2005 International Conference on Analysis of Algorithms DMTCS proc. **AD**, 2005, 105–114

A repertoire for additive functionals of uniformly distributed m-ary search trees (Extended Abstract)

James Allen Fill[1] † and Nevin Kapur[2] ‡

[1] *Department of Applied Mathematics and Statistics, The Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218-2682.* `jimfill@jhu.edu`

[2] *Department of Computer Science, California Institute of Technology, MC 256-80, 1200 E. California Blvd., Pasadena CA 91125.* `nkapur@cs.caltech.edu`

Using recent results on singularity analysis for Hadamard products of generating functions, we obtain the limiting distributions for additive functionals on m-ary search trees on n keys with toll sequence (i) n^α with $\alpha \geq 0$ ($\alpha = 0$ and $\alpha = 1$ correspond roughly to the space requirement and total path length, respectively); (ii) $\ln \binom{n}{m-1}$, which corresponds to the so-called shape functional; and (iii) $\mathbf{1}_{n=m-1}$, which corresponds to the number of leaves.

Keywords: additive functionals, Hadamard products, limit laws, method of moments, search trees, shape functional, singularity analysis, space requirement, leaves

1 Introduction

We begin by providing a brief overview of m-ary search trees. For integer $m \geq 2$, the m-ary search tree, or multiway tree, generalizes the binary search tree. The quantity m is called the *branching factor*. According to [17], search trees with branching factors higher than 2 were first suggested by Muntz and Uzgalis [20] "to solve internal memory problems with large quantities of data." For further background we refer the reader to [14, 15] and [17].

We consider the space of *m-ary search trees* on n keys, and assume that the keys can be linearly ordered. Since we shall be concerned only with the structure of the tree and not its specific contents, we can then without loss of generality take the set of keys to be $[n] := \{1, 2, \ldots, n\}$. An m-ary search tree can be constructed from a sequence s of n distinct keys in the following way:

(a) If $n < m$, then all the keys are stored in the root node in increasing order.

(b) If $n \geq m$, then the first $m - 1$ keys in the sequence are stored in the root in increasing order, and the remaining $n - (m - 1)$ keys are stored in the m subtrees subject to the condition that if $\kappa_1 < \kappa_2 < \cdots < \kappa_{m-1}$ denotes the ordered sequence of keys in the root, then the keys in the jth subtree are those that lie between κ_{j-1} and κ_j, where $\kappa_0 := 0$ and $\kappa_m := n + 1$, sequenced as in s.

(c) Recursively, all the subtrees are m-ary search trees that satisfy conditions (a), (b), and (c).

In this work we consider additive functionals on m-ary search trees, as we describe next.

Fix $m \geq 2$. Given an m-ary search tree T, let $L_1(T), \ldots, L_m(T)$ denote the subtrees rooted at the children of the root of T. The *size* $|T|$ of a tree T is the number of keys in it. We will call a functional f on m-ary search trees *additive* if it satisfies the recurrence

$$f(T) = \sum_{i=1}^{m} f(L_i(T)) + b_{|T|}, \tag{1.1}$$

†The research of both authors was supported by NSF Grants DMS–0104167 and DMS–0406104, and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics.

‡Nevin Kapur's research was also partially supported by NSF Grant 0049092, and the by Center for Mathematics of Information at the California Institute of Technology.

for any tree T with $|T| \geq m-1$. Here $(b_n)_{n \geq m-1}$ is a given sequence, henceforth called the *toll sequence* or *toll function*. Note that the recurrence (1.1) does not make any reference to b_n for $0 \leq n \leq m-2$ nor specify the initial conditions $f(T)$ for $0 \leq |T| \leq m-2$.

Several interesting examples can be cast as additive functionals.

Example 1.1. If we specify $f(T)$ arbitrarily for $0 \leq |T| \leq m-2$ and take $b_n \equiv c$ for $n \geq m-1$, we obtain the "additive functional" framework of [17, §3.1]. (Our definition of an additive functional substantially generalizes this notion.) In particular if we define $f(\emptyset) := 0$ and $f(T) := 1$ for the unique m-ary search tree T on n keys for $1 \leq n \leq m-2$ and let $b_n \equiv 1$ for $n \geq m-1$, then $f(T)$ counts the number of nodes in T and thus gives the *space requirement* functional discussed in [17, §3.4].

Example 1.2. If we define $f(T) := 0$ when $|T| = 0$, $f(T) := 1$ when $1 \leq |T| \leq m-2$, and $b_n := \mathbf{1}_{n=m-1}$, then f is the *number of leaves* in the m-ary search tree.

Example 1.3. If we define $f(T) := 0$ when $0 \leq |T| \leq m-2$ and $b_n := n-(m-1)$ for $n \geq m-1$, then f is the *internal path length* functional discussed in [17, §3.5]: $f(T)$ is the sum of all root-to-key distances in T.

In this work we choose to treat explicitly the toll n, rather than $n-(m-1)$. However our techniques reveal that the lead-order asymptotics of moments and the limiting distributions of these two additive functionals are the same.

