then

$$N \leq_{st} \sum_{i=1}^{n} Z^i; \quad \text{where } Z^i \overset{d}{=} Z; \tag{3.22}$$

hence $\mathbf{E}(N) \leq n\mathbf{E}(Z)$. We have

$$\mathbf{E}(Z) = \sum_{k=1}^{\infty} k\mathbf{P}(Z = k). \tag{3.23}$$

Note that once a node gives balls to at least 2 children the splitting process ends. Thus, for $k > b$

$$\mathbf{P}(Z^1 = k|\mathscr{G}_k) \leq \sum_{\substack{v;\\ d(v)=k-b-1}} \prod_{j=1}^{k-b-1} V_{j,v}^{s+1}.$$

Hence, (3.23) implies,

$$\mathbf{E}(Z) \leq \sum_{k=b+2}^{\infty} k(b\mathbf{E}(V^{s+1}))^{k-b-1} + (b+1)^2. \tag{3.24}$$

There is a $\delta > 0$ such that $b\mathbf{E}(V^{s+1}) \leq b^{-\delta}$ since $\mathbf{E}(V^{s+1}) < \mathbf{E}(V) = \frac{1}{b}$, for $V \in (0,1)$. Thus, (3.24) gives that there exists a constant C such that

$$\mathbf{E}(Z) \leq C\sum_{k=1}^{\infty} kb^{-k\delta} = C\frac{b^{-\delta}}{(1-b^{-\delta})^2} < \infty. \tag{3.25}$$

This shows (3.1).

Now we show (3.2). By applying the well-known Minkowski's inequality to (3.22) we get

$$\mathbf{E}(N^2) \leq n^2 \mathbf{E}(Z^2). \tag{3.26}$$

By similar calculations as in (3.24)–(3.25) we get that there exist constants C_1 and $\delta > 0$ such that

$$\mathbf{E}(Z^2) \leq \sum_{k=1}^{\infty} k^2 \mathbf{P}(Z = k) \leq \sum_{k=1}^{\infty} k^2 C_1 b^{-k\delta} < \infty. \tag{3.27}$$

Thus, (3.2) follows from (3.26) and (3.27).

□

Proof of Lemma 3.2. The proof of this lemma is in analogy with the proof of (3.1) in Lemma 3.1. Adding one ball to the tree will only increase the number of nodes if it is added to a leaf with s balls. Recall that Z is the number of nodes that one gets when a node of $s+1$ balls splits. From (3.24) we have that $\mathbf{E}(Z) \leq C^{'}$ for some constant $C^{'}$ implying that K balls can create at most $K\mathbf{E}(Z) \leq C^{'}K$ additional nodes. □

Proof of Lemma 3.3. By applying (3.4) we get for B large enough, that with probability at least $1 - \frac{1}{B^{0.1}}$,

$$|n_r - M_r^n| \leq B^{0.6}. \tag{3.28}$$

We have

$$\mathbf{E}\Big(\sum_{r\in R} n_r I\{|n_r - M_r^n| \geq B^{0.6}\}\Big) = E_1 + E_2, \tag{3.29}$$

where

$$E_1 = \mathbf{E}\Big(\sum_{r\in R} n_r I\{|n_r - M_r^n| \geq B^{0.6}\} I\{n_r \leq 2M_r^n\}\Big),$$
$$E_2 = \mathbf{E}\Big(\sum_{r\in R} n_r I\{|n_r - M_r^n| \geq B^{0.6}\} I\{n_r > 2M_r^n\}\Big).$$

Hence, the facts that $\sum_{r\in R} M_r^n = \mathcal{O}(n)$ and that the bound in (3.28) holds with probability $1 - \frac{1}{B^{0.1}}$, give

$$E_1 \leq \mathbf{E}\Big(\sum_{r\in R} 2M_r^n I\{|n_r - M_r^n| \geq B^{0.6}\}\Big) = \mathcal{O}\Big(\frac{n}{B^{0.1}}\Big).$$

Recall that R is the set of nodes such that $r \in R$, if r is the root of a T_r, $r \in R$, subtree. We obviously have

$$E_2 \leq \mathbf{E}\Big(\sum_{v} 2(n_v - M_v^n) I\{n_v > 2M_v^n\} I\{v \in R\}\Big).$$

By summing over nodes v at depth k we get

$$E_2 \leq \sum_{k=0}^{\infty} 2b^k \mathbf{E}\big((n_v - M_v^n) I\{n_v > 2M_v^n\}\big) \mathbf{P}(v \in R). \tag{3.30}$$

We write F for the expected value in (3.30), i.e., $F := \mathbf{E}\big((n_v - M_v^n) I\{n_v > 2M_v^n\}\big)$.

Hence, the conditional Cauchy-Schwarz and the conditional Markov inequalities give

$$\begin{aligned} F :&\leq \mathbf{E}\left(\sqrt{\mathbf{E}\big(\big(n_v - M_v^n\big)^2|\mathscr{G}_d\big)}\sqrt{\mathbf{P}\big(n_v > 2M_v^n|\mathscr{G}_d\big)}\right) \\ &\leq \min\left(\mathbf{E}\left(\frac{\mathbf{E}\big(\big(n_v - M_v^n\big)^2|\mathscr{G}_d\big)}{M_v^n}\right), \mathbf{E}\left(\sqrt{\mathbf{E}\big(\big(n_v - M_v^n\big)^2|\mathscr{G}_d\big)}\right)\right). \end{aligned} \tag{3.31}$$

From (1.12) we have that for all v with $d(v) = d$, conditioned on $\mathscr{G}_d$, $n_v \leq_{st} n'_v + n''_v$, where

$$\begin{aligned} n'_v &:= \mathrm{Bin}(n, \prod_{j=1}^{d} V_{j,v}), \\ n''_v &:= \mathrm{Bin}(s_1, \prod_{j=2}^{d} V_{j,v}) + \cdots + \mathrm{Bin}(s_1, V_{d,v}) + s_1. \end{aligned}$$

