A REPERTOIRE FOR ADDITIVE FUNCTIONALS OF UNIFORMLY DISTRIBUTED m-ARY SEARCH TREES

JAMES ALLEN FILL AND NEVIN KAPUR

Abstract. 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.

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.

2000 *Mathematics Subject Classification.* Primary: 68W40; Secondary: 60F05, 60C05.

Key words and phrases. additive functionals, Hadamard products, limit laws, method of moments, search trees, shape functional, singularity analysis, space requirement, leaves.

The research of James Allen Fill 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 partially supported by NSF Grant 0049092, and the by Center for Mathematics of Mathematics of Information at the California Institute of Technology.

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

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

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 [3] 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 [3] and [5]. 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 paper we will refer to $-\ln Q$ (rather than Q) as the *shape functional*.

Several authors [18, 16, 2, 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 paper 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 [5], who derived (limited) asymptotic information about its mean and variance. Limiting distributions for the shape functional and other additive functionals treated in the present paper were identified in [8] 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 paper 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 [3] 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 4.5, 4.6, 5.2, 6.2, and 7.2) 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:

	Model	
Toll function n^α	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	$n^{\alpha+\frac{1}{2}}$

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

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 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" [7] 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 [4] 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.

The paper 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, 5, 6, and 7 derive limiting distributions for the additive functionals corresponding to the tolls n^α ($\alpha > 0$), $\ln \binom{n}{m-1}$ (shape functional), 1 (space requirement), and $\mathbf{1}_{n=m-1}$ (number of leaves), respectively.

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 $\stackrel{\mathcal{L}}{\to}$ 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 $\mathbf{J} \equiv (J_1, \dots, J_m)$ 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 $(X_0, \dots, X_{m-2}) =: \mathbf{x}$ 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*. In (2.1), $\stackrel{\mathcal{L}}{=}$ denotes equality in law (i.e., in distribution), and on the right,

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

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

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

(Throughout we will take $m \geq 2$ to be fixed and so will 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

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

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

$$\begin{aligned} \tau_n \mu_n^{[s]} &= m \sum\nolimits^* \tau_{j_1} \cdots \tau_{j_m}\, \mu_{j_1}^{[s]} \\ &\quad + \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_0, \ldots, s_m} b_n^{s_0} \sum\nolimits^* \tau_{j_1} \mu_{j_1}^{[s_1]} \cdots \tau_{j_m} \mu_{j_m}^{[s_m]} \\ &= m \sum_{j_1=0}^{n-(m-1)} \tau_{j_1} \mu_{j_1}^{[s_1]} \sum_{j_2+\cdots+j_m=n-(m-1)-j_1} \tau_{j_2} \cdots \tau_{j_m} + r_n^{[s]}, \end{aligned} \tag{2.3}$$

where

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

Introduce generating functions

$$\mu^{[s]}(z) := \sum_{n=0}^{\infty} \tau_n \mu_n^{[s]} z^n, \quad r^{[s]}(z) := \sum_{n=0}^{\infty} r_n^{[s]} z^n, \quad \tau(z) := \sum_{n=0}^{\infty} \tau_n z^n.$$

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

$$\begin{aligned} r^{[s]}(z) &= \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_0, \ldots, s_m} \sum_{n=0}^{\infty} b_n^{s_0} \left\{ [z^{n-(m-1)}] \left[\mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) \right] \right\} z^n \\ &= \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_0, \ldots, s_m} b^{\odot s_0}(z) \odot \left(z^{m-1} \mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) \right), \end{aligned} \tag{2.6}$$

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 $0 \le n \le m-2$ 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, 7] 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 [6] 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 [6], 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

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

The singularities of $\tau(z)$ lie in the exceptional set. It is clear [6, 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

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

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

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

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 [6], we define the key quantity

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

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. **Generalized polylogarithms.** For α an arbitrary complex number and r a nonnegative integer, the *generalized polylogarithm* function $\mathrm{Li}_{\alpha,r}$ is defined for $|z| < 1$ by

$$\mathrm{Li}_{\alpha,r}(z) := \sum_{n=1}^{\infty} \frac{(\ln n)^r}{n^\alpha} z^n.$$

We record here three singular expansions (as $z \to \rho$) that are computed using singularity analysis of generalized polylogarithms [10]:

$$\begin{aligned}
\mathrm{Li}_{3/2,0}(\rho^{-1}z) &\sim \zeta(3/2) - 2\sqrt{\pi}Z^{1/2} + Z\mathcal{A} + Z^{3/2}\mathcal{A},\\
\mathrm{Li}_{3/2,1}(\rho^{-1}z) &\sim -\zeta'(3/2) - 2\sqrt{\pi}Z^{1/2}\ln Z^{-1} - 2\sqrt{\pi}[2(1-\ln 2)-\gamma]Z^{1/2}\\
&\qquad + Z\mathcal{A} + (Z^{3/2}\log Z)\mathcal{A} + Z^{3/2}\mathcal{A},\\
\mathrm{Li}_{1,1}(\rho^{-1}z) &\sim \frac{1}{2}\ln^2 Z^{-1} - \gamma\ln Z^{-1} + \mathcal{A} + (Z\log Z)\mathcal{A}.
\end{aligned} \tag{3.15}$$

These expansions will be utilized in the analysis of the shape functional in Section 5.

3.3. **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 [7], 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 $\mathcal{B} = \{(1-z)^\beta[\ln(1-z)]^k\}$, or the polylogarithm basis $\mathcal{B} = \{\mathrm{Li}_{\beta,k}(z)\}$. 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

The main theorems of this section are Theorems 4.5 and 4.6, which give limiting distributions for the additive functionals corresponding to the tolls n^α with $\alpha > 0$. Although the normalization required to produce a limiting distribution depends on m, our results exhibit a striking *invariance principle*: The limiting distributions themselves do not depend on the value of m (and thus in particular, have already arisen when $m = 2$ in [8]).

4.1. **Mean.** We consider the mean for the toll $b_n \equiv n^\alpha$, where $\alpha > 0$. 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),

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

Until further notice, assume that $\alpha \notin \{1/2, 3/2, \dots\}$. (The contrary cases are considered later in this section.) We employ the Zigzag Algorithm outlined in Section 3.3. A compatible singular expansion for $r^{[1]}(z)$ is given by

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

If $\alpha > 1/2$, then using (3.10), (3.13), 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

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

On the other hand, if $\alpha < 1/2$, the dominant term in (4.2) is now the constant term so that

$$r^{[1]}(z) + \sum_{j=0}^{m-2} x_j z^j \sim C_\alpha + \frac{-a_1}{2\sqrt{\pi}}\Gamma(\alpha - \tfrac{1}{2})Z^{-\alpha+\frac{1}{2}} + Z\mathcal{A} + Z^{-\alpha+\frac{3}{2}}\mathcal{A}, \tag{4.4}$$

where

$$C_\alpha := r^{[1]}(\rho) + \sum_{j=0}^{m-2} x_j\rho^j = \sum_{n=m-1}^{\infty}\rho^n n^\alpha\tau_n + \sum_{j=0}^{m-2} x_j\rho^j. \tag{4.5}$$