Example 1.4. As described above, each permutation of $[n]$ gives rise to an m-ary search tree. Suppose we place the uniform distribution on such permutations. This induces a distribution on m-ary search trees called the *random permutation model*. Denote its probability mass function by Q. Dobrow and Fill [2] noted that

$$Q(T) = \frac{1}{\prod_x \binom{|T_x|}{m-1}}, \tag{1.2}$$

where the product in (1.2) is over all nodes in T that contain $m-1$ keys. This functional is sometimes called the "shape functional" as it serves as a crude measure of the "shape" of the tree, with "full" trees (such as the complete tree) achieving the larger values of Q. For further discussions along these lines, consult [2] and [4]. If we define $f(T) := 0$ for $0 \leq |T| \leq m-2$ and $b_n := \ln \binom{n}{m-1}$ for $n \geq m-1$, then $f(T) = -\ln Q(T)$. Henceforth throughout this extended abstract we will refer to $-\ln Q$ (rather than Q) as the *shape functional*.

Several authors [18, 16, 1, 9] have studied additive functionals under the random permutation model. Clearly the random permutation model does not induce the uniform distribution on m-ary search trees with n keys since different permutations can give rise to the same tree. In this extended abstract we consider additive functionals under the uniform model, i.e., when each tree on n keys is considered equally likely. The shape functional for the case $m = 2$ (uniformly distributed binary search trees) was considered by Fill [4], who derived (limited) asymptotic information about its mean and variance. Limiting distributions for the shape functional and other additive functionals treated in the present extended abstract were identified in [7] for $m = 2$. We now generalize these results to include all values of m. What makes the analysis for general m significantly more intricate is that several key quantities (such as the number ρ discussed at the beginning of Section 3) are for general m known only implicitly.

One motivation for the present work can be understood in the context of the shape functional. The probability mass function Q corresponding to the random permutation model (a reasonably realistic model in practice) is an object of natural interest. Dobrow and Fill [2] determined the smallest and largest values of Q; but what are "typical" values? We can study this question probabilistically by placing a distribution on T and considering the distribution of $Q(T)$. Two rather natural choices for this distribution are Q itself (as treated in [9]) and the uniform distribution on trees (as treated herein).

We follow the "repertoire" approach of Greene and Knuth [13], determining the effect of a family of basic tolls (for example, those of the form n^α). Then the effect of a new toll could be determined by expressing it in terms of the basic tolls.

For tolls of the form n^α with $\alpha \geq 0$ and the tolls $\ln \binom{n}{m-1}$ and $\mathbf{1}_{n=m-1}$, we determine asymptotics of moments of all orders and our main results (Theorems 4.1, 5.1, 5.3, and 5.4) use these to yield limiting distributions. Here, in broad terms for the toll n^α, is a summary of lead-order results under both the random permutation model and the uniform model:

It is not surprising that the orders of magnitude under the uniform model are at least as large as under the random permutation model. Indeed, it is well known that trees produced by the uniform model are generally much "stringier" than trees produced by the random permutation model; for example, height is

Toll function n^α	Model: Random permutation	Uniform
α smaller than 1/2	n	n
α between 1/2 and 1	n	$n^{\alpha+\frac{1}{2}}$
α bigger than 1	n^α	$n^{\alpha+\frac{1}{2}}$

Tab. 1: Order of magnitude of the additive functional corresponding to the toll n^α.

of order $\sqrt{n}$ under the uniform model and order $\log n$ under the random permutation model. Furthermore "stringy" trees tend to give large values of the functional.

Qualitatively the uniform model differs significantly from the random permutation model, where, for example, there is a "phase change" in the limiting behavior at $m = 26$ from asymptotic normality to non-existence of a limiting distribution, for any toll whose order of growth does not exceed $n^{1/2}$; see [9] for precise results. On the other hand, for all m the uniform model leads to the normal distribution for the shape functional, space requirement, and number of leaves, and to (apparently) non-normal distributions for tolls of the form n^α with $\alpha > 0$.

We use methods from analytic combinatorics, in particular singularity analysis of generating functions [11], to derive the asymptotics of moments of the functional under consideration and then the method of moments to characterize the limiting distribution. A key singularity analysis tool is the newly-developed "Zigzag algorithm" [6] to handle Hadamard products of generating functions.

The limiting distributions (and even local limit theorems) for the space requirement and the number of leaves presumably can also be derived using Theorem 2 of [3] since the bivariate generating function for these parameters satisfy suitable functional equations. (This is *not* the case for the other tolls that we consider.) We include our proofs of these results for completeness and uniformity of treatment of tolls.

This extended abstract is organized as follows. In Section 2 we set up the problem using generating functions. In Section 3, a singular expansion for the generating function of the number of m-ary search trees on n keys is obtained. Sections 4 and 5 treat Examples 1.1–1.4.

