LIMITING DISTRIBUTIONS FOR ADDITIVE FUNCTIONALS ON CATALAN TREES

JAMES ALLEN FILL AND NEVIN KAPUR

Abstract. Additive tree functionals represent the cost of many divide-and-conquer algorithms. We derive the limiting distribution of the additive functionals induced by toll functions of the form (a) n^α when $\alpha > 0$ and (b) $\log n$ (the so-called shape functional) on uniformly distributed binary trees, sometimes called Catalan trees. The Gaussian law obtained in the latter case complements the central limit theorem for the shape functional under the random permutation model. Our results give rise to an apparently new family of distributions containing the Airy distribution ($\alpha = 1$) and the normal distribution [case (b), and case (a) as $\alpha \downarrow 0$]. The main theoretical tools employed are recent results relating asymptotics of the generating functions of sequences to those of their Hadamard product, and the method of moments.

1. Introduction

Binary trees are fundamental data structures in computer science, with primary application in searching and sorting. For background we refer the reader to Chapter 2 of the excellent book [18]. In this article we consider additive functionals defined on uniformly distributed binary trees (sometimes called Catalan trees) induced by two types of toll sequences [(n^α) and $(\log n)$]. (See the simple Definition 2.1.) Our main results, Theorems 3.10 and 4.2, establish the limiting distribution for these induced functionals.

A competing model of randomness for binary trees—one used for binary search trees—is the *random permutation model* (RPM); see Section 2.3 of [18]. While there has been much study of additive functionals under the RPM (see, for example, [18, Section 3.3] and [21, 5, 13, 3]), little attention has been paid to the distribution of functionals defined on binary trees under the uniform (Catalan) model of randomness. Fill [5] argued that the functional corresponding to the toll sequence $(\log n)$ serves as a crude measure of the "shape" of a binary tree, and explained how this functional arises in connection with the move-to-root self-organizing scheme for dynamic maintenance of binary search trees. He derived a central limit theorem under the RPM, but obtained only asymptotic information about the mean and variance

Date: June 4, 2003. Revised April 1, 2004.

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

Key words and phrases. Catalan trees, additive functionals, limiting distributions, divide-and-conquer, shape functional, Hadamard product of functions, singularity analysis, Airy distribution, generalized polylogarithm, method of moments, central limit theorem.

Research for both authors supported by NSF grants DMS-9803780 and DMS-0104167, and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics. Research for the second author supported by NSF grant 0049092 and carried out primarily while this author was affiliated with what is now the Department of Applied Mathematics and Statistics at The Johns Hopkins University.

under the Catalan model. (The latter results were rederived in the extension [19] from binary trees to simply generated rooted trees.) In this paper (Theorem 4.2) we show that there is again asymptotic normality under the Catalan model.

In [11, Prop. 2] Flajolet and Steyaert gave order-of-growth information about the mean of functionals induced by tolls of the form n^α. (The motivation is to build a "repertoire" of tolls from which the behavior of more complicated tolls can be deduced by combining elements from the repertoire. The corresponding results under the random permutation model were derived by Neininger [20].) Takács established the limiting (Airy) distribution of path length in Catalan trees [23, 24, 25], which is the additive functional for the toll $n-1$. The additive functional for the toll n^2 arises in the study of the Wiener index of the tree and has been analyzed by Janson [15]. In this paper (Theorem 3.10) we obtain the limiting distribution for Catalan trees for toll n^α for any $\alpha > 0$. The family of limiting distributions appears to be new. In most cases we have a description of the distribution only in terms of its moments, although other descriptions in terms of Brownian excursion, as for the Airy distribution and the limiting distribution for the Wiener index, may be possible. This is currently under investigation by the authors in collaboration with others.

The uniform model on binary trees has also been used recently by Janson [14] in the analysis of an algorithm of Koda and Ruskey [17] for listing ideals in a forest poset.

This paper serves as the first example of the application of recent results [6], extending singularity analysis [10], to obtain limiting distributions. In [6], it is shown how the asymptotics of generating functions of sequences relate to those of their Hadamard product. First moments for our problems were treated in [6] and a sketch of the technique we employ was presented there. (Our approach to obtaining asymptotics of Hadamard products of generating functions differs only marginally from the Zigzag Algorithm as presented in [6].) As will be evident soon, Hadamard products occur naturally when one is analyzing moments of additive tree functionals. The program we carry out allows a fairly mechanical derivation of the asymptotics of moments of each order, thereby facilitating application of the method of moments. Indeed, preliminary investigations suggest that the techniques we develop are likewise applicable to the wider class of simply generated trees; this is work in progress.