Then, using (3.10) and (4.4) in (2.5) we obtain

$$\mu^{[1]}(z) \sim \frac{a_0 C_\alpha}{(m-1)(-a_1)}Z^{-1/2} + \frac{a_0\Gamma(\alpha - \frac{1}{2})}{2\sqrt{\pi}(m-1)}Z^{-\alpha} + \mathcal{A} + Z^{-\alpha+\frac{1}{2}}\mathcal{A} + Z^{-\alpha+1}\mathcal{A} + Z^{1/2}\mathcal{A}, \tag{4.6}$$

whence singularity analysis and the singular expansion of τ_n yields

$$\mu_n^{[1]} \sim \frac{2a_0C_\alpha}{(m-1)a_1^2}n + \frac{a_0\Gamma(\alpha-\frac{1}{2})}{(-a_1)(m-1)\Gamma(\alpha)}n^{\alpha+\frac{1}{2}} + n^\alpha\mathcal{N} + \mathcal{N} + n^{\alpha-\frac{1}{2}}\mathcal{N}. \tag{4.7}$$

If $\alpha \in \{3/2, 5/2, \dots\}$, then logarithmic terms appear in the singular expansion compatible with (4.1), so that

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

This leads to

$$\mu^{[1]}(z) \sim \frac{a_0\Gamma(\alpha-\frac{1}{2})}{2\sqrt{\pi}(m-1)}Z^{-\alpha}+Z^{-\alpha+\frac{1}{2}}\mathcal{A}+Z^{-\alpha+1}\mathcal{A}+(Z^{-1/2}\log Z)\mathcal{A}+(\log Z)\mathcal{A} \tag{4.8}$$

and consequently

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

Observe that the lead-order term and the order of growth of the remainder $[O(|Z|^{-\alpha+\frac{1}{2}})]$ in the expansion of $\mu^{[1]}(z)$ at (4.8) are the same as at (4.3).

Finally, we consider $\alpha = 1/2$. Now, a singular expansion compatible with (4.1) is given by

$$r^{[1]}(z) \sim \frac{-a_1}{2\sqrt{\pi}}\ln Z^{-1} + C_{1/2} + (Z\log Z)\mathcal{A} + Z\mathcal{A},$$

where the constant term is

$$C_{1/2} := \sum_{n=m-1}^{\infty}\left(n^{1/2}\rho^n\tau_n + \frac{a_1}{2\sqrt{\pi n}}\right). \tag{4.9}$$

Thus

$$r^{[1]}(z) + \sum_{j=0}^{m-2}x_jz^j \sim \frac{-a_1}{2\sqrt{\pi}}\ln Z^{-1} + C'_{1/2} + (Z\log Z)\mathcal{A} + Z\mathcal{A},$$

where

$$C'_{1/2} := C_{1/2} + \sum_{j=0}^{m-2}x_j\rho^j. \tag{4.10}$$

Using (2.5) and (3.10), we get

$$\begin{aligned}\mu^{[1]}(z) \sim{}& \frac{a_0}{2\sqrt{\pi}(m-1)}Z^{-1/2}\ln Z^{-1} + \frac{a_0C'_{1/2}}{-a_1(m-1)}Z^{-1/2}\\ &+ (\log Z)\mathcal{A} + \mathcal{A} + (Z^{1/2}\log Z)\mathcal{A} + Z^{1/2}\mathcal{A}.\end{aligned} \tag{4.11}$$

Using singularity analysis and (3.4) we conclude

$$\mu_n^{[1]} \sim \frac{a_0}{-a_1\sqrt{\pi}(m-1)}n\ln n + \eta_{1/2}n + n^{1/2}\mathcal{N} + (\log n)\mathcal{N} + \mathcal{N}, \tag{4.12}$$

where

$$\eta_{1/2} := \frac{2\sqrt{\pi}}{-a_1}\left(\frac{a_0(\gamma+2\ln 2)}{2\pi(m-1)} + \frac{a_0C'_{1/2}}{-a_1\sqrt{\pi}(m-1)}\right).$$

4.2. **Higher moments.** We will use induction to obtain asymptotics for higher-order moments. Throughout $\alpha' := \alpha + \frac{1}{2}$. We consider the case $\alpha > 1/2$ in Proposition 4.1 and handle the remaining cases in Propositions 4.2 and 4.3.

Proposition 4.1. *Let $\alpha > 1/2$. 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 $s \geq 2$,

$$D_s = \frac{a_0}{(m-1)(-a_1)} \left[\frac{m-1}{2a_0} \sum_{j=1}^{s-1} \binom{s}{j} D_j D_{s-j} + \frac{\Gamma(s\alpha' - 1)}{\Gamma((s-1)\alpha' - \frac{1}{2})} s D_{s-1} \right]. \tag{4.13}$$

Proof. We proceed by induction on s. For $s = 1$ the claim was proved as (4.3) and (4.8). [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 [7], 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

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

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

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

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). □

For $\alpha < 1/2$, it will be convenient to consider instead the "approximately centered" random variable

$$\widetilde{X}_n := X_n - \frac{2a_0 C_\alpha}{(m-1)a_1^2}(n+1) = X_n - \frac{\rho m^{\frac{m}{m-1}} C_\alpha}{(m-1)\alpha^*}(n+1), \tag{4.14}$$

where C_α is defined at (4.5) and α^* at (3.8). See (4.7) for the motivation behind this definition. The choice of centering by a multiple of $n+1$ rather than n is motivated by the fact that with this centering $\widetilde{X}_n$ satisfies the same distributional recurrence (2.1) as X_n [with appropriate initial conditions $(\widetilde{X}_0, \ldots, \widetilde{X}_{m-2})$]. We will use $\tilde{r}^{[s]}(z)$, $\tilde{\mu}^{[s]}(z)$, and $\tilde{\mu}_n^{[s]}$ to denote the analogous quantities for $(\widetilde{X}_n)$.

Proposition 4.2. *Let $\alpha < 1/2$. Then, for $s \geq 1$, and $\epsilon > 0$ small enough,*

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

where c_s is a constant and D_s is defined as in Proposition 4.1.

Proof outline. The basis of the induction is (4.6) and the induction step is identical to the one in the proof of Proposition 4.1. We omit the details. □

When $\alpha = 1/2$, we define

$$\widetilde{X}_n := X_n - \frac{2a_0 C'_{1/2}}{(m-1)a_1^2}(n+1) = X_n - \frac{\rho m^{\frac{m}{m-1}} C'_{1/2}}{(m-1)\alpha^*}(n+1). \tag{4.15}$$

The constant $C'_{1/2}$ is defined at (4.10) using (4.9). The key result here is the following.

Proposition 4.3. *Let $\alpha = 1/2$. Define $\sigma_m := -a_1(m-1)/(\sqrt{2}a_0)$. Then, for $s \geq 1$,*

$$\tilde{\mu}^{[s]}(z) = -a_1 \sigma_m^{-s} Z^{-s+\frac{1}{2}} \sum_{r=0}^{s} C_{s,r} (\ln Z^{-1})^{s-r} + O(|Z|^{-s+1-\epsilon}),$$

where the constants $C_{s,r}$ do not depend on m.

Proof sketch. The form of the proof is the same as those of Propositions 4.1 and 4.2. The basis of the induction is (4.11). [Note that without the centering we have done we would be saddled with the $Z^{-1/2}$ term, whose coefficient depends on m, in (4.11).] For the induction step, in estimating $\tilde{r}^{[s]}(z)$, unless

(1) $s_0 = 0$, and exactly two of $s_1, \ldots, s_m$ are nonzero; or
(2) $s_0 = 1$, and exactly one of $s_1, \ldots, s_m$ is nonzero,

the contribution is $O(|Z|^{-s+\frac{3}{2}-\epsilon})$.