Thus, (3.31) gives for $M_v^n \geq 1$,

$$\begin{aligned} F &\leq \mathbf{E}\left(\frac{\mathbf{E}\big(\big(n_v - M_v^n\big)^2|\mathscr{G}_d\big)}{M_v^n}\right) \\ &= \mathbf{E}\left(\frac{\mathbf{E}\big(\big(n'_v - M_v^n\big)^2|\mathscr{G}_d\big)}{M_v^n} + \frac{\mathbf{E}\big((n''_v)^2 + 2n'_v n''_v - 2n''_v M_v^n|\mathscr{G}_d\big)}{M_v^n}\right) \\ &\leq \mathbf{E}\left(\frac{\mathbf{E}\big(\big(n'_v - M_v^n\big)^2|\mathscr{G}_d\big)}{M_v^n}\right) + \mathbf{E}\big((n''_v)^2\big) + 2\mathbf{E}\big(n''_v\big), \end{aligned} \tag{3.32}$$

where we in the last inequality applied that $\mathbf{E}(n'_v|\mathscr{G}_d) = M_v^n$. For $M_v^n < 1$ we apply that (3.31) gives

$$F \leq \mathbf{E}\left(\sqrt{\mathbf{E}\big(\big(n_v - M_v^n\big)^2|\mathscr{G}_d\big)}\right). \tag{3.33}$$

By using that the variance of a Bin(m, p) distribution is $m(p - p^2)$ we get

$$\mathbf{E}\big(\big(n'_v - M_v^n\big)^2\big|\mathscr{G}_d\big) \leq M_v^n,$$

and from Minkowski's inequality we easily deduce that $\mathbf{E}\big((n''_v)^2\big)$ is bounded by some constant C not depending on d. Hence, by using that we can bound F as in (3.32) for $M_v^n \geq 1$, and by the bound in (3.33) for $M_v^n < 1$, we get that $F \leq C_2$ for some constant C_2. Thus, from (3.30) we get $E_2 \leq C_2 \sum_{k=0}^{\infty} b^k \mathbf{P}(v \in R)$. Note that $v \in R$ only if $M_w^n \geq B$ for all strict ancestors w of v. Hence, by applying Corollary 2.1 for $K_n = B$, we get that $E_2 = \mathcal{O}\big(\frac{n}{B}\big)$. By applying Lemma 3.1 in combination with (3.29) and using the bounds of E_1 and E_2 we get

$$\mathbf{E}\big(\sum_{r\in R} N_r I\{|n_r - M_r^n| \geq B^{0.6}\}\big) = \mathcal{O}\big(\mathbf{E}\big(\sum_{r\in R} n_r I\{|n_r - M_r^n| \geq B^{0.6}\}\big)\big) = \mathcal{O}\big(\frac{n}{B^{0.1}}\big).$$

□

Proof of Lemma 3.4. Recall the definition of $Y_k = -\sum_{j=1}^{k} \ln V_j$ and

$$\nu(t) = b\mathbf{P}(-\ln V \leq t).$$

Also recall that we write $S = \{1, 1-\gamma, 1-2\gamma, \ldots, \epsilon\}$ for $\gamma = \epsilon^2$ and $\epsilon = \frac{1}{k}$ for some positive integer k. We have for $\zeta \in S$

$$\mathbf{E}(|R_{\zeta B}|) = \sum_{k=0}^{\infty} b^{k+1}\Big(\mathbf{P}\big(\{Y_k - \ln V_{k+1} > \ln\frac{n}{B} - \ln\zeta\} \cap \{Y_k \le \ln\frac{n}{B}\}\big)$$
$$- \mathbf{P}\big(\{Y_k - \ln V_{k+1} > \ln\frac{n}{B} - \ln\big(\zeta-\gamma\big)\} \cap \{Y_k \le \ln\frac{n}{B}\}\big)\Big).$$

We write $q := \ln\frac{n}{B}$. We have that $\mathbf{E}(|R_{\zeta B}|)$ is equal to

$$Z(q) := \int_0^q b\Big(\mathbf{P}\Big(-\ln V > q - t - \ln\zeta\Big) - \mathbf{P}\Big(-\ln V > q - t - \ln\big(\zeta-\gamma\big)\Big)\Big)dU(t);$$

recalling the definition of $U(t)$ in (1.17). Hence,

$$Z(q) := \int_0^q b\mathbf{P}\big(q - t - \ln\zeta < -\ln V \le q - t - \ln\big(\zeta-\gamma\big)\big)dU(t). \tag{3.34}$$

We write

$$G(t) := b\mathbf{P}\big(t - \ln\zeta < -\ln V \le t - \ln\big(\zeta-\gamma\big)\big).$$

Thus, $Z(q) = (G * dU)(q)$. Recall that we write $d\omega(t) = e^{-t}d\nu(t)$ where $\omega(t)$ is a probability measure. Recall from (2.2) that we have

$$\widehat{U}(t) = \widehat{\nu}(t) + (\widehat{U} * d\omega)(t),$$

where $\widehat{U}(t) := e^{-t}U(t)$ and $\widehat{\nu}(t) := e^{-t}\nu(t)$. Thus, by using [1, Theorem VI.5.1] we have for $\widehat{Z}(x) = e^{-x}Z(x)$ and $\widehat{G}(x) = e^{-x}G(x)$ that $\widehat{Z}(q) = (\widehat{G} * d\omega)(q)$. By using (3.34) this implies that

$$\widehat{Z}(q) = \int_0^q be^{t-q}\mathbf{P}\big(q - t - \ln\zeta < -\ln V \le q - t - \ln\big(\zeta-\gamma\big)\big)d\omega(t).$$

Applying the key renewal theorem [12, Theorem II.4.3] to $\widehat{U}(t)$ we get

$$\lim_{q\to\infty}\widehat{Z}(q) = \frac{b}{\mu}\int_0^{\infty} e^{-t}\mathbf{P}\big(t - \ln\zeta < -\ln V \le t - \ln\big(\zeta-\gamma\big)\big)dt.$$

Note that $\lim_{q\to\infty}\widehat{Z}(q) := c_\zeta$, for some constant c_ζ only depending on ζ. Thus, by using $\widehat{Z}(x) = e^{-x}Z(x)$ we get that $\mathbf{E}(|R_{\zeta B}|) = \frac{n}{B}c_\zeta + o\big(\frac{n}{B}\big)$, which shows (3.7).