The organization of this paper is as follows. Section 2 establishes notation and states certain preliminaries that will be used in the subsequent proofs. In Section 3 we consider the toll sequence (n^α) for general $\alpha > 0$. In Section 3.1 we compute the asymptotics of the mean of the corresponding additive functional. In Section 3.2 the analysis diverges slightly as the nature of asymptotics of the higher moments differs depending on the value of α. Section 3.3 employs singularity analysis [10] to derive the asymptotics of moments of each order. In Section 3.4 we use the results of Section 3.3 and the method of moments to derive the limiting distribution of the additive tree functional. In Section 4 we employ the approach again to obtain a normal limit theorem for the shape functional. Finally, in Section 5, we present heuristic arguments that may lead to the identification of toll sequences giving rise to a normal limit.

2. Notation and Preliminaries

2.1. **Additive tree functionals.** We first establish some notation. Let T be a binary tree. We use $|T|$ to denote the number of nodes in T. Let $L(T)$ and $R(T)$ denote, respectively, the left and right subtrees rooted at the children of the root of T.

Definition 2.1. A functional f on binary trees is called an *additive tree functional* if it satisfies the recurrence

$$f(T) = f(L(T)) + f(R(T)) + b_{|T|},$$

for any tree T with $|T| \geq 1$. Here $(b_n)_{n\geq 1}$ is a given sequence, henceforth called the *toll function*.

We analyze additive functionals defined on binary trees uniformly distributed over $\{T\colon |T| = n\}$ for given n. Let X_n be such an additive functional induced by the toll sequence (b_n). It is well known that the number of binary trees on n nodes is counted by the nth Catalan number

$$\beta_n := \frac{1}{n+1}\binom{2n}{n},$$

with generating function

$$\mathrm{CAT}(z) := \sum_{n=0}^{\infty} \beta_n z^n = \frac{1}{2z}(1 - \sqrt{1-4z}).$$

In our subsequent analysis we will make use of the identity

$$z\,\mathrm{CAT}^2(z) = \mathrm{CAT}(z) - 1. \tag{2.1}$$

The mean of the cost function $a_n := \mathbf{E}\,X_n$ can be obtained recursively by conditioning on the size of $L(T)$ as

$$a_n = \sum_{j=1}^{n} \frac{\beta_{j-1}\beta_{n-j}}{\beta_n}(a_{j-1} + a_{n-j}) + b_n, \qquad n \geq 1.$$

This recurrence can be rewritten as

$$(\beta_n a_n) = 2\sum_{j=1}^{n}(\beta_{j-1}a_{j-1})\beta_{n-j} + (\beta_n b_n), \qquad n \geq 1. \tag{2.2}$$

Recall that the *Hadamard product* of two power series F and G, denoted by $F(z) \odot G(z)$, is the power series defined by

$$(F \odot G)(z) \equiv F(z) \odot G(z) := \sum_n f_n g_n z^n,$$

where

$$F(z) = \sum_n f_n z^n \qquad \text{and} \qquad G(z) = \sum_n g_n z^n.$$

Multiplying (2.2) by $z^n/4^n$ and summing over $n \geq 1$ we get

$$A(z) \odot \mathrm{CAT}(z/4) = \frac{B(z) \odot \mathrm{CAT}(z/4)}{\sqrt{1-z}}, \tag{2.3}$$

where $A(z)$ and $B(z)$ are the ordinary generating functions of (a_n) and (b_n) respectively.

Remark 2.2. Catalan numbers are ubiquitous in combinatorial applications; see [22] for a list of 66 instances and `http://www-math.mit.edu/~rstan/ec/` for more.

In the sequel the notation $[\cdots]$ is used both for Iverson's convention [16, 1.2.3(16)] and for the coefficient of certain terms in the succeeding expression. The interpretation will be clear from the context. For example, $[\alpha > 0]$ has the value 1 when $\alpha > 0$ and the value 0 otherwise. In contrast, $[z^n]F(z)$ denotes the coefficient of z^n in the series expansion of $F(z)$. Throughout this paper Γ and ζ denote Euler's gamma function and the Riemann zeta function, respectively.

2.2. **Singularity analysis.** *Singularity analysis* is a systematic complex-analytic technique that relates asymptotics of sequences to singularities of their generating functions. The applicability of singularity analysis rests on the technical condition of Δ-*regularity*. Here is the definition. See [6] or [10] for further background.