In case (1), by induction, the contribution to $\tilde{r}^{[s]}(z)$ is

$$a_1^2 \sigma_m^{-s} \binom{m}{2} \rho^{m-1} a_0^{m-2} Z^{-s+1} \sum_{r=0}^{s} F_{s,r} (\ln Z^{-1})^{s-r} + O(|Z|^{-s+\frac{3}{2}-\epsilon}),$$

where the $(F_{s,r})$ do not depend on m. Using (3.5) and the definition of σ_m, we see that the constant multiplying the lead sum is

$$a_1^2 \sigma_m^{-s} \binom{m}{2} \rho^{m-1} a_0^{m-2} = -a_1 \sigma_m^{-(s-1)}/\sqrt{2}.$$

In case (2), the contribution is

$$-a_1\sigma_m^{-(s-1)} m\rho^{m-1} a_0^{m-1} Z^{-s+1} \sum_{r=0}^{s-1} G_{s,r} (\ln Z^{-1})^{s-1-r} + O(|Z|^{-s+\frac{3}{2}-\epsilon}),$$

where the $(G_{s,r})$ do not depend on m. Again, using (3.5), the constant multiplying the lead sum is

$$-a_1\sigma_m^{-(s-1)} m\rho^{m-1} a_0^{m-1} = -a_1\sigma_m^{-(s-1)}.$$

Summing all the contributions we get

$$\tilde{r}^{[s]}(z) = -a_1\sigma_m^{-(s-1)} Z^{-s+1} \sum_{r=0}^{s} H_{s,r} (\ln Z^{-1})^{s-r} + O(|Z|^{-s+\frac{3}{2}-\epsilon}),$$

where the $(H_{s,r})$ do not depend on m. Recalling (2.5) and (3.10), and the definition of σ_m once again, completes the induction. □

Remark 4.4. The significance of Proposition 4.3 is that the case $m = 2$ has already been considered in [8], allowing us to determine the desired limiting distribution (Theorem 4.6) without computing the constants $(C_{s,r})$ in Proposition 4.3.

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

Theorem 4.5. *Let $\alpha \neq 1/2$, and let X_n denote the additive functional that satisfies the distributional recurrence* (2.1) *with $b_n \equiv n^\alpha$. Define $\alpha' := \alpha + \frac{1}{2}$.*

(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

$$\frac{(m-1)(m\alpha^*)^{1/2}}{n^{\alpha'}} \left[X_n - \frac{\rho m^{\frac{m}{m-1}} C_\alpha}{(m-1)\alpha^*}(n+1) \right] \xrightarrow{\mathcal{L}} Y_\alpha,$$

with C_α defined at (4.5) *and α^* at* (3.8).

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 $s \geq 2$,

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

Proof. If $\alpha > 1/2$, then by Proposition 4.1, 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 (4.13), 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 [8].

The same proof holds if $\alpha < 1/2$, now by considering $\widetilde{X}_n$ defined at (4.14) and using Proposition 4.2. $\square$

The case $\alpha = 1/2$ is covered by the following result. As alluded to in Remark 4.4, we will use the known results for $m = 2$ to derive the distribution for all m.

Theorem 4.6. *Let X_n denote the additive functional that satisfies the distributional recurrence* (2.1) *with $b_n \equiv n^{1/2}$. Define*

$$d_0(m) := \frac{\rho m^{\frac{m}{m-1}} C'_{1/2}}{(m-1)\alpha^*},$$

where $C'_{1/2}$ is defined at (4.10) *using* (4.9)*. Then, as $n \to \infty$,*

$$n^{-1}\left\{\sigma_m X_n - \frac{1}{\sqrt{2\pi}}\left(n \ln n + \left[(2\ln 2 + \gamma) + \sqrt{2\pi}\sigma_m d_0(m)\right] n\right)\right\} \xrightarrow{\mathcal{L}} Y_{1/2},$$

with convergence of all moments, where $Y_{1/2}$ has the unique distribution with moments $m_k := \mathbf{E}\, Y_{1/2}^k$ given by $m_0 = 1$, $m_1 = 0$, and for $k \geq 2$,

$$m_k = \frac{1}{4\sqrt{\pi}} \frac{\Gamma(k-1)}{\Gamma(k-\frac{1}{2})} \times \left[\sum_{\substack{k_1+k_2+k_3=k \\ k_1,k_2<k}} \binom{k}{k_1, k_2, k_3} m_{k_1} m_{k_2} \left(\frac{1}{\sqrt{2\pi}}\right)^{k_3} J_{k_1,k_2,k_3} + 4\sqrt{\frac{\pi}{2}} k m_{k-1}\right].$$

Here

$$J_{k_1,k_2,k_3} := \int_0^1 x^{k_1 - \frac{3}{2}} (1-x)^{k_2 - \frac{3}{2}} \left[x \ln x + (1-x)\ln(1-x)\right]^{k_3} \, dx.$$

Proof. Recall the definition of $\widetilde{X}_n \equiv \widetilde{X}_n(m)$ at (4.15). Using Proposition 4.3, singularity analysis, and (3.4), we see that

$$\tilde{\mu}_n^{[s]} = \sigma_m^{-s} n^s \sum_{r=0}^{s} \widehat{C}_{s,r} (\log n)^{s-r} + O(n^{s-\frac{1}{2}+\epsilon}),$$

where the $(\widehat{C}_{s,r})$ do not depend on m. Thus, for all (integer) $s \geq 0$,

$$\mathbf{E}\left[\sigma_m \widetilde{X}_n(m)\right]^s = \mathbf{E}\left[\sigma_2 \widetilde{X}_n(2)\right]^s + O(n^{s-\frac{1}{2}+\epsilon}).$$

It follows that

$$\begin{aligned}
&\mathbf{E}\left\{\sigma_m X_n - \frac{1}{\sqrt{2\pi}}\left(n\ln n + \left[(2\ln 2+\gamma)+\sqrt{2\pi}\sigma_m d_0(m)\right]n\right)\right\}^s \\
&= \mathbf{E}\left\{\sigma_m \widetilde{X}_n(m) - \frac{1}{\sqrt{2\pi}}[n\ln n + (2\ln 2+\gamma)n]\right\}^s \\
&= \mathbf{E}\left\{\sigma_2 \widetilde{X}_n(2) - \frac{1}{\sqrt{2\pi}}[n\ln n + (2\ln 2+\gamma)n]\right\}^s + O(n^{s-\frac{1}{2}+\epsilon}) \\
&= 2^{-s/2}\mathbf{E}\left\{X_n(2) - \frac{1}{\sqrt{\pi}}n\ln n - D_1 n\right\}^s + O(n^{s-\frac{1}{2}+\epsilon}),
\end{aligned}$$

the last equality using $\sigma_2 = 2^{-1/2}$ and

$$D_1 := \frac{1}{\sqrt{\pi}}[2\ln 2 + \gamma + \sqrt{\pi}d_0(2)].$$

Observe that

$$-\delta_n := \left(\frac{1}{\sqrt{\pi}}n\ln n + D_1 n\right) - \left[\frac{1}{\sqrt{\pi}}(n+1)\ln(n+1) + D_1(n+1)\right] = O(\log n).$$

Hence, for $l \geq 0$,

$$\begin{aligned}
&\mathbf{E}\left[X_n(2) - \frac{1}{\sqrt{\pi}}n\ln n - D_1 n\right]^s \\
&= \mathbf{E}\left[X_n(2) - \frac{1}{\sqrt{\pi}}(n+1)\ln(n+1) - D_1(n+1) + \delta_n\right]^s \\
&= \sum_{k=0}^{s}\binom{s}{r}[m_k n^k + o(n^k)]O((\log n)^{s-k}) \\
&= [m_s + o(1)]n^s,
\end{aligned}$$

where

$$m_k := \lim_{n\to\infty}\mathbf{E}\left\{n^{-1}\left[X_n(2) - \frac{1}{\sqrt{\pi}}(n+1)\ln(n+1) - D_1(n+1)\right]\right\}^k.$$