Notation. Throughout, we will use $[z^n]f(z)$ to denote the coefficient of z^n in the Taylor series expansion of $f(z)$ around $z = 0$. We use $\mathcal{L}(Y)$ to denote the law (or distribution) of a random variable Y, the symbol $\stackrel{\mathcal{L}}{=}$ to denote equality in law, and $\stackrel{\mathcal{L}}{\longrightarrow}$ to denote convergence in law. We denote the (univariate) normal distribution with mean μ and variance σ^2 by $N(\mu, \sigma^2)$.

2 Preliminaries

Our starting point is the recursive construction of m-ary search trees. Let $X_n \equiv X_n(T)$ denote an additive functional on a random m-ary search tree T on n keys. Let $\mathbf{J} \equiv (J_1, \ldots, J_m)$ be the (random) vector of sizes of the subtrees rooted at the children of the root of T. If T is a uniformly distributed m-ary search tree on n keys, then X_n satisfies the distributional recurrence

$$X_n \stackrel{\mathcal{L}}{=} \sum_{k=1}^{m} X^{(k)}_{J_k} + b_n, \quad n \geq m-1, \tag{2.1}$$

with $(X_0, \ldots, X_{m-2}) =: \mathbf{x}$ denoting the vector of deterministic values of the functional for trees with fewer than $m-1$ keys. The sequence $(b_n)_{n \geq m-1}$ is called the *toll sequence*. On the right in (2.1),

- for each $k = 1, \ldots, m$, we have $X^{(k)}_j \stackrel{\mathcal{L}}{=} X_j$;
- the quantities $\mathbf{J}$; $X^{(1)}_0, \ldots, X^{(1)}_{n-(m-1)}$; $X^{(2)}_0, \ldots, X^{(2)}_{n-(m-1)}$; $\ldots$; $X^{(m)}_0, \ldots, X^{(m)}_{n-(m-1)}$ are all independent;
- the distribution of $\mathbf{J}$ if given by

$$\mathbf{P}\left[J_1 = j_1, \ldots, J_m = j_m\right] = \frac{\tau_{j_1} \cdots \tau_{j_m}}{\tau_n}, \tag{2.2}$$

for $(j_1, \ldots, j_m) \geq \mathbf{0}$ with $j_1 + \cdots + j_m = n - (m-1)$, where $\tau_k \equiv \tau_k(m)$ is the number of m-ary search trees on k keys.

(Throughout we take $m \geq 2$ to be fixed and so suppress the dependence of various parameters on m.)

Denote the sth moment of X_n by $\mu_n^{[s]} := \mathbf{E}\, X_n^s$. Now taking the sth power of (2.1) and conditioning on $(J_1, \ldots, J_m)$ gives

$$\mu_n^{[s]} = \sum_{s_0+\cdots+s_m=s} \binom{s}{s_0, \ldots, s_m} b_n^{s_0} \sum{}^{*} \frac{\tau_{j_1} \cdots \tau_{j_m}}{\tau_n} \mu_{j_1}^{[s_1]} \cdots \mu_{j_m}^{[s_m]},$$

where $\sum^{*}$ denotes the sum over all m-tuples $(j_1, \ldots, j_m) \geq \mathbf{0}$ such that $\sum_{i=1}^{m} j_i = n - (m-1)$. Isolating the terms in the sum where $s_i = s$ for some $i \in [m]$, we get (after some rearrangement)

$$\tau_n \mu_n^{[s]} = m \sum_{j_1=0}^{n-(m-1)} \tau_{j_1} \mu_{j_1}^{[s_1]} \sum_{j_2+\cdots+j_m=n-(m-1)-j_1} \tau_{j_2} \cdots \tau_{j_m} + r_n^{[s]}, \tag{2.3}$$

where

$$r_n^{[s]} := \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_0, \ldots, s_m} b_n^{s_0} \sum{}^{*} \tau_{j_1} \mu_{j_1}^{[s_1]} \cdots \tau_{j_m} \mu_{j_m}^{[s_m]}. \tag{2.4}$$

Let $\mu^{[s]}(z)$, $r^{[s]}(z)$, $\tau(z)$ denote the ordinary generating functions of $(\tau_n \mu_n^{[s]})_{n\geq 0}$, $(r_n^{[s]})_{n\geq 0}$, $(\tau_n)_{n\geq 0}$ respectively. Multiplying (2.3) by z^n and summing over $n \geq m-1$ yields (observe that $\tau_0 = \cdots = \tau_{m-2} = 1$ and $r_0^{[s]} = \cdots = r_{m-2}^{[s]} = 0$)

$$\mu^{[s]}(z) - \sum_{j=0}^{m-2} x_j^s z^j = m z^{m-1} \mu^{[s]}(z) \tau^{m-1}(z) + r^{[s]}(z),$$

so that

$$\mu^{[s]}(z) = \frac{r^{[s]}(z) + \sum_{j=0}^{m-2} x_j^s z^j}{1 - m[z\tau(z)]^{m-1}}. \tag{2.5}$$

Furthermore

$$r^{[s]}(z) = \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_0, \ldots, s_m} b^{\odot s_0}(z) \odot \left(z^{m-1} \mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) \right), \tag{2.6}$$

where $b(z) := \sum_{n=0}^{\infty} b_n z^n$ and $f(z) \odot g(z) \equiv (f \odot g)(z)$ is the Hadamard product of the power series f and g. Note that since $[z^n] \left(z^{m-1} \mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) \right) = 0$ for $0 \leq n \leq m-2$ we may instead use $b(z) := \sum_{n=m-1}^{\infty} b_n z^n$ when convenient.