Also note that we have

$$\sum_{\zeta\in S} c_\zeta = \frac{b}{\mu}\int_0^{\infty} e^{-t}\sum_{\zeta\in S}\mathbf{P}\big(t - \ln\zeta < -\ln V \le t - \ln\big(\zeta-\gamma\big)\big)dt$$
$$= \frac{b}{\mu}\int_0^{\infty} e^{-t}\mathbf{P}\big(t < -\ln V \le t - \ln\big(\epsilon-\gamma\big)\big)dt \le \frac{b}{\mu}.$$

□

3.2 Proof of Theorem 1.2

Proof of Theorem 1.2. We use large deviations to show this theorem.

Note that a node v belongs to T^n, if and only if, $n_v \ge 1$. Recall from (1.12) that given $\mathscr{G}_d$,

$$n_v \le_{st} \mathrm{Bin}(n, \prod_{j=1}^{d} V_j) + \mathrm{Bin}(s_1, \prod_{j=2}^{d} V_j) + \cdots + \mathrm{Bin}(s_1, V_d) + s_1, \tag{3.35}$$

where $V_j,\ j \in \{1,\dots,d\}$, are i.i.d. random variables distributed as V. It is enough to consider the first term $\mathrm{Bin}(n,\prod_{j=1}^{d} V_j)$ in (3.35), and prove that the expected number of bad nodes with $\mathrm{Bin}(n,\prod_{j=1}^{d} V_j) \geq 1$ is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$. If $s_1 = 0$, then $\mathrm{Bin}(n,\prod_{j=1}^{d} V_j)$ is the only term in (3.35). We now explain the fact that we can ignore the terms in n_v that occur because of the parameter s_1. Assume that for split trees with $s_1 = 0$, the expected number of bad nodes is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$. We first consider the nodes with $d \leq \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon}$. If $s_1 > 0$, we assume that we first add the n balls as in the construction of a split tree with $s_1 = 0$. Hence, the expected number of nodes v with $d \leq \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon}$, is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$. We now repay the subtree sizes for their potential loss of balls because of $s_1 > 0$. A node v at depth d can at most have a loss of $s_1 d$ balls in the subtree rooted at v. These balls cannot give more than $s_1 b d$ nodes to the tree (since only if $s_0 = s_1 = 0$ it is possible for an increment of more than b nodes when a new ball is added to the tree). Thus, since $d \leq \frac{\ln n}{\mu}$ and the fact that we assume that we have $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$ nodes in expectation before the repayment of the loss of balls, these additional balls cannot give more than $\mathcal{O}\big(\frac{n}{\ln^{k} n}\big)$ nodes in expectation. Now we consider the nodes with $d \geq \frac{\ln n}{\mu} + (\ln n)^{\frac{1}{2}+\epsilon}$. Again we first distribute the n balls assuming that $s_1 = 0$, and then repay for the potential loss of balls in the subtrees if $s_1 > 0$. Note that for $d = \mathcal{O}(\ln n)$ we can argue as in the previous case. This means that the expected number of nodes with $\frac{\ln n}{\mu} + (\ln n)^{\frac{1}{2}+\epsilon} \leq d \leq K \ln n$ for some arbitrary constant K is $\mathcal{O}\big(\frac{n}{\ln^{k} n}\big)$. For larger d we argue as follows: For any constant $K > 0$,

$$\begin{aligned}
&\mathrm{mBin}(s_1,\prod_{j=2}^{d} V_j) + \mathrm{mBin}(s_1,\prod_{j=3}^{d} V_j) + \cdots + \mathrm{mBin}(s_1, V_{d,v}) + s_1 \\
&\leq \mathrm{mBin}(s_1,\prod_{j=2}^{d} V_j) + \cdots + \mathrm{mBin}(s_1,\prod_{j=d-\lfloor K\ln n\rfloor}^{d} V_j) + K s_1 \ln n.
\end{aligned}$$

The Markov inequality gives that for any constant $K > 0$ there exists a constant C such that

$$\begin{aligned}
&\mathbf{P}\big(\mathrm{mBin}(s_1,\prod_{j=2}^{d} V_j) + \cdots + \mathrm{mBin}(s_1,\prod_{j=d-\lfloor K\ln n\rfloor}^{d} V_j) \geq 1\big) \\
&\leq \mathbf{E}\big(\mathrm{Bin}(s_1,\prod_{j=2}^{d} V_j) + \cdots + \mathrm{Bin}(s_1,\prod_{j=d-\lfloor K\ln n\rfloor}^{d} V_j)\big) \leq C b^{-K\ln n},
\end{aligned} \tag{3.36}$$

where the last equality is obtained by first condition on $\mathscr{G}_d$ and then take the expected value twice. Thus, the expected number of nodes that gets a repayment of at least $K s_1 \ln n + 1$ balls is $\mathcal{O}\big(\frac{n}{b^{K\ln n}}\big)$. Since $s_1 > 0$, we can assume that $d \leq n$. Hence, the expected number of balls of this contribution is $\mathcal{O}\big(\frac{n^2}{b^{K\ln n}}\big)$; choosing K large enough this number is $o(1)$.

It remains to prove that if $s_1 = 0$ the expected number of nodes v, where $d(v) \leq \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon}$ or $d(v) \geq \frac{\ln n}{\mu} + (\ln n)^{\frac{1}{2}+\epsilon}$, with $n_v \geq 1$ is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$ for any constant k. Note that an upper bound of the expected number of nodes at depth d is given by

$$b^d \mathbf{P}(n_v \geq 2), \tag{3.37}$$

where v is a node at depth $d-1$. Note that this is true also when $s_0 = 0$, since for all internal nodes $n_v \geq s+1$. Choosing $t > 0$, an application of the Markov inequality implies that

$$\mathbf{P}(n_v \geq 2) \leq \mathbf{P}(n_v(n_v-1) \geq 2) \leq \mathbf{P}(n_v^t (n_v-1)^t \geq 2^t) \leq \frac{\mathbf{E}(n_v^t (n_v-1)^t)}{2^t}. \tag{3.38}$$

Thus, an upper bound of the expected profile for the nodes at depth d is

$$b^d\mathbf{E}(n_v^t(n_v-1)^t), \tag{3.39}$$

where v is a is a node at depth $d-1$.