But the m_k's have already been determined [8, Proposition 3.8] and the claim follows from there. □

5. The shape functional

Recall that the shape functional X_n is the additive functional on m-ary search trees induced by the toll

$$b_n = \ln\binom{n}{m-1},$$

with $(X_0, \ldots, X_{m-2}) = \mathbf{0}$. By (2.5), we have

$$\mu^{[s]}(z) = \frac{r^{[s]}(z)}{1 - m[z\tau(z)]^{m-1}}, \tag{5.1}$$

with $r^{[s]}(z)$ given by (2.6), where

$$b(z) = \sum_{n=m-1}^{\infty} \ln\binom{n}{m-1} z^n.$$

It follows from Theorem 1 of [10] that $b(z)$ is amenable to singularity analysis. In the sequel we will make use of the following asymptotic expansion of b_n as $n \to \infty$:

$$b_n \sim (m-1)\ln n - \ln[(m-1)!] + n^{-1}\mathcal{N}. \tag{5.2}$$

5.1. **Mean.** Using (3.14) and (5.2) we have

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

A compatible singular expansion can be computed using (3.15):

$$\begin{aligned} r^{[1]}(z) &\sim C_{\ln} - (-a_1)(m-1)Z^{1/2}\ln Z^{-1} \\ &\quad - \{(-a_1)(m-1)[2(1-\ln 2)-\gamma] - (-a_1)\ln[(m-1)!]\}\, Z^{1/2} \\ &\quad + Z\mathcal{A} + (Z^{3/2}\log Z)\mathcal{A} + Z^{3/2}\mathcal{A}, \end{aligned}$$

where

$$C_{\ln} := r^{[1]}(\rho) = \sum_{n=m-1}^{\infty}\rho^n\left[\ln\binom{n}{m-1}\right]\tau_n.$$

Now using (5.1) and (3.10) we get

$$\begin{aligned} \mu^{[1]}(z) &\sim \frac{a_0 C_{\ln}}{(-a_1)(m-1)}Z^{-1/2} \\ &\quad - a_0\ln Z^{-1} - a_0\left[(2(1-\ln 2)-\gamma) - \frac{\ln[(m-1)!]}{m-1}\right] + \frac{(m-2)C_{\ln}}{3(m-1)} \\ &\quad + (Z^{1/2}\log Z)\mathcal{A} + Z^{1/2}\mathcal{A} + Z\mathcal{A} + (Z\log Z)\mathcal{A}, \end{aligned} \tag{5.3}$$

whence singularity analysis and (3.4) yield

$$\mu_n^{[1]} \sim \frac{2a_0C_{\ln}}{(m-1)a_1^2}n - 2\sqrt{\pi}\left(\frac{a_0}{-a_1}\right)n^{1/2} + (\log n)\mathcal{N} + \mathcal{N} + n^{-1/2}\mathcal{N}.$$

5.2. **Second moment and variance.** As in the case of the toll n^α when $\alpha < 1/2$, it will be convenient to consider the random variable

$$\widetilde{X}_n := X_n - d_1(n+1),$$

where here

$$d_1 := \frac{2a_0C_{\ln}}{(m-1)a_1^2}.$$

Thus, $\widetilde{X}_n$ satisfies the same distributional recurrence (2.1) as X_n with initial conditions $(\widetilde{X}_0,\dots,\widetilde{X}_{m-2}) = -d_1(1,\dots,m-1)$. Again, we use $\tilde{r}^{[s]}(z)$, $\tilde{\mu}^{[s]}(z)$, and $\tilde{\mu}_n^{[s]}$ to denote the analogous quantities for $\widetilde{X}_n$. Then, noting that $\mathbf{E}\,\widetilde{X}_n = \mathbf{E}\,X_n - d_1(n+1)$, a singular expansion for $\tilde{\mu}^{[1]}(z)$ can be obtained using (5.3), namely,

$$\tilde{\mu}^{[1]}(z) = -a_0\ln Z^{-1} - d_2 + O(|Z|^{\frac{1}{2}-\epsilon}), \tag{5.4}$$

where

$$d_2 := a_0\left[(2(1-\ln 2)-\gamma) - \frac{\ln[(m-1)!]}{m-1}\right] - \frac{(m-2)C_{\ln}}{3(m-1)} + (a_0-a_2)d_1$$

We begin the variance computation by obtaining asymptotics for the second moment $\tilde{\mu}_n^{[2]}$. To that end, we calculate the contribution of the terms in the sum

$$\tilde{r}^{[2]}(z) = \sum_{\substack{s_0+\cdots+s_m=2 \\ s_1,\ldots,s_m<2}} \binom{2}{s_0,\ldots,s_m} b^{\odot s_0}(z) \odot \left[z^{m-1}\tilde{\mu}^{[s_1]}(z)\cdots\tilde{\mu}^{[s_m]}(z)\right].$$

When $s_0 = 0$, exactly two of $s_1,\ldots,s_m$ equal 1. The contribution to $\tilde{r}^{[2]}(z)$ from such terms is

$$\begin{aligned} &2\binom{m}{2} z^{m-1}\tau^{m-2}(z)\left[\tilde{\mu}^{[1]}(z)\right]^2 \\ &\qquad = m(m-1)\rho^{m-1}a_0^{m-2}\left[a_0^2\ln^2 Z^{-1} + 2a_0d_2\ln Z^{-1} + d_2^2\right] + O(|Z|^{\frac{1}{2}-2\epsilon}). \end{aligned}$$

When $s_0 = 1$, exactly one of $s_1,\ldots,s_m$ equals 1. The contribution to $\tilde{r}^{[2]}(z)$ from such terms is obtained (after some routine calculations using the Zigzag algorithm) using the expansion for $\mathrm{Li}_{1,1}$ at (3.15) as

$$\begin{aligned} &2mb(z)\odot\left[z^{m-1}\tau^{m-1}(z)\tilde{\mu}^{[1]}(z)\right] \\ &= 2m\rho^{m-1}a_0^m\left[-\frac{m-1}{2}\ln^2 Z^{-1} + (\gamma(m-1)+\ln[(m-1)!])\ln Z^{-1}\right] \\ &\qquad + \text{constant} + O(|Z|^{\frac{1}{2}-\epsilon}). \end{aligned}$$

Finally, when $s_0 = 2$ the contribution to $r^{[1]}(z)$ is $b^{\odot 2}(z)\odot\left[z^{m-1}\tau^m(z)\right]$, which equals constant $+ O(|Z|^{\frac{1}{2}-\epsilon})$.