3 Singular expansions

We will employ singularity analysis [11, 10, 6] to derive asymptotics of $\mu_n^{[s]}$ using (2.5). In order to do so we need a singular expansion for $\tau(z)$ around its dominant singularity. We will use the theory of analytic continuation of algebraic functions (see, for example, [19, §III.45] or [12, §VII.4]) to derive such an expansion. The terminology used is from [12, §VII.4].

Before we begin, we note that Fill and Dobrow [5] were able to use large-deviations techniques to obtain lead-order asymptotics of τ_n. However their techniques do not seem to be sufficient to derive the higher-order results we will need.

We now proceed with our analytic approach. As observed by Fill and Dobrow [5], it follows from the recursive definition of m-ary search trees that

$$\tau(z) - \sum_{j=0}^{m-2} z^j = z^{m-1} \tau^m(z). \tag{3.1}$$

Thus $\tau(z)$ is an algebraic series satisfying $P(z, \tau(z)) = 0$, where

$$P(z, w) := z^{m-1} w^m - w + \sum_{j=0}^{m-2} z^j. \tag{3.2}$$

The exceptional set of P [excluding $z = 0$, at which $\tau(z)$ clearly has no singularity] is

$$\bigcup_{w\in\mathbb{C}}\Big\{z\colon P(z,w)=0 \text{ and } \frac{\partial}{\partial w}P(z,w)=0\Big\} = \bigcup_{w\in\mathbb{C}}\Big\{z\colon z^{m-1}w^m - w + \sum_{j=0}^{m-2} z^j = 0 \text{ and } m(zw)^{m-1}-1=0\Big\}$$
$$= \Big\{z\colon m^m\Big(\sum_{j=1}^{m-1} z^j\Big)^{m-1} = (m-1)^{m-1}\Big\}.$$

The singularities of $\tau(z)$ lie in the exceptional set. It is clear [5, Theorem 3.1] that there exists a unique $\rho \in (0,1)$ contained in this set. Furthermore, since the Taylor coefficients of $\tau(z)$ are nonnegative, by Pringsheim's theorem [19, Theorem I.17.13], ρ is a dominant singularity of $\tau(z)$. It is straightforward to check that the polynomial system given by writing $P(z,w) = 0$ in the form $w = \Phi(z,w)$ is a-proper, a-positive, a-irreducible, and a-aperiodic (cf. [12, §VII.4.2]), so that by Theorem VII.7 of [12] we have that ρ is the unique dominant singularity and as $z \to \rho$ a singular expansion of the form

$$\tau(z) \sim \sum_{l\geq 0} a_l(1-\rho^{-1}z)^{l/2}. \tag{3.3}$$

Remark 3.1. Singularity analysis immediately yields from (3.3) a complete asymptotic expansion for τ_n, the number of m-ary search trees on n keys:

$$\tau_n \sim \rho^{-n}\sum_{l\geq 0}\frac{a_{2l+1}}{\Gamma(-l-\frac{1}{2})}n^{-l-\frac{3}{2}}. \tag{3.4}$$

In particular,

$$\tau_n = [1+O(n^{-1})]\frac{-a_1}{2\sqrt{\pi}}n^{-3/2}\rho^{-n}.$$

3.1 Determination of the coefficients a_l

Define $w_\rho := \frac{m}{m-1}\sum_{j=0}^{m-2}\rho^j$, so that

$$P(\rho, w_\rho) = 0 \text{ and } \frac{\partial}{\partial w}P(\rho,w)\Big|_{w=w_\rho} = 0.$$

Using the definition of ρ and the fact that $w_\rho > 0$ by definition, we have $w_\rho = m^{-\frac{1}{m-1}}\rho^{-1}$. Now $\frac{\partial}{\partial w}P(\rho,w)$ is negative, zero, or positive as $w > 0$ is less than, equal to, or greater than w_ρ. Hence, for $w > 0$, $P(\rho,w) = 0$ if and only if $w = w_\rho$. But $a_0 > 0$ and $0 = P(\rho,\tau(\rho)) = P(\rho,a_0)$, so that

$$a_0 = w_\rho = m^{-\frac{1}{m-1}}\rho^{-1}. \tag{3.5}$$

To obtain values of a_l for $l \geq 1$, we rewrite (3.1) for $z \neq 1$ as

$$z^{m-1}\tau^m(z) - \tau(z) + \frac{1-z^{m-1}}{1-z} = 0$$

and, then defining $Z := 1-\rho^{-1}z$, equivalently as

$$1+\rho^{m-1}(1-Z)^{m-1}\big[(1-\rho+\rho Z)\tau^m(z)-1\big] - (1-\rho+\rho Z)\tau(z). \tag{3.6}$$