First we show (assuming $s_1=0$) that the expected number of nodes v where

$$d(v) \geq \frac{\ln n}{\mu} + (\ln n)^{\frac{1}{2}+\epsilon}$$

is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$. We prove this by choosing $t=\frac{1+\epsilon(n)}{2}$, where $\epsilon(n)>0$ is a decreasing function of n that we specify below, and show that

$$\sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} b^d\mathbf{E}\big(n_v^{\frac{1+\epsilon(n)}{2}}(n_v-1)^{\frac{1+\epsilon(n)}{2}}\big) = \mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big). \tag{3.40}$$

Let X_d be distributed as $\mathrm{mBin}(n,\prod_{j=1}^d V_j)$. To show (3.40) it is enough to show that the expected value of

$$\sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} b^d\mathbf{E}\big(X_d^{\frac{1+\epsilon(n)}{2}}(X_d-1)^{\frac{1+\epsilon(n)}{2}}\big|\mathscr{G}_d\big), \tag{3.41}$$

is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$. That this is enough follows from (3.35), since we assume that $s_1=0$.

Suppose that $\epsilon(n)<1$, thus the Lyapounov inequality (which is a special case of the well-known Hölder inequality) gives

$$\mathbf{E}\big(X_d^{\frac{1+\epsilon(n)}{2}}(X_d-1)^{\frac{1+\epsilon(n)}{2}}\big|\mathscr{G}_d\big) \leq (n^2-n)^{\frac{1+\epsilon(n)}{2}}\prod_{j=1}^d V_j^{1+\epsilon(n)} \leq \big(n\prod_{j=1}^d V_j\big)^{1+\epsilon(n)}. \tag{3.42}$$

Hence, to show (3.40) we deduce from the right hand-side of the second inequality in (3.42) that it is enough to show that

$$S_1 := \sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} b^d\big(\mathbf{E}(V^{1+\epsilon(n)})\big)^d n^{1+\epsilon(n)} = \mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big). \tag{3.43}$$

Taylor expansion gives

$$V^{1+\epsilon(n)} = Ve^{\epsilon(n)\ln V} = V\Big(1+\epsilon(n)\ln V + \frac{\ln^2 V}{2}\epsilon^2(n)\Big) + \mathcal{O}\big(V\ln^3 V\epsilon^3(n)\big). \tag{3.44}$$

Thus, by taking expectations in (3.44) we get

$$\begin{aligned} S_1 &= \sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} \Big(1-\mu\epsilon(n)+\frac{\sigma^2+\mu^2}{2}\epsilon^2(n)+\mathcal{O}(\epsilon^3(n))\Big)^d n^{1+\epsilon(n)} \\ &= \sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} e^{d\ln\big(1-\mu\epsilon(n)+\frac{\sigma^2+\mu^2}{2}\epsilon^2(n)+\mathcal{O}(\epsilon^3(n))\big)+(1+\epsilon(n))\ln n} \\ &= \sum_{d=\lfloor \frac{\ln n}{\mu}+(\ln n)^{\frac{1}{2}+\epsilon}\rfloor -1}^{\infty} e^{d\big(-\mu\epsilon(n)+\frac{\sigma^2}{2}\epsilon^2(n)+\mathcal{O}(\epsilon^3(n))\big)+(1+\epsilon(n))\ln n} \\ &= \mathcal{O}\Big(\frac{n^{1-\mu\epsilon(n)(\ln n)^{-\frac{1}{2}+\epsilon}+\mathcal{O}(\epsilon^2(n))}}{\epsilon(n)}\Big). \end{aligned} \tag{3.45}$$

Let $\epsilon(n) := \delta(\ln n)^{-\frac{1}{2}+\epsilon}$ for some small enough constant $\delta > 0$, then the last inequality in (3.45) implies that for some constant $C > 0$ and any constant k,

$$S_1 = \mathcal{O}\Big(ne^{-C\ln^{2\epsilon} n}\Big) = \mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big). \tag{3.46}$$

We argue similarly for the nodes v, $d(v) \le \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon}$. In (3.39) let $t = \frac{1-\epsilon(n)}{2}$ where $\epsilon(n) = \delta(\ln n)^{-\frac{1}{2}+\epsilon}$ as above. In analogy with (3.40) the expected number of nodes v such that $d(v) \le \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon}$ is bounded by

$$S_2 := \sum_{d=0}^{\lfloor \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon} \rfloor} b^d \mathbf{E}\big(n_v^{\frac{1-\epsilon(n)}{2}} (n_v - 1)^{\frac{1-\epsilon(n)}{2}}\big).$$

We use similar calculations as in (3.43)–(3.46) to show that

$$\sum_{d=0}^{\lfloor \frac{\ln n}{\mu} - (\ln n)^{\frac{1}{2}+\epsilon} \rfloor} b^d \big(\mathbf{E}(V^{1-\epsilon(n)})\big)^d n^{1-\epsilon(n)} = \mathcal{O}\Big(ne^{-B\ln^{2\epsilon} n}\Big).$$

This implies in analogy with (3.40)–(3.43) that for some constant C and any constant k,

$$S_2 = \mathcal{O}\big(ne^{-C\ln^{2\epsilon} n}\big) = \mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big). \tag{3.47}$$

Hence, if $s_1 = 0$ the expected number of bad nodes is $\mathcal{O}\big(\frac{n}{\ln^{k+1} n}\big)$, for any constant k, and from our previous explanation it follows that the expected number of bad nodes for $s_1 \ge 0$ is $\mathcal{O}\big(\frac{n}{\ln^k n}\big)$. $\square$

Remark 3.3. We note from (3.45), (3.46) and (3.47) that a sharper bound for the expected number of bad nodes is $\mathcal{O}\big(ne^{-C'\ln^{2\epsilon} n}\big)$ for some constant $C' > 0$.

Remark 3.4. From the calculations in (3.45), we see that a smaller error term holds for larger depths, i.e., for any constant $r > 0$ there is a constant $K > 0$, such that the expected number of nodes with $d(v) \ge K \ln n$ is $\mathcal{O}\big(\frac{1}{n^r}\big)$.