Summing these contributions and using (5.4) and (3.5), we conclude (note the cancellation of the ostensible lead term) that

$$\tilde{r}^{[2]}(z) = 4a_0(m-1)(1-\ln 2)\ln Z^{-1} + d_3 + O(|Z|^{\frac{1}{2}-\epsilon}),$$

where

$$d_3 := \lim_{z\to\rho}\left[r^{[2]}(z) - 4a_0(m-1)(1-\ln 2)\ln Z^{-1}\right].$$

This leads, using (3.10), to

$$\tilde{\mu}^{[2]}(z) = \frac{4a_0^2}{-a_1}(1-\ln 2)Z^{-1/2}\ln Z^{-1} + \frac{d_3a_0}{(-a_1)(m-1)}Z^{-1/2} + O(|Z|^{-\epsilon}). \tag{5.5}$$

By singularity analysis

$$\rho^n\tilde{\mu}_n^{[2]}\tau_n = \frac{4a_0^2}{-\sqrt{\pi}a_1}(1-\ln 2)n^{-1/2}\ln n + d_4n^{-1/2} + O(n^{-1+\epsilon}),$$

where

$$d_4 := \frac{4a_0^2}{-\sqrt{\pi}a_1}(1-\ln 2)(\gamma + 2\ln 2) + \frac{d_3a_0}{\sqrt{\pi}(-a_1)(m-1)}.$$

Using the asymptotics of τ_n at (3.4) we get

$$\tilde{\mu}_n^{[2]} = 8\left(\frac{a_0}{a_1}\right)^2(1-\ln 2)n\ln n + \frac{2\sqrt{\pi}d_4}{-a_1}n + O(n^{\frac{1}{2}+\epsilon}).$$

Thus

$$\begin{aligned}\mathbf{Var}\, X_n = \mathbf{Var}\, \widetilde{X}_n &= \tilde{\mu}_n^{[2]} - \left(\tilde{\mu}_n^{[1]}\right)^2 \\ &= 8\left(\frac{a_0}{a_1}\right)^2 (1-\ln 2) n \ln n + \left(\frac{2\sqrt{\pi} d_4}{-a_1} - \frac{4\pi a_0^2}{a_1^2}\right) n + O(n^{\frac{1}{2}+\epsilon}).\end{aligned}$$

5.3. **Higher moments and limiting distribution.**

Proposition 5.1. *For $s \geq 2$ and $\epsilon > 0$ small enough,*

$$\tilde{\mu}^{[s]}(z) = Z^{-\frac{s}{2}+\frac{1}{2}} \sum_{j=0}^{\lfloor s/2 \rfloor} C_{s,j} \left(\ln^{\lfloor s/2 \rfloor - j} Z^{-1}\right) + O(|Z|^{-\frac{s}{2}+1-\epsilon}),$$

with

$$C_{2l,0} = \frac{1}{-2a_1} \sum_{j=1}^{l-1} \binom{2l}{2j} C_{2j,0} C_{2l-2j,0}, \quad l \geq 2; \qquad C_{2,0} = \frac{4a_0^2}{-a_1}(1-\ln 2). \tag{5.6}$$

Proof sketch. We proceed by induction. For $s = 2$ the claim is true by (5.5). [Recalling (5.4), we note in passing that the claim is *not* true for $s = 1$.] In the rest of the proof we will explicitly compute $C_{s,j}$ only when s is even and $j = 0$. The rest of the coefficients will appear as unspecified constants.

For the induction step consider $s \geq 3$. In a manner analogous to the proof of Proposition 4.1 we obtain the asymptotics of $\tilde{r}^{[s]}(z)$ by first analyzing the contributions of the terms in the sum (2.6). As in that proof one can check that unless

(a) $s_0 = 0$ and exactly two of $s_1, \ldots, s_m$ are nonzero, or
(b) $s_0 = 1$ and exactly one of $s_1, \ldots, s_m$ is $s-1$

in (2.6), the contribution to $\tilde{r}^{[s]}(z)$ is $O(|Z|^{-\frac{s}{2}+\frac{3}{2}-(s+1)\epsilon})$.

In case (a), the contribution to $\tilde{r}^{[s]}(z)$ is

$$\binom{m}{2} z^{m-1} \tau^{m-2}(z) \sum_{j=1}^{s-1} \binom{s}{j} \tilde{\mu}^{[j]}(z) \tilde{\mu}^{[s-j]}(z). \tag{5.7}$$

In this sum unless j is 1 or $s-1$, the induction hypothesis implies a contribution of

$$\binom{m}{2} \rho^{m-1} a_0^{m-2} Z^{-\frac{s}{2}+1} \sum_{l=0}^{\lfloor j/2 \rfloor + \lfloor \frac{s-j}{2} \rfloor} \binom{s}{j} A_{s,j,l} \left(\ln^{\lfloor j/2 \rfloor + \lfloor \frac{s-j}{2} \rfloor - l} Z^{-1}\right) + O(|Z|^{-\frac{s}{2}+\frac{3}{2}-2\epsilon}), \tag{5.8}$$

with $A_{s,j,0} = C_{j,0} C_{s-j,0}$. Notice that when s is even and j is odd, this contribution is $O\left(|Z|^{-\frac{s}{2}+1} \left(\ln^{\lfloor s/2 \rfloor - 1} Z^{-1}\right)\right)$. In all other parity cases the contribution is $O\left(|Z|^{-\frac{s}{2}+1} \ln^{\lfloor s/2 \rfloor} Z^{-1}\right)$.

On the other hand the total contribution to $\tilde{r}^{[s]}(z)$ from the terms where j is 1 or $s-1$ in the sum in (5.7) is obtained using the induction hypothesis and (5.4) as

$$-2s \binom{m}{2} \rho^{m-1} a_0^{m-1} Z^{-\frac{s}{2}+1} \sum_{j=0}^{\lfloor \frac{s-1}{2} \rfloor} E_{s-1,j} \ln^{\lfloor \frac{s-1}{2} \rfloor + 1 - j} Z^{-1} + O(|Z|^{-\frac{s}{2}+\frac{3}{2}-2\epsilon}), \tag{5.9}$$

where $E_{s-1,0} = C_{s-1,0}$.

In case (b), the total contribution to $\tilde{r}^{[s]}(z)$ is

$$msb(z) \odot [z^{m-1}\tau^{m-1}(z)\tilde{\mu}^{[s-1]}(z)].$$

Now another application of the Zigzag algorithm yields this contribution to be

$$sm(m-1)\rho^{m-1}a_0^{m-1}Z^{-\frac{s}{2}+1}\sum_{j=0}^{\lfloor\frac{s-1}{2}\rfloor} D_{s,j}\ln^{\lfloor\frac{s-1}{2}\rfloor+1-j} Z^{-1} + O(|Z|^{-\frac{s}{2}+\frac{3}{2}-2\epsilon}), \tag{5.10}$$

where $D_{s,0} = C_{s-1,0}$. Note that the lead term here is exactly the same as in (5.9) but with opposite sign, so that (5.9) and (5.10) cancel each other to lead order.