By comparing the coefficients of Z in this equation and observing that $a_1 < 0$ we obtain

$$a_1 = -\sqrt{2m\alpha^*}m^{-\frac{m}{m-1}}\rho^{-1}, \tag{3.7}$$

where, matching the notation of [5], we define the key quantity

$$\alpha^* := m - \big(m^{\frac{m}{m-1}}-1\big)(\rho^{-1}-1)^{-1}. \tag{3.8}$$

In the sequel we will also need the following relation, which follows from comparing coefficients of $Z^{3/2}$ in (3.6):

$$\frac{a_0(a_0-a_2)}{a_1^2} = \frac{m-2}{6}. \tag{3.9}$$

Let $\mathcal{A}$ denote a generic (formal) power series in Z, possibly different at each appearance. Similarly, let $\mathcal{P}_d$ denote a generic polynomial in Z of degree at most d. In the sequel we will likewise use $\mathcal{N}$ to denote a generic (formal) power series in powers of n^{-1}. Then, using (3.3) and (3.5), we have

$$(1 - m[z\tau(z)]^{m-1})^{-1} \sim \frac{a_0}{-a_1(m-1)} Z^{-1/2} + c_0 + Z^{1/2}\mathcal{A} + Z\mathcal{A}, \tag{3.10}$$

where, using (3.9), we have

$$c_0 := \frac{m-2}{3(m-1)}; \tag{3.11}$$

$$z^{m-1}\tau^m(z) \sim a_0 m^{-1} + a_1 Z^{1/2} + Z\mathcal{A} + Z^{3/2}\mathcal{A}; \tag{3.12}$$

and

$$\sum_{j=0}^{m-2} x_j^s z^j = \sum_{j=0}^{m-2} x_j^s \rho^j + Z\mathcal{P}_{m-3}. \tag{3.13}$$

Thus, by singularity analysis,

$$[z^n]\left[z^{m-1}\tau^m(z)\right] \sim n^{-3/2}\rho^{-n}\left(\frac{-a_1}{2\sqrt{\pi}} + n^{-1}\mathcal{N}\right). \tag{3.14}$$

3.2 Zigzag algorithm

For the reader's convenience we present the Zigzag algorithm, which is used extensively in the rest of this paper to determine singular expansions of Hadamard products. The validity of the algorithm was established recently in [6], to which the reader is referred for further background discussion.

"Zigzag" Algorithm. [Computes the singular expansion of $f \odot g$ up to $O(|1-z|^C)$.]

1. Use singularity analysis to determine separately the asymptotic expansions of $f_n = [z^n]f(z)$ and $g_n = [z^n]g(z)$ into descending powers of n.

2. Multiply the resulting expansions and reorganize to obtain an asymptotic expansion for the product $f_n g_n$.

3. Choose a basis $\mathcal{B}$ of singular functions, for instance, the standard basis $\mathcal{B} = \left\{(1-z)^\beta \ln[(1-z)]^k\right\}$. Construct a function $H(z)$ expressed in terms of $\mathcal{B}$ whose singular behavior is such that the asymptotic form of its coefficients h_n is compatible with that of $f_n g_n$ up to the needed error terms.

4. Output the singular expansion of $f \odot g$ as the quantity $H(z) + P(z) + O(|1-z|^C)$, where P is a polynomial in $(1-z)$ of degree less than C.

The reason for the addition of a polynomial in Step **4** is that integral powers of $(1-z)$ do not leave a trace in coefficient asymptotics since their contribution is asymptotically null. The Zigzag Algorithm is principally useful for determining the divergent part of expansions. If needed, the coefficients in the polynomial P can be expressed as values of the function $f \odot g$ and its derivatives at 1 once it has been stripped of its nondifferentiable terms.

4 The toll n^α

Here is the main theorem of this section. The case $\alpha = 1/2$ is treated by special means in [8].

Theorem 4.1. *Let $\alpha \neq 1/2$, and let X_n denote the additive functional that satisfies the distributional recurrence* (2.1) *with $b_n \equiv n^\alpha$ and initial conditions $(x_0, \dots, x_{m-2})$. Define $\alpha' := \alpha + \frac{1}{2}$ and recall* (3.8).

(a) If $\alpha > 1/2$, then

$$(m-1)(m\alpha^*)^{1/2}\frac{X_n}{n^{\alpha'}} \xrightarrow{\mathcal{L}} Y_\alpha;$$

(b) if $\alpha < 1/2$, then

$$\frac{(m-1)(m\alpha^*)^{1/2}}{n^{\alpha'}}\left[X_n - \frac{\rho m^{\frac{m}{m-1}} C_\alpha}{(m-1)\alpha^*}(n+1)\right] \xrightarrow{\mathcal{L}} Y_\alpha, \quad \textit{where } C_\alpha := \sum_{n=m-1}^{\infty} \rho^n n^\alpha \tau_n + \sum_{j=0}^{m-2} x_j \rho^j.$$