3.3 Proof of Theorem 1.3 and Proposition 1.1

Proof of Theorem 1.3. We write $Z_n := \frac{D_n - \mu^{-1}\ln n}{\sqrt{\ln n}}$. By a classical result in probability theory, see e.g. [13, Theorem 5.5.4], the limit law in (1.6) implies that (1.9) holds if Z_n is uniformly integrable. In particular this is true if Z_n^2 is uniformly integrable. This uniformly integrability also gives

$$\mathbf{E}\big(Z_n^2\big) := \frac{\mathbf{E}\big(\big(D_n - \mu^{-1}\ln n\big)^2\big)}{\ln n} \to \mathbf{E}\big(N\big(0, \sigma^2\mu^{-3}\big)^2\big) = \sigma^2\mu^{-3}. \tag{3.48}$$

Furthermore, the convergence results in (1.9) and (3.48) imply (1.11) for D_n. By using the same coupling argument as in (1.4) it is easy to show that the convergence result of the expected depth in (1.9) implies the convergence result of the expected average depth in (1.10).

Thus, it remains to show that Z_n^2 is uniformly integrable and that (1.11) for D_n implies that (1.11) also holds for $D_{k,n}$, where $\frac{n}{\ln n} \le k < n$. By a standard argument, see e.g. [13, Theorem 5.5.4], Z_n^2 is uniformly integrable if for some $p > 1$ and n_0 large enough,

$$\sup_{n > n_0} \mathbf{E}\big(|Z_n^2|^p\big) := \sup_{n > n_0} \mathbf{E}\Big(\Big|\frac{\big(D_n - \mu^{-1}\ln n\big)^2}{\ln n}\Big|^p\Big), \tag{3.49}$$

is uniformly bounded. We choose $p = \frac{3}{2}$. We show that this is true by using similar calculations as Devroye used in [5] for proving the limit law of D_n in (1.6). First, consider an infinite random path $u_1, u_2, \dots,$ in the skeleton tree S_b, where u_1 is the root. Given u_1 and the split vector $\mathcal{V}_{u_i} = (V_1, \dots, V_b)$ for u_i, then u_{i+1} is the jth child of i with probability V_j. Construct a random split tree with n balls and let $u*$ be the unique leaf in the infinite path. Then by a natural coupling, letting the nth ball follow the random path, D_n is in stochastic sense less than or equal to the distance between $u*$ and the root. In the coupling D_n is less than this distance, if the nth ball is sent to a leaf which splits and does not send this ball to one of its children (i.e, the nth ball is one of the s_0 balls). If the nth ball is one of the s_1 balls it is added to a sister of $u*$, i.e., it ends up at the same depth as $u*$. For all $\beta > 0$ we have

$$\mathbf{P}\big(D_n > k + \beta\big) \leq \mathbf{P}\big(n(u_k) > \beta\big) + \mathbf{P}\big(H_\beta > \beta\big); \tag{3.50}$$

where H_j denotes the height of a split tree with j balls. Furthermore,

$$\mathbf{P}\big(D_n < k\big) \leq \mathbf{P}\big(n(u_k) \leq s + 1\big). \tag{3.51}$$

Recall that $\Delta = V_S$, where given $(V_1, \dots, V_b)$, $S = j$ with probability V_j. Then

$$n(u_k) \leq_{st} \mathrm{mBin}(n, \prod_{j=1}^{k} \Delta_j) + \mathrm{mBin}(s_1, \prod_{j=2}^{k} \Delta_j) + \cdots + \mathrm{mBin}(s_1, \Delta_k) + s_1, \tag{3.52}$$

where Δ_j are i.i.d random variables distributed as Δ.

Consider the probability $\mathbf{P}(D_n > k + \beta)$, where $k = \lfloor \mu^{-1} \ln n + \frac{x}{2}\sqrt{\ln n} \rfloor$ for $x \in R^+$. We bound this by bounding the probabilities in the right hand-side of (3.50), choosing $\beta = \lfloor \frac{x}{2} \ln^{0.2} n \rfloor$. First note that (3.52) implies that

$$n(u_k) \leq_{st} \mathrm{mBin}\big(n, \prod_{j=1}^{k} \Delta_j\big) + \cdots + \mathrm{mBin}\big(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n \rfloor + 1}^{k} \Delta_j\big) + \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor.$$

Thus, we can bound the first probability in the right hand-side of (3.50) by

$$\begin{aligned} \mathbf{P}(n(u_k) > \beta) \leq{}& \mathbf{P}\Big(\mathrm{mBin}\big(n, \prod_{j=1}^{k} \Delta_j\big) + \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor \geq \beta - 1\Big) \\ &+ \mathbf{P}\Big(\mathrm{mBin}\big(s_1, \prod_{j=2}^{k} \Delta_j\big) + \cdots + \mathrm{mBin}\big(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n \rfloor}^{k} \Delta_j\big) > 1\Big). \end{aligned} \tag{3.53}$$

We bound the first probability in the right hand-side of the inequality in (3.53), by using [5, Lemma 4] which states a general result for bounding tail probabilities for mixed binomial (m, Z) distributions where Z is a random variable. Thus, we obtain

$$\begin{aligned} P_1 :={}& \mathbf{P}\Big(\mathrm{mBin}\big(n, \prod_{j=1}^{k} \Delta_j\big) > \beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1\Big) \\ \leq{}& \mathbf{P}\Big(\sum_{j=1}^{k} \ln \Delta_j > \ln\big(\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2n}\big)\Big) + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}}. \end{aligned} \tag{3.54}$$

Recall that we write $\mu = \mathbf{E}(-\ln \Delta)$ and $\sigma^2 = \mathbf{Var}(\ln \Delta)$. From (3.54) we deduce that for n large enough

$$\begin{aligned} P_1 \leq{}& \mathbf{P}\Big(\frac{\sum_{j=1}^{k} \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} > \frac{\ln\big(\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2n}\big) + k\mu}{\sqrt{k\sigma^2}}\Big) + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}} \\ \leq{}& \mathbf{P}\Big(\frac{\sum_{j=1}^{k} \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} > \frac{x\mu^{\frac{3}{2}}}{3\sigma}\Big) + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}}. \end{aligned} \tag{3.55}$$