Summing the various contributions, we find that $\tilde{r}^{[s]}(z)$ is of the form

$$\tilde{r}^{[s]}(z) = Z^{-\frac{s}{2}+1}\sum_{j=0}^{\lfloor s/2\rfloor}\widehat{C}_{s,j}\left(\ln^{\lfloor s/2\rfloor-j} Z^{-1}\right) + O\left(|Z|^{-\frac{s}{2}+\frac{3}{2}-2\epsilon}\right),$$

where, for s even,

$$\widehat{C}_{s,0} = \binom{m}{2}\rho^{m-1}a_0^{m-2}\sum_{\substack{0<j<s \\ j \text{ even}}}\binom{s}{j}C_{j,0}C_{s-j,0}.$$

Using (3.10), the result follows. $\square$

It follows from Proposition 5.1, singularity analysis, and the asymptotics of τ_n at (3.4) that

$$\tilde{\mu}_n^{[s]} = \frac{2\sqrt{\pi}C_{s,0}}{-a_1\Gamma(\frac{s-1}{2})}n^{s/2}\ln^{\lfloor s/2\rfloor} n + O(n^{s/2}\ln^{\lfloor s/2\rfloor-1} n),$$

whence for $s \geq 1$, as $n \to \infty$,

$$\mathbf{E}\left[\frac{\widetilde{X}_n}{\sqrt{n\ln n}}\right]^{2s} \to \frac{2\sqrt{\pi}C_{2s,0}}{(-a_1)\Gamma(s-\frac{1}{2})} \quad \text{and} \quad \mathbf{E}\left[\frac{\widetilde{X}_n}{\sqrt{n\ln n}}\right]^{2s-1} = o(1).$$

Solving the recurrence (5.6) for $C_{2s,0}$ yields

$$C_{2s,0} = \frac{-a_1\Gamma(\frac{s-1}{2})}{2\sqrt{\pi}}\frac{(2s)!}{2^s s!}\sigma^{2s} = \frac{-a_1}{2}\frac{(2s)!(2s-2)!}{2^s 2^{2s-2}s!(s-1)!}\sigma^{2s},$$

where $\sigma^2 := 8\,(a_0/a_1)^2\,(1-\ln 2)$. The method of moments (see, for example, [1, Theorem 30.1]) implies then that the shape functional is asymptotically normal.

Theorem 5.2. *Let X_n denote the shape functional for uniformly distributed m-ary search trees on n keys. Then*

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

where

$$d_1 := \frac{2a_0}{(m-1)a_1^2}\sum_{n=m-1}^{\infty}\rho^n\left[\ln\binom{n}{m-1}\right]\tau_n$$

and $\sigma^2 := 8\,(a_0/a_1)^2\,(1-\ln 2)$.

Remark 5.3. It is known [2, 9] that under the random permutation model the shape functional centered by its mean and scaled by its 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 $m \ge 2$.

6. The space requirement

The space requirement for m-ary search trees is the number of nodes in the tree [17]. (For a binary search tree the space requirement for n keys is clearly n, so in this section we assume $m \ge 3$.) The limiting distribution of this parameter under the random permutation model has been considered by several authors [18, 16, 2, 9]. In our framework it is the additive functional X_n corresponding to the toll $\mathbf{1}_{n \ge m-1}$ with initial conditions $(X_0, \dots, X_{m-2}) = (0, 1, \dots, 1)$.

The sth moment $\mu^{[s]}(z) := \mathbf{E}\, X_n^s$ can be computed as usual using (2.5), where now

$$r^{[s]}(z) = z^{m-1} \sum_{\substack{s_0+\cdots+s_m=s \\ s_1,\dots,s_m<s}} \binom{s}{s_0, \dots, s_m} \mu^{[s_1]}(z) \cdots \mu^{[s_m]}(z) \tag{6.1}$$

since the toll generating function $b(z) = (1-z)^{-1}$ serves as the identity for Hadamard products.

6.1. Mean.

Substituting $s = 1$ in (6.1) and using (3.1) and (3.3) yields

$$r^{[1]}(z) + \sum_{j=1}^{m-2} z^j = z^{m-1}\tau^m(z) + \sum_{j=1}^{m-2} z^j = \tau(z) - 1 \sim (a_0 - 1) + a_1 Z^{1/2} + Z\mathcal{A} + Z^{3/2}\mathcal{A}.$$

Then, by (2.5) and (3.10),

$$\mu^{[1]}(z) \sim \frac{a_0(a_0-1)}{-a_1(m-1)} Z^{-1/2} + \left[c_0(a_0-1) - \frac{a_0}{m-1}\right] + Z^{1/2}\mathcal{A} + Z\mathcal{A}. \tag{6.2}$$

Singularity analysis and the asymptotics of τ_n immediately lead to

$$\mu_n^{[1]} \sim \frac{2a_0(a_0-1)}{a_1^2(m-1)} n + \mathcal{N}. \tag{6.3}$$

6.2. Variance.

As for the shape functional it is convenient to consider instead the "centered" functional $\widetilde{X}_n := X_n - d_1(n+1)$, where now

$$d_1 := \frac{2a_0(a_0-1)}{a_1^2(m-1)} = \frac{m(1 - \rho m^{\frac{1}{m-1}})}{(m-1)\alpha^*}.$$

The "centered" space requirement satisfies the same distributional recurrence as the space requirement with initial conditions

$$(\widetilde{X}_0, \dots, \widetilde{X}_{m-2}) = -(d_1, 2d_1 - 1, \dots, (m-1)d_1 - 1).$$

We employ the same notation as Section 5, with $\tilde{r}^{[s]}(z)$, $\tilde{\mu}^{[s]}(z)$, and $\tilde{\mu}_n^{[s]}$ denoting quantities analogous to $r^{[s]}(z)$, $\mu^{[s]}(z)$, and $\mu_n^{[s]}$.

By definition

$$\tilde{\mu}^{[1]}(z) = \mu^{[1]}(z) - d_1 \sum_{n=0}^{\infty} (n+1)\tau_n z^n. \tag{6.4}$$

Now

$$\sum_{n=0}^{\infty}(n+1)\tau_n z^n = z\tau'(z) + \tau(z) \sim \frac{-a_1}{2}Z^{-1/2} + (a_0 - a_2) + Z^{1/2}\mathcal{A} + Z\mathcal{A}. \tag{6.5}$$

Thus using (6.2), (6.4), and (6.5) we find

$$\tilde{\mu}^{[1]}(z) \sim B_1 + Z^{1/2}\mathcal{A} + Z\mathcal{A}, \tag{6.6}$$

where, using (3.9), we have

$$B_1 := c_0(a_0 - 1) - \frac{a_0}{m-1} - d_1(a_0 - a_2) = -\frac{a_0}{m-1}. \tag{6.7}$$

Using (6.1),

$$\tilde{r}^{[2]}(z) = z^{m-1}\left[m(m-1)\left(\tilde{\mu}^{[1]}(z)\right)^2 \tau^{m-2}(z) + 2m\tilde{\mu}^{[1]}(z)\tau^{m-1}(z) + \tau^m(z)\right]. \tag{6.8}$$

Also,

$$\sum_{j=0}^{m-2}\tilde{x}_j^2\rho^j = d_1^2 + \sum_{j=1}^{m-2}[-(j+1)d_1 + 1]^2\rho^j =: \delta_1. \tag{6.9}$$

By (6.6)

$$\left(\tilde{\mu}^{[1]}(z)\right)^2 \sim B_1^2 + Z^{1/2}\mathcal{A} + Z\mathcal{A},$$

and, for $k \ge 1$,

$$\tau^k(z) \sim a_0^k + k a_0^{k-1}a_1 Z^{1/2} + Z\mathcal{A} + Z^{3/2}\mathcal{A},$$

so that

$$\left(\tilde{\mu}^{[1]}(z)\right)^2 \tau^{m-2}(z) \sim a_0^{m-2}B_1^2 + Z^{1/2}\mathcal{A} + Z\mathcal{A}.$$

Similarly

$$\tilde{\mu}^{[1]}(z)\tau^{m-1}(z) \sim a_0^{m-1}B_1 + Z^{1/2}\mathcal{A} + Z\mathcal{A}.$$