In either case we have convergence of all moments, where Y_α has the unique distribution whose moments are given by $\mathbf{E}\, Y_\alpha^s = M_s \equiv M_s(\alpha)$. Here

$$M_1 = \frac{\Gamma(\alpha - \frac{1}{2})}{\sqrt{2}\Gamma(\alpha)},$$

and, for $s \geq 2$,

$$M_s = \frac{1}{4\sqrt{\pi}} \sum_{j=1}^{s-1} \binom{s}{j} \frac{\Gamma(j\alpha' - \frac{1}{2})\Gamma((s-j)\alpha' - \frac{1}{2})}{\Gamma(s\alpha' - \frac{1}{2})} M_j M_{s-j} + \frac{s\Gamma(s\alpha' - 1)}{\sqrt{2}\Gamma(s\alpha' - \frac{1}{2})} M_{s-1}.$$

Although the normalization required to produce a limiting distribution depends on m, Theorem 4.1 exhibits a striking *invariance principle*: the distributions $\mathcal{L}(Y_\alpha)$ do not depend on the value of m (and thus in particular, have already arisen when $m = 2$ in [7]).

We will present the proof of Theorem 4.1 only for the simplest case $\alpha \in (1/2, \infty) \setminus \{3/2, 5/2, \ldots\}$; consult [8] for the other cases.

4.1 Mean

Using $s = 1$ in (2.6) we have

$$r^{[1]}(z) = b(z) \odot \left[z^{m-1}\tau^m(z)\right],$$

and consequently, by (2.4) and (3.4),

$$[z^n]r^{[1]}(z) = r_n^{[1]} = b_n\tau_n \sim n^{\alpha - \frac{3}{2}}\rho^{-n}\left(\frac{-a_1}{2\sqrt{\pi}} + n^{-1}\mathcal{N}\right). \tag{4.1}$$

We employ the Zigzag Algorithm outlined in Section 3.2. A compatible singular expansion for $r^{[1]}(z)$ is

$$r^{[1]}(z) \sim \frac{-a_1}{2\sqrt{\pi}}\Gamma(\alpha - \tfrac{1}{2})Z^{-\alpha+\frac{1}{2}} + Z^{-\alpha+\frac{3}{2}}\mathcal{A} + \mathcal{A}. \tag{4.2}$$

Using (3.10) and (4.2) in (2.5) we obtain

$$\mu^{[1]}(z) \sim \frac{a_0\Gamma(\alpha - \frac{1}{2})}{2\sqrt{\pi}(m-1)}Z^{-\alpha} + Z^{-\alpha+\frac{1}{2}}\mathcal{A} + Z^{-\alpha+1}\mathcal{A} + Z^{-1/2}\mathcal{A} + \mathcal{A}, \tag{4.3}$$

whence, by singularity analysis,

$$\rho^n\mu_n^{[1]}\tau_n \sim \frac{a_0\Gamma(\alpha - \frac{1}{2})}{2\sqrt{\pi}(m-1)\Gamma(\alpha)}n^{\alpha-1} + n^{\alpha-\frac{3}{2}}\mathcal{N} + n^{\alpha-2}\mathcal{N} + n^{-1/2}\mathcal{N}.$$

The singular expansion for τ_n at (3.4) then gives

$$\mu_n^{[1]} \sim \frac{a_0\Gamma(\alpha - \frac{1}{2})}{(-a_1)(m-1)\Gamma(\alpha)}n^{\alpha+\frac{1}{2}} + n^\alpha\mathcal{N} + n^{\alpha-\frac{1}{2}}\mathcal{N} + n\mathcal{N}.$$

4.2 Higher moments

We will use induction to obtain asymptotics for higher-order moments. Throughout $\alpha' := \alpha + \frac{1}{2}$.

Proposition 4.2. *Let $\alpha \in (1/2, \infty) \setminus \{3/2, 5/2, \ldots\}$. Then, for $s \geq 1$, and $\epsilon > 0$ small enough,*

$$\mu^{[s]}(z) = D_s Z^{-s\alpha'+\frac{1}{2}} + O(|Z|^{-s\alpha'+\frac{1}{2}+q}),$$

where $q := \min\{\alpha - \frac{1}{2}, \frac{1}{2}\} - \epsilon$ with

$$D_1 := \frac{a_0\Gamma(\alpha - \frac{1}{2})}{2(m-1)\sqrt{\pi}},$$

and, for $s \geq 2$,

$$D_s = \frac{a_0}{(m-1)(-a_1)}\left[\frac{m-1}{2a_0}\sum_{j=1}^{s-1}\binom{s}{j}D_jD_{s-j} + \frac{\Gamma(s\alpha' - 1)}{\Gamma((s-1)\alpha' - \frac{1}{2})}sD_{s-1}\right]. \tag{4.4}$$

Proof. We proceed by induction on s. For $s = 1$ the claim was proved as (4.3). [Note that $\mu^{[0]}(z) = \tau(z) \sim a_0$.] Suppose $s \geq 2$. We will first obtain the asymptotics of $r^{[s]}(z)$ at (2.6) by analyzing each of the terms in the sum there.