Since the Δ_j, $j \in \{1, \ldots, k\}$, are i.i.d random variables we can use the Marcinkiewicz-Zygmund inequality, see e.g. [13, Corollary 3.8.2], which gives for $q \geq 2$,

$$\mathbf{E}\Big(\Big|\sum_{j=1}^{k} \ln \Delta_j + k\mu\Big|^q\Big) \leq B_q k^{\frac{q}{2}} \mathbf{E}\Big(\big|\ln \Delta_j + \mu\big|^q\Big), \tag{3.56}$$

where B_q is a constant only depending on q. By using the Markov inequality and (3.56) we get from (3.55) that for n large enough

$$\begin{aligned} P_1 &\leq \frac{\mathbf{E}\Big(\Big(\frac{\sum_{j=1}^{k} \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}}\Big)^4\Big)}{\Big(\frac{x\mu^{\frac{3}{2}}}{3\sigma}\Big)^4} + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n\rfloor - 1}{2}} \\ &\leq \frac{B_4 \mathbf{E}\Big(\big|\ln \Delta_j + \mu\big|^4\Big)}{\Big(\frac{x\mu^{\frac{3}{2}}}{3}\Big)^4} + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n\rfloor - 1}{2}} = \frac{C}{x^4} + \big(\frac{e}{4}\big)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n\rfloor - 1}{2}}, \end{aligned} \tag{3.57}$$

for the constant $C = \frac{B_4 \mathbf{E}(|\ln \Delta_j + \mu|^4) 3^4}{\mu^6} < \infty$ (recall from Section 1.1.2 that all moments of $|\ln \Delta|$ are bounded). Note that $c := \mathbf{E}(\Delta) = b\mathbf{E}(V^2) < 1$. The Markov inequality implies that

$$\begin{aligned} &\mathbf{P}\Big(\mathrm{mBin}\big(s_1, \prod_{j=2}^{k} \Delta_j\big) + \cdots + \mathrm{mBin}\big(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n\rfloor + 1}^{k} \Delta_j\big) \geq 1\Big) \\ &\leq \mathbf{E}\Big(\mathrm{mBin}\big(s_1, \prod_{j=2}^{k} \Delta_j\big) + \cdots + \mathrm{mBin}\big(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n\rfloor + 1}^{k} \Delta_j\big)\Big) \\ &= \mathcal{O}\big(c^{\lfloor \frac{x}{2} \ln^{0.1} n\rfloor}\big), \quad \text{for } \; c = \mathbf{E}(\Delta) < 1. \end{aligned} \tag{3.58}$$

We now consider the other probability i.e., $\mathbf{P}\big(H_\beta > \beta\big)$. (Note that this probability is 0 if $s_0 > 0$ or $s_1 > 0$.) By applying (3.37) we get $\mathbf{P}\big(H_\beta > \beta\big) \leq b^\beta \mathbf{P}\big(n(v) \geq 2\big)$, where v is a node at depth $\beta - 1$. From (3.38) we deduce for $t = 0.75$,

$$\mathbf{P}(n_v \geq 2) \leq \mathbf{E}(n_v^{0.75}(n_v - 1)^{0.75}). \tag{3.59}$$

Let X_β be distributed as $\mathrm{mBin}(n, \prod_{j=1}^{\beta} V_j)$. Note similarly as in (3.41) that (3.59) is bounded by the expectation of $\mathbf{E}\big(X_\beta^{0.75}(X_\beta - 1)^{0.75}\big|\mathscr{G}_\beta\big)$. In analogy to (3.42) the Lyapounov inequality gives

$$\mathbf{E}\big(X_\beta^{0.75}(X_\beta - 1)^{0.75}\big|\mathscr{G}_\beta\big) \leq \big(\beta \prod_{j=1}^{\beta} V_j\big)^{1.5}.$$

Again the fact that $\mathbf{E}(V^2) < \mathbf{E}(V) = \frac{1}{b}$ (since $V \in (0, 1)$), gives that there is a $\delta > 0$ such that

$$\mathbf{P}\Big(H_\beta > \beta\Big) \leq b^{-\delta\beta} \beta^{1.5}. \tag{3.60}$$

We now consider the probability $\mathbf{P}(D_n < k)$, where $k = \lfloor \mu^{-1} \ln n - x\sqrt{\ln n}\rfloor$ for $x \in R^+$, and use the bound of the larger probability in (3.51). We have

$$\mathbf{P}(n(u_k) \leq s + 1) \leq \mathbf{P}\Big(-ks + \mathrm{Bin}(n, \prod_{j=1}^{k} \Delta_j) \leq s + 1\Big).$$

Again by applying [5, Lemma 4] and using similar calculations as in (3.54)–(3.57), we get that for n large enough

$$\begin{aligned}
P_2 &\le \mathbf{P}\Big(\frac{\sum_{j=1}^{k}\ln\Delta_j + k\mu}{\sqrt{k\sigma^2}} < \frac{\ln\big(\frac{2(s(k+1)+1)}{n}\big)+k\mu}{\sqrt{k\sigma^2}}\Big) + \big(\frac{2}{e}\big)^{s(k+1)+1} \\
&\le \mathbf{P}\Big(\frac{\sum_{j=1}^{k}\ln\Delta_j + k\mu}{\sqrt{k\sigma^2}} < \frac{x\mu^{\frac{3}{2}}}{3\sigma}\Big) + \big(\frac{2}{e}\big)^{s(k+1)+1} \\
&\le \frac{B_4\mathbf{E}\Big(\big|\ln\Delta_j+\mu\big|^4\Big)}{\Big(\frac{x\mu^{\frac{3}{2}}}{3}\Big)^4} + \big(\frac{2}{e}\big)^{s(k+1)+1} \\
&= C\frac{1}{x^4} + \big(\frac{2}{e}\big)^{s(k+1)+1}, \qquad \text{for } \ C = \frac{B_4\mathbf{E}(|\ln\Delta_j+\mu|^4)3^4}{\mu^6} < \infty.
\end{aligned} \tag{3.61}$$