Using these expansions and (3.5) in (6.8) gives [recalling (3.13) and (6.9)]

$$\tilde{r}^{[2]}(z) + \sum_{j=0}^{m-2}\tilde{x}_j^2 z^j \sim \left[\frac{m-1}{a_0}B_1^2 + 2B_1 + \frac{a_0}{m} + \delta_1\right] + Z^{1/2}\mathcal{A} + Z\mathcal{A},$$

whence

$$\tilde{\mu}^{[2]}(z) \sim B_2 Z^{-1/2} + \mathcal{A} + Z^{1/2}\mathcal{A}, \tag{6.10}$$

where

$$B_2 := \frac{a_0}{-a_1(m-1)}\left[\frac{m-1}{a_0}B_1^2 + 2B_1 + \frac{a_0}{m} + \delta_1\right] = \frac{a_0}{-a_1(m-1)}\left[\delta_1 - \frac{a_0}{m(m-1)}\right]. \tag{6.11}$$

By singularity analysis and the asymptotics of τ_n, then

$$\tilde{\mu}_n^{[2]} \sim \frac{2B_2}{-a_1}n + \mathcal{N}.$$

Recalling (6.3), we observe that $\tilde{\mu}_n^{[1]} \sim \mathcal{N}$ so that

$$\mathbf{Var}\,X_n = \mathbf{Var}\,\widetilde{X}_n = \tilde{\mu}_n^{[2]} - \left(\tilde{\mu}_n^{[1]}\right)^2 \sim \frac{2B_2}{-a_1}n + \mathcal{N}. \tag{6.12}$$

We pause here to remark that since $\mathbf{Var}\, X_n \to \infty$ as $n \to \infty$ for $m \geq 3$ (see Remark A.2 in the Appendix), we must have $B_2 > 0$. We do not know a direct proof of this fact.

6.3. **Higher moments and limiting distribution.**

Proposition 6.1. *For $s \geq 1$,*

$$\tilde{\mu}^{[s]}(z) = B_s Z^{-\frac{s}{2}+\frac{1}{2}} + O(|Z|^{-\frac{s}{2}+1}),$$

where B_1 and B_2 are given at (6.7) *and* (6.11)*, respectively, and, for $s \geq 3$,*

$$B_s = \frac{a_0}{-a_1(m-1)} \left[\frac{m-1}{2a_0} \sum_{j=1}^{s-1} \binom{s}{j} B_j B_{s-j} + s B_{s-1} \right]. \tag{6.13}$$

Proof Sketch. We proceed by induction. For $s = 1$ and $s = 2$, the result has been established at (6.6) and (6.10), respectively.

Suppose $s \geq 3$. We will first obtain a singular expansion for $\tilde{r}^{[s]}(z)$ by analyzing the contributions of the terms in the sum at (6.1). As in the proofs of Propositions 4.1 and 5.1 unless

(a) $s_0 = 0$ and exactly two of $s_1, \dots, s_m$ are nonzero, or
(b) $s_0 = 1$ and exactly one of $s_1, \dots, s_m$ is $s - 1$,

the contribution to $\tilde{r}^{[s]}(z)$ is $O(|Z|^{-\frac{s}{2}+\frac{3}{2}})$.

In case (a), the contribution to $\tilde{r}^{[s]}(z)$ is

$$\binom{m}{2} \rho^{m-1} a_0^{m-2} Z^{-\frac{s}{2}+1} \sum_{j=1}^{s-1} \binom{s}{j} B_j B_{s-j} + O(|Z|^{-\frac{s}{2}+\frac{3}{2}}).$$

In case (b), the contribution to $\tilde{r}^{[s]}(z)$ is

$$m \rho^{m-1} s a_0^{m-1} B_{s-1} Z^{-\frac{s}{2}+1} + O(|Z|^{-\frac{s}{2}+\frac{3}{2}})$$

This leads to

$$\tilde{r}^{[s]}(z) + \sum_{j=0}^{m-2} \tilde{x}_j^s z^j = \left(\frac{m-1}{2a_0} \sum_{j=1}^{s-1} \binom{s}{j} B_j B_{s-j} + s B_{s-1} \right) Z^{-\frac{s}{2}+1} + O(|Z|^{-\frac{s}{2}+\frac{3}{2}}).$$

whence (2.5) and (3.10) complete the induction. $\square$

Now using (6.13) and (6.7) we have $B_3 = 0$ and by induction $B_{2s+1} = 0$ for $s = 1, 2, \dots$. Then [compare (5.6)]

$$B_{2l} = \frac{1}{-2a_1} \sum_{j=1}^{l-1} \binom{2l}{2j} B_{2j} B_{2l-2j}, \quad l \geq 2,$$

with B_2 given at (6.11). Now following the development leading to Theorem 5.2 we can conclude asymptotic normality for the space requirement.

Theorem 6.2. *Let X_n denote the space requirement for uniformly distributed m-ary search trees on n keys. Then*

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

where

$$d_1 = \frac{m(1-\rho m^{\frac{1}{m-1}})}{(m-1)\alpha^*} \quad \text{and} \quad \sigma^2 = 2\frac{B_2}{-a_1}.$$

Here B_2 is given by (6.11)*, with δ_1 defined at* (6.9) *and a_0 and a_1 at* (3.5) *and* (3.7)*, respectively.*

7. Number of leaves

Lastly we consider the number of leaves in an m-ary search tree. This is the additive functional X_n corresponding to the toll $\mathbf{1}_{n=m-1}$ with initial conditions $(X_0, \ldots, X_{m-2}) = (0, 1, \ldots, 1)$. Under the random permutation model, the number of leaves is asymptotically normal [9, Theorem 2.5]. We will establish an analogous result (Theorem 7.2) under the uniform model.

The toll generating function is now $b(z) = z^{m-1}$. Note that

$$b^{\odot s}(z) = \begin{cases} (1-z)^{-1} & s = 0, \\ z^{m-1} & s \geq 1. \end{cases}$$

Thus in (2.6)

$$r^{[s]}(z) = z^{m-1}\left[\sum_{\substack{s_1+\cdots+s_m=s \\ s_1,\ldots,s_m<s}} \binom{s}{s_1,\ldots,s_m}\mu^{[s_1]}(z)\cdots\mu^{[s_m]}(z) + 1\right]. \tag{7.1}$$

Given the similarity of the calculations with those of Section 6, we will be brief. The interested reader is invited to flesh out the development of this section along the lines of Section 6.

7.1. Mean.

Substituting $s = 1$ in (7.1) and using (2.5) we get

$$\mu^{[1]}(z) \sim \frac{a_0}{-a_1 m^{\frac{m}{m-1}}} Z^{-1/2} + \frac{m-2}{3m^{\frac{m}{m-1}}} + Z^{1/2}\mathcal{A} + Z\mathcal{A}$$

so that

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

7.2. Variance.

For the variance we consider again $\widetilde{X}_n := X_n - d_1(n+1)$, where now $d_1 := \frac{\rho}{\alpha^*}$. Then

$$\tilde{\mu}^{[1]}(z) \sim B_1 + Z^{1/2}\mathcal{A} + Z\mathcal{A},$$

where [compare (6.7)] $B_1 = 0$. Substituting $s = 2$ in (7.1) we get

$$\tilde{r}^{[2]}(z) = z^{m-1}\left[1 + m(m-1)\left(\tilde{\mu}^{[1]}(z)\right)^2 \tau^{m-2}(z)\right] \sim \rho^{m-1} + Z\mathcal{A} + Z^{3/2}\mathcal{A}.$$

Thus

$$\tilde{\mu}^{[2]}(z) \sim B_2 Z^{-1/2} + \mathcal{A} + Z^{1/2}\mathcal{A},$$

where

$$B_2 = \frac{a_0(\rho^{m-1} + \delta_1)}{-a_1(m-1)}, \tag{7.2}$$

with δ_1 defined as at (6.9) (but now with $d_1 = \rho/\alpha^*$). This leads to

$$\mathbf{Var}\, X_n \sim \frac{2B_2}{-a_1} n + \mathcal{N} \sim \frac{\rho m^{\frac{m}{m-1}}(\rho^{m-1} + \delta_1)}{\alpha^*(m-1)} n + \mathcal{N}.$$

Clearly $B_2 > 0$, and therefore $\mathbf{Var}\, X_n$ grows at an asymptotically linear rate.