Suppose exactly $k \geq 1$ of $s_1, \ldots, s_m$, say $s_1, \ldots, s_k$, are nonzero. Then, by induction,

$$z^{m-1}\mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) = O(|Z|^{-(s-s_0)\alpha' + \frac{k}{2}}).$$

Moreover, if $s_0 = 0$ then the contribution to $r^{[s]}(z)$ is $O(|Z|^{-s\alpha' + \frac{3}{2}})$ unless $k = 1$ or $k = 2$. (Observe, however, that if $k = 1$ then s_0 cannot be zero as that would imply $s_1 = s$.) On the other hand, if $s_0 \neq 0$, then using singularity analysis for polylogarithms [10] and Hadamard products [6], we see that

$$b^{\odot s_0}(z) \odot [z^{m-1}\mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z)] = O(|Z|^{-s\alpha' + \frac{s_0}{2} + \frac{k}{2}}),$$

which is $O(|Z|^{-s\alpha' + \frac{3}{2} - \epsilon})$ unless $k = 1$ and $s_0 = 1$. (The ϵ term in the exponent avoids logarithmic factors that arise when $-s\alpha' + \frac{s_0}{2} + \frac{k}{2}$ is a nonnegative integer.)

If all of $s_1, \ldots, s_m$ are zero, then $s_0 = s$ and, using (3.12), the contribution to $r^{[s]}(z)$ is $O(|Z|^{-s\alpha' + \frac{s}{2} + \frac{1}{2}})$ which is $O(|Z|^{-s\alpha' + \frac{3}{2}})$.

Hence unless $s_0 = 0$ and exactly two of $s_1, \ldots, s_m$ are nonzero or $s_0 = 1$ and exactly one of $s_1, \ldots, s_m$ is $s-1$ in (2.6), the contribution to $r^{[s]}(z)$ is $O(|Z|^{-s\alpha' + \frac{3}{2} - \epsilon})$. In the former case the contribution to $r^{[s]}(z)$ is gotten by using the induction hypothesis as

$$\binom{m}{2}\rho^{m-1}Z^{-s\alpha'+1}a_0^{m-2}\sum_{j=1}^{s-1}\binom{s}{j}D_jD_{s-j} + O(|Z|^{-s\alpha'+1+q}).$$

In the latter case, again using the induction hypothesis and singularity analysis for Hadamard products we get the contribution to $r^{[s]}(z)$ as

$$m\rho^{m-1}a_0^{m-1}sD_{s-1}\frac{\Gamma(s\alpha' - 1)}{\Gamma((s-1)\alpha' - \frac{1}{2})}Z^{-s\alpha'+1} + O(|Z|^{-s\alpha'+1+q}).$$

Finally, noting that the contribution from $\sum_{j=0}^{m-2} x_j^s z^j$ to the numerator on the right side in (2.5) is negligible, we complete the induction by using (3.5) and (3.10). $\square$

4.3 Limiting distributions

We can now use the method of moments to derive limiting distributions for the additive functional.

Proof of Theorem 4.1 when $\alpha \in (1/2, \infty) \setminus \{3/2, 5/2, \ldots\}$. By Proposition 4.2, singularity analysis, and the asymptotics of τ_n at (3.4), we have

$$\mathbf{E}\, X_n^s = \mu_n^{[s]} = \frac{D_s 2\sqrt{\pi}}{(-a_1)\Gamma(s\alpha' - \frac{1}{2})}n^{s\alpha'} + O(n^{s\alpha'-q}).$$

Define $\sigma \equiv \sigma_m := -a_1(m-1)/(\sqrt{2}a_0) = (m-1)(\alpha^*/m)^{1/2}$, where the last equality uses (3.5), (3.7), and (3.8). Then, for fixed m, as $n \to \infty$,

$$\mathbf{E}\left[\sigma_m \frac{X_n}{n^{\alpha'}}\right]^s \to M_s,$$

where, for $s \geq 1$,

$$M_s := \frac{\sigma^s D_s 2\sqrt{\pi}}{(-a_1)\Gamma(s\alpha' - \frac{1}{2})}.$$

In particular, $M_1 = \Gamma(\alpha - \frac{1}{2})/[\sqrt{2}\Gamma(\alpha)]$. Furthermore, using (4.4), we obtain the recurrence for M_s.

Convergence in distribution follows from the fact that (M_s) satisfies Carleman's condition, as has been established in [7]. $\square$

5 Three asymptotically normal additive functionals

Examples 1.1, 1.2, and 1.4 can be treated in much the same way as the toll n^α; consult [8] for the details, which vary substantially from one example to another. Here are the results.