It now follows that $\sup_{n>n_0}\mathbf{E}\big(|Z_n^2|^{\frac{3}{2}}\big)$ in (3.49) is uniformly bounded: By the choice of k and β, we get from (3.57), (3.58), (3.60) and (3.61) that for for n_0 large enough

$$\begin{aligned}
\sup_{n>n_0}\mathbf{E}\big(|Z_n^2|^{\frac{3}{2}}\big) &:= \sup_{n>n_0}\mathbf{E}\Big(\Big|\frac{\big(D_n-\mu^{-1}\ln n\big)^2}{\ln n}\Big|^{\frac{3}{2}}\Big) \\
&= \sup_{n>n_0}\int_{x=0}^{\infty} 3x^2\mathbf{P}\Big(\Big|\frac{D_n-\mu^{-1}\ln n}{\sqrt{\ln n}}\Big| > x\Big)dx \\
&\le \sup_{n>n_0}\Big\{\int_{x=1}^{\infty}\Big(\frac{6C}{x^2} + 3x^2\big(\frac{e}{4}\big)^{\frac{\beta-\lfloor\frac{s_1 x}{2}\ln^{0.1}n\rfloor-1}{2}} + 3x^2\big(\frac{2}{e}\big)^{s(k+1)+1} \\
&\quad + \mathcal{O}\big(x^2c^{\lfloor\frac{x}{2}\ln^{0.1}n\rfloor}\big) + 3x^2b^{-\delta\beta}\beta^{1.5}\Big)dx\Big\} + 1 < \infty,
\end{aligned} \tag{3.62}$$

and thus Z_n^2 is uniformly integrable so that (3.48) holds, which shows (1.11) for D_n.

It is now easy to show that (1.11) also holds for $D_{k,n}$, $\frac{n}{\ln n} \le k < n$. Proposition 1.1 implies that

$$D_k \le_{st} D_{k,n} \le_{st} D_n, \qquad \text{for } \ k \le n. \tag{3.63}$$

From (1.6) it follows that for all $\frac{n}{\ln n} \le k \le n$,

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

By using this and (3.63), for $\frac{n}{\ln n} \le k \le n$, we have

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

We need to show that for $\frac{n}{\ln n} \le k \le n$, we have

$$\frac{\mathbf{E}\big(\big(D_{k,n}-\mu^{-1}\ln n\big)^2\big)}{\ln n} \to \mathbf{E}\big(N\big(0,\sigma^2\mu^{-3}\big)^2\big).$$

As for D_n this follows if for n_0 large enough,

$$\sup_{n>n_0}\mathbf{E}\Big(\Big|\frac{\big(D_{k,n}-\mu^{-1}\ln n\big)^2}{\ln n}\Big|^{\frac{3}{2}}\Big) < \infty. \tag{3.64}$$

We have for $k \leq n$,

$$\mathbf{P}\Big(\frac{D_k - \mu^{-1}\ln n}{\sqrt{\ln n}} \geq x\Big) \leq \mathbf{P}\Big(\frac{D_{k,n} - \mu^{-1}\ln n}{\sqrt{\ln n}} \geq x\Big) \leq \mathbf{P}\Big(\frac{D_n - \mu^{-1}\ln n}{\sqrt{\ln n}} \geq x\Big) \text{ and}$$
$$\mathbf{P}\Big(\frac{D_n - \mu^{-1}\ln n}{\sqrt{\ln n}} < x\Big) \leq \mathbf{P}\Big(\frac{D_{k,n} - \mu^{-1}\ln n}{\sqrt{\ln n}} < x\Big) \leq \mathbf{P}\Big(\frac{D_k - \mu^{-1}\ln n}{\sqrt{\ln n}} < x\Big).$$

Thus, (3.64) follows from the calculations in (3.62). This shows that (1.11) holds for $D_{k,n}$, $\frac{n}{\ln n} \leq k < n$, follows from the fact that (1.11) holds for D_n. □

Proof of Proposition 1.1. We will prove that for an arbitrary $i \in \{1, \ldots, n-1\}$, $D_{i,n} \leq_{st} D_{i+1,n}$. In the proof coupling arguments will be used.

First consider two identical copies T and $\widehat{T}$ of the split tree when $i-1$ balls have been added, where we let $\widehat{v}$ in $\widehat{T}$ denote the corresponding node of v in T. More precisely, we consider two split trees T and $\widehat{T}$ with the same split vectors in all nodes of the infinite skeleton tree, and if a ball k, $k \leq i-1$, is added to v in T then ball k is added to $\widehat{v}$ in $\widehat{T}$. Now assume that we add the two balls i and $i+1$ to T and $\widehat{T}$.

If ball i and ball $i+1$ are added to different leaves l_1 and l_2 in T then in $\widehat{T}$ we let them switch positions, i.e., ball i is added to $\widehat{l_2}$ and ball $i+1$ is added to $\widehat{l_1}$. Recall that D_j is the last ball added in a tree with j balls. It is obvious for reasons of symmetry that $D_i \stackrel{d}{=} D_{i+1}$. When the balls $\in \{i+2, \ldots, n\}$ are added, we add them to the corresponding nodes in T and $\widehat{T}$. Thus, the two trees are identical in the whole process except for that ball i and ball $i+1$ always have switched positions in T and $\widehat{T}$. By symmetry $D_{i,n} \stackrel{d}{=} D_{i+1,n}$.

If ball i and ball $i+1$ are added to the same leaf l in T then there are three different cases:

If $n_l \leq s-2$, so that l does not split when also ball i and ball $i+1$ have been added, then T and $\widehat{T}$ are still identical since ball i and ball $i+1$ stay in l. When more balls are added we can again assume that ball i and ball $i+1$ have switched positions in T and $\widehat{T}$ at every step of the the iterative construction until all n balls are added. Hence, by symmetry $D_{i,n} \stackrel{d}{=} D_{i+1,n}$.