7.3. **Higher moments and limiting distribution.** The analog of Proposition 6.1 is the following. We omit the proof.

Proposition 7.1. *For $s \geq 1$,*

$$\tilde{\mu}^{[a]}(z) = B_s Z^{-\frac{s}{2}+\frac{1}{2}} + O(|Z|^{-\frac{s}{2}+1}),$$

where $B_1 = 0$, B_2 is given at (7.2), and, for $s \geq 3$,

$$B_s = -\frac{1}{-2a_1} \sum_{j=1}^{s-1} \binom{s}{j} B_j B_{s-j}.$$

It follows easily from Proposition 7.1 that the number of leaves is asymptotically normal.

Theorem 7.2. *Let X_n denote the number of leaves in a uniformly distributed m-ary search tree on n keys. Then*

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

where

$$\sigma^2 = \frac{\rho m^{\frac{m}{m-1}} (\rho^{m-1} + \delta_1)}{\alpha^*(m-1)}.$$

Appendix A. Growth of variance for nondegenerate functionals

In Section 6.2 we claimed that the variance of the space requirement tends to infinity as $n \to \infty$ for $m \geq 3$. In this appendix we identify additive functionals that are degenerate, i.e., for each fixed n have the same value for all m-ary search trees with n keys, and provide a lower bound on the rate of growth of the variance of any nondegenerate additive functional.

Theorem A.1. *Consider an additive functional $(X_n)_{n \geq 0}$ [as at (2.1)] with toll $(b_n)_{n \geq 0}$, with initial conditions $(x_0, \ldots, x_{m-2}) = (b_0, \ldots, b_{m-2})$. Then the following statements are equivalent:*

(a) $\mathcal{L}(X_n)$ is degenerate for every $n \geq 0$.
(b) The toll satisfies

$$b_n = \begin{cases} n b_1 - (n-1) b_0, & n = 2, \ldots, m-2 \\ (m-1)(b_1 - 2b_0), & n \geq m-1. \end{cases}$$

(c) $X_n = n b_1 - (n-1) b_0$ for every $n \geq 0$.

Moreover, if (a) does not hold then $\sigma_n^2 := \mathbf{Var}\, X_n = \Omega(\log n)$ as $n \to \infty$.

Remark A.2. Before we prove the theorem we apply it to the space requirement. It is easily checked that, for this additive functional, condition (b) holds only when $m = 2$. Thus the space requirement is not degenerate (and its variance tends to infinity as $n \to \infty$) for $m \geq 3$.

Proof of Theorem A.1. We will show below that (b) and (c) are equivalent.

To show the equivalence with (a), suppose first that (a) holds, so that $X_n = x_n$ deterministically for all $n \geq 0$. Then

$$\text{(A.1)} \qquad \begin{aligned} x_{n+(m-1)} &= x_n + (m-1)x_0 + b_{n+(m-1)} && \text{for } n \geq 0 \\ &= x_{n-1} + x_1 + (m-2)x_0 + b_{n+(m-1)} && \text{for } n \geq 1, \end{aligned}$$

and so $x_n = x_{n-1} + (x_1 - x_0) = x_{n-1} + (b_1 - b_0)$ for $n \geq 1$. Condition (c) then follows by induction. Conversely, condition (c) trivially implies (a), and the equivalent (b) shows that X_n is indeed an additive functional.

We now show the equivalence of (b) and (c). If (c), holds then so does (A.1) [because $(X_n) = (x_n)$ is an additive functional], from which [by solving for $b_{n+(m-1)}$] it is easy to check that (b) holds. Conversely, if (b) holds, then (c) is trivially true for $n = 0, \dots, m-2$, and holds by induction for $n \geq m-1$.

Suppose now that (b_n) does not satisfy (b). Let n_0 be such that $\mathcal{L}(X_{n_0})$ is not degenerate. Then for any $n \geq n_1 := n_0 + (m-1)$, there is positive probability of having a subtree containing precisely n_0 keys, so that by the law of total variance we must $\sigma_n^2 > 0$.

Finally we show that if $\mathcal{L}(X_n)$ is nondegenerate for all $n \geq n_1$, then $\sigma_n^2 = \Omega(\log n)$ as $n \to \infty$. First, it is clear that in this case there exists $\epsilon > 0$ such that $\sigma_n^2 \geq \epsilon$ for all $n \in [n_1, mn_1 + (m-2)]$. Now suppose $n \geq mn_1 + (m-1)$. Then at least one subtree must have size in the range $[n_1, n-(m-1)] \subseteq [n_1, n-1]$, so that by induction and the law of total variance we have $\sigma_n^2 \geq \epsilon$ for all $n \geq n_1$.

For $n \geq n_1 + (m-1)$, let p_n denote the probability that a tree of size n has its first subtree of size $n - n_1 - (m-1)$, its second of size n_1, and the rest of size 0. Then by (2.2) and the asymptotics of τ_n (see Remark 3.1), as $n \to \infty$ we have

$$p_n = \frac{\tau_{n-n_1-(m-1)}\tau_{n_1}}{\tau_n} \to \tau_{n_1}\rho^{n_1+(m-1)} > 0.$$

Also $p_n > 0$ for $n \geq n_1 + (m-1)$, and so $\delta := \inf\{p_n : n \geq n_1 + (m-1)\} > 0$.

Define $\alpha_0 := n_1$, $\alpha_1 := mn_1 + (m-1)$, and, for $k \geq 2$, $\alpha_k := m\alpha_{k-1} + n_1 + (m-1)$. We will show, for each $k \geq 0$ and any $n \geq \alpha_k$, that $\sigma_n^2 \geq (1 + k\delta)\epsilon$, and the logarithmic-growth claim follows.

For $k = 0$ the result has been shown above. For $k = 1$, using the law of total variance and $p_n \geq \delta$ the result follows (compare the case $k \geq 2$ to follow). Suppose $k \geq 2$ and $n \geq \alpha_k$. Then $n-(m-1) \geq m\alpha_{k-1}$, so there must be at least one subtree of size at least α_{k-1}. Hence $\sigma_n^2 \geq (1 + (k-1)\delta)\epsilon$. But with probability $p_n \geq \delta$, the first two subtrees are each of size at least n_1; then by the law of total variance we have $\sigma_n^2 \geq [1 + (k-1)\delta]\epsilon + \delta\epsilon = (1 + k\delta)\epsilon$, as desired. □

References

[1] P. Billingsley. *Probability and measure*. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[2] 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).

[3] 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.

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

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

[6] 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.
[7] 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.
[8] 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.
[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.

(James Allen Fill) Department of Applied Mathematics and Statistics, The Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218-2682

URL: `http://www.ams.jhu.edu/~fill/`

E-mail address: `jimfill@jhu.edu`

(Nevin Kapur) Department of Computer Science, California Institute of Technology, MC 256-80, 1200 E. California Blvd., Pasadena CA 91125

URL: `http://www.cs.caltech.edu/~nkapur/`

E-mail address: `nkapur@cs.caltech.edu`