Theorem 5.1. *Let X_n denote the shape functional for uniformly distributed m-ary search trees on n keys. Then*

$$\frac{X_n - d_1(n+1)}{\sqrt{n \ln n}} \xrightarrow{\mathcal{L}} N(0,\sigma^2) \quad \text{and} \quad \frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} N(0,1),$$

$$\text{where} \quad d_1 := \frac{2a_0}{(m-1)a_1^2} \sum_{n=m-1}^{\infty} \rho^n \left[\ln \binom{n}{m-1} \right] \tau_n \quad \text{and} \quad \sigma^2 := 8\,(a_0/a_1)^2\,(1-\ln 2).$$

Remark 5.2. It is known [1, 9] that under the random permutation model the shape functional normalized by its mean and standard deviation is asymptotically normal for $2 \le m \le 26$ and does not have a limiting distribution for $m > 26$. In contrast, under the uniform model we have asymptotic normality for all $m \ge 2$.

Theorem 5.3. *Let X_n denote the space requirement for uniformly distributed m-ary search trees on n keys. Then*

$$\frac{X_n - d_1(n+1)}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,\sigma^2) \quad \text{and} \quad \frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} N(0,1),$$

$$\text{where} \quad d_1 := \frac{m(1-\rho m^{\frac{1}{m-1}})}{(m-1)\alpha^*} \quad \text{and} \quad \sigma^2 := \frac{2a_0}{a_1^2(m-1)} \left[d_1^2 + \sum_{j=1}^{m-2} [-(j+1)d_1 + 1]^2 \rho^j - \frac{a_0}{m(m-1)} \right].$$

Theorem 5.4. *Let X_n denote the number of leaves in a uniformly distributed m-ary search tree on n keys. Then*

$$\frac{X_n - \frac{\rho}{\alpha^*}(n+1)}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,\sigma^2) \quad \text{and} \quad \frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} N(0,1),$$

$$\text{where} \quad \sigma^2 := \frac{\rho m^{\frac{m}{m-1}}(\rho^{m-1} + \delta_1)}{\alpha^*(m-1)}, \quad \text{with} \quad \delta_1 := (\rho/\alpha^*)^2 + \sum_{j=1}^{m-2} \left[-(j+1)(\rho/\alpha^*) + 1 \right]^2 \rho^j.$$

References

[1] H.-H. Chern and H.-K. Hwang. Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms*, 19(3-4):316–358, 2001. Analysis of algorithms (Krynica Morska, 2000).

[2] R. P. Dobrow and J. A. Fill. Multiway trees of maximum and minimum probability under the random permutation model. *Combin. Probab. Comput.*, 5(4):351–371, 1996.

[3] M. Drmota. Asymptotic distributions and a multivariate Darboux method in enumeration problems. *J. Combin. Theory Ser. A*, 67(2):169–184, 1994.

[4] J. A. Fill. On the distribution of binary search trees under the random permutation model. *Random Structures Algorithms*, 8(1):1–25, 1996.

[5] J. A. Fill and R. P. Dobrow. The number of m-ary search trees on n keys. *Combin. Probab. Comput.*, 6(4):435–453, 1997.

[6] J. A. Fill, P. Flajolet, and N. Kapur. Singularity analysis, Hadamard products, and tree recurrences. *J. Comput. Appl. Math.*, 174(2):271–313, 2005, arXiv:math.CO/0306225.

[7] J. A. Fill and N. Kapur. Limiting distributions for additive functionals on Catalan trees. *Theoret. Comput. Sci.*, 326:69–102, 2004, arXiv:math.PR/0306226.

[8] J. A. Fill and N. Kapur. A repertoire for additive functionals of uniformly distributed m-ary search trees, 2005, arXiv:math.PR/0502422. Full-length paper corresponding to this extended abstract.

[9] J. A. Fill and N. Kapur. Transfer theorems and asymptotic distributional results for m-ary search trees, 2005, arXiv:math.PR/0306050. To appear, *Random Structures Algorithms*.

[10] P. Flajolet. Singularity analysis and asymptotics of Bernoulli sums. *Theoret. Comput. Sci.*, 215(1-2):371–381, 1999.

[11] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[12] P. Flajolet and R. Sedgewick. The Average Case Analysis of Algorithms. Book in preparation. A draft is available from `http://pauillac.inria.fr/algo/flajolet/Publications/publist.html`, 200x.

[13] D. H. Greene and D. E. Knuth. *Mathematics for the analysis of algorithms*, volume 1 of *Progress in Computer Science and Applied Logic*. Birkhäuser Boston Inc., Boston, MA, 1990.

[14] D. E. Knuth. *The art of computer programming. Volume 1*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 3rd edition, 1997.

[15] D. E. Knuth. *The art of computer programming. Volume 3*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 2nd edition, 1998.

[16] W. Lew and H. M. Mahmoud. The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.*, 23(5):1050–1074, 1994.

[17] H. M. Mahmoud. *Evolution of random search trees*. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[18] H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms*, 10(1):52–75, 1989.

[19] A. I. Markushevich. *Theory of functions of a complex variable*. Translated and edited by Richard A. Silverman. Chelsea Publishing Company, New York, N.Y., 1977.

[20] R. R. Muntz and R. C. Uzgalis. Dynamic storage allocation for binary search trees in a two-level memory. In *Proceedings of Princeton Conference on Information Sciences and Systems*, volume 4, pages 345–349, 1971.