If $n_l = s-1$, so that l gets $s+1$ balls when the new balls are added, l splits according to the usual splitting process when ball $i+1$ is added. Again we let ball i and ball $i+1$ switch positions in T and $\widehat{T}$. This means that if ball i is added to v_1 and ball $i+1$ is added to v_2 in T, then in $\widehat{T}$ ball i is added to $\widehat{v}_2$ and ball $i+1$ is added to $\widehat{v}_1$. Again by symmetry $D_i \stackrel{d}{=} D_{i+1}$. By the same type of argument as in the cases above $D_{i,n} \stackrel{d}{=} D_{i+1,n}$.

If $n_l = s$, so that l in T gets $s+2$ balls when the new balls are added, let l split according to the usual splitting process where l keeps s_0 balls and sends the other balls to its children.

If ball i is one of the s_0 balls then it is obvious without using the coupling that $D_i \leq D_{i+1}$ and also $D_{i,n} \leq D_{i+1,n}$.

If ball i is not one of the s_1 balls in the children of l in T and ball i is added to v_1 and ball $i+1$ is added to v_2, then in $\widehat{T}$ we can again assume that ball i is added to $\widehat{v}_2$ and ball $i+1$ is added to $\widehat{v}_1$. Thus, $D_i \stackrel{d}{=} D_{i+1}$, and $D_{i,n} \stackrel{d}{=} D_{i+1,n}$.

If ball i is one of the s_1 balls in T, we use a related but not an identical type of coupling argument as in the previous cases. In this case ball i is added by uniformly choosing one of the b children of l each with probability $\frac{1}{b}$, while ball $i+1$ is added by using the probabilities given by the components in the split vector $\mathcal{V}_l$ of l. Again T and $\widehat{T}$ are identical until $i-1$ balls are added barring the possibility of variation in the split vectors of the nodes below the leaves as described below. If ball i in T goes to a child v_1 of l related to a component V_j in $\mathcal{V}_l$, then we add ball $i+1$ in $\widehat{T}$ to $\widehat{v}_1$ with probability $\min\{1, \frac{V_j}{1/b}\}$ and to one of the other children related to a component $V_k > 1/b$ with probability $\max\{0, 1 - \frac{V_j}{1/b}\}$, so that the sum of the probabilities gives the right marginal distribution. Assume that ball i is added to the child v of l in T and ball $i+1$ is added to the child $\widehat{w}$ of $\widehat{l}$ in $\widehat{T}$. This means that $\widehat{w}$ relates to a component of the split vector of $\widehat{l}$ at least as large as the component of the split vector of l related to v. Now we can assume that the split vectors in the nodes in T_v correspond to the split vectors in the nodes in $\widehat{T}_{\widehat{w}}$. This means that we can assume that when ball number j in the subtrees, is added it goes to the corresponding node in both of the subtrees. However, note that the balls could have different labels if we consider their original label in the whole tree, since $\widehat{T}_{\widehat{w}}$ could have more balls than T_v. Thus, as long as the subtrees have the same number of balls, new balls are added to the corresponding positions in these subtrees, and ball i and ball $i+1$ are also held by nodes of corresponding positions. This construction shows that if the subtrees $\widehat{T}_{\widehat{w}}$ and T_v have k and l balls, respectively, where $k > l$, and ball i in T_v is in node $h \in T$, then ball $i+1$ in $\widehat{T}$ is in a subtree of $\widehat{T}_{\widehat{w}}$ with root corresponding to the position of h. This shows that $D_{i,n} \leq_{st} D_{i+1,n}$.

Hence, in all cases, $D_{i,n} \leq_{st} D_{i+1,n}$ and thus for $i < j$, it follows that

$$D_{i,n} \leq_{st} D_{j,n}.$$

□

Acknowledgments. Professor Svante Janson is gratefully acknowledged for invaluable support and advice. I also thank Dr Nicolas Broutin for helpful discussions.

References

[1] S. Asmussen, *Applied Probability and Queues.* John Wiley Sons, Chichester, 1987. MR-0889893

[2] C.J. Bell, *An Investigation into the Principles of the Classification and Analysis of Data on an Automatic Digital Computer. PhD Thesis* 1965.

[3] J. Clément, P. Flajolet, and B. Vallée, Dynamical source in information theory: a general analysis of trie structures. *Algorithmica* **29** (2001), 307–369. MR-1887308

[4] E. G. Coffman, and J. Eve, File structures using hashing functions. *Communications of the ACM* **13** (1970), 427–436.

[5] L. Devroye, Universal limit laws for depths in random trees. *SIAM J. Comput.* **28** (1998), no 2, 409–432. MR-1634354

[6] L. Devroye Applications of Stein's method in the analysis of random binary search trees. *Stein's Method and Applications*, Inst. for Math. Sci. Lect. Notes Ser. **5,** World Scientific Press, Singapore, (2005), 47–297. MR-2205340

[7] W. Feller, Fluctuation theory and recurrent events. *Trans. Amer. Math. Soc.* **67** (1949), 98–119. MR-0032114

[8] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. 1.* 3rd ed., Wiley, New York, 1968. MR-0228020

[9] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II.* 2nd ed., Wiley, New York, 1971. MR-0270403

[10] R.A. Finkel and J.L. Bentley, Quad trees, a data structure for retrieval on composite keys. *Acta Inform* **4** (1974), 1–9.

[11] E. Fredkin, Trie memory. *Communications of the ACM* **3** (1960), no. 9, 490–499.

[12] A. Gut, *Stopped Random Walks.* Springer Verlag, New York, Berlin, Heidelberg, 1988. MR-0916870

[13] A. Gut, *Probability: A Graduate Course*, Springer, New York, 2005. MR-2125120

[14] C.A.R. Hoare, Quicksort. *The Computer Journal* **5**, (1962), 391–424. MR-0142216

[15] C. Holmgren, A weakly 1-stable limiting distribution for the number of random records and cuttings in split trees. Adv. in Appl. Probab. **43** (2011), 151-177. MR-2761152

[16] H. Mahmoud and B. Pittel, Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10** (1989), no. 1, 52–75. MR-0987097

[17] R. Pyke, Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)* **27** (1965), no 3, 395–449. MR-0216622

[18] H. Mohamed and P. Robert, A probabilistic analysis of some tree algorithms. *Ann. Appl. Probab.* **15** (2005), no. 4, 2445–2471. MR-2187300