Definition 2.3. A function defined by a Taylor series about the origin with radius of convergence equal to 1 is Δ-*regular* if it can be analytically continued in a domain

$$\Delta(\phi, \eta) := \{z : |z| < 1 + \eta, |\arg(z-1)| > \phi\},$$

for some $\eta > 0$ and $0 < \phi < \pi/2$. A function f is said to admit a *singular expansion* at $z = 1$ if it is Δ-regular and

$$f(z) = \sum_{j=0}^{J} c_j (1-z)^{\alpha_j} + O(|1-z|^A)$$

uniformly in $z \in \Delta(\phi, \eta)$, for a sequence of complex numbers $(c_j)_{0 \le j \le J}$ and an increasing sequence of real numbers $(\alpha_j)_{0 \le j \le J}$ satisfying $\alpha_j < A$. It is said to satisfy a singular expansion *"with logarithmic terms"* if, similarly,

$$f(z) = \sum_{j=0}^{J} c_j\,(L(z))\,(1-z)^{\alpha_j} + O(|1-z|^A), \qquad L(z) := \log \frac{1}{1-z},$$

where each $c_j(\cdot)$ is a polynomial.

Following established terminology, when a function has a singular expansion with logarithmic terms we shall say that it is *amenable* to singularity analysis.

Recall the definition of the *generalized polylogarithm*:

Definition 2.4. 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

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

The key property of the generalized polylogarithm that we will employ is

$$\mathrm{Li}_{\alpha,r} \odot \mathrm{Li}_{\beta,s} = \mathrm{Li}_{\alpha+\beta,r+s}\,.$$

We will also make extensive use of the following consequences of the singular expansion of the generalized polylogarithm. Neither this lemma nor the ones following make any claims about uniformity in α or r. Note that $\mathrm{Li}_{1,0}(z) = L(z) = \log\big((1-z)^{-1}\big)$.

Lemma 2.5. *For any real $\alpha < 1$ and nonnegative integer r, we have the singular expansion*

$$\mathrm{Li}_{\alpha,r}(z) = \sum_{k=0}^{r} \lambda_k^{(\alpha,r)} (1-z)^{\alpha-1} L^{r-k}(z) + O(|1-z|^{\alpha-\epsilon}) + (-1)^r \zeta^{(r)}(\alpha)[\alpha > 0],$$

where $\lambda_k^{(\alpha,r)} \equiv \binom{r}{k}\Gamma^{(k)}(1-\alpha)$ and $\epsilon > 0$ is arbitrarily small.

Proof. By Theorem 1 in [8],

$$\text{(2.4)}\quad \mathrm{Li}_{\alpha,0}(z) \sim \Gamma(1-\alpha)t^{\alpha-1} + \sum_{j\geq 0} \frac{(-1)^j}{j!}\zeta(\alpha-j)t^j, \qquad t = -\log z = \sum_{l=1}^{\infty} \frac{(1-z)^l}{l},$$

and for any positive integer r,

$$\mathrm{Li}_{\alpha,r}(z) = (-1)^r \frac{\partial^r}{\partial\alpha^r} \mathrm{Li}_{\alpha,0}(z).$$

Moreover, as also shown in [8], the singular expansion for $\mathrm{Li}_{\alpha,r}$ is obtained by performing the indicated differentiation of (2.4) term-by-term. To establish the claim we set $f = \Gamma(1-\alpha)$ and $g = t^{\alpha-1}$ in the general formula for the rth derivative of a product:

$$(fg)^{(r)} = \sum_{k=0}^{r} \binom{r}{k} f^{(k)} g^{(r-k)}$$

to first obtain

$$(-1)^r \frac{\partial^r}{\partial\alpha^r}[\Gamma(1-\alpha)t^{\alpha-1}] = (-1)^r \sum_{k=0}^{r} \binom{r}{k} (-1)^k \Gamma^{(k)}(1-\alpha) t^{\alpha-1} (\log t)^{r-k}$$

The claim then follows easily. □

The following "inverse" of Lemma 2.5 is very useful for computing with Hadamard products.

Lemma 2.6. *For any real $\alpha < 1$ and nonnegative integer r, there exists a region $\Delta(\phi,\eta)$ as in Defintion 2.3 such that*

$$(1-z)^{\alpha-1} L^r(z) = \sum_{k=0}^{r} \mu_k^{(\alpha,r)} \mathrm{Li}_{\alpha,r-k}(z) + O(|1-z|^{\alpha-\epsilon}) + c_r(\alpha)[\alpha > 0]$$

holds uniformly in $z \in \Delta(\phi,\eta)$, where $\mu_0^{(\alpha,r)} = 1/\Gamma(1-\alpha)$, $c_r(\alpha)$ is a constant, and $\epsilon > 0$ is arbitrarily small.

Proof. We use induction on r. For $r = 0$ we have

$$\mathrm{Li}_{\alpha,0}(z) = \Gamma(1-\alpha)(1-z)^{\alpha-1} + O(|1-z|^{\alpha-\epsilon}) + \zeta(\alpha)[\alpha > 0]$$

and the claim is verified with

$$\mu_0^{(\alpha,0)} = \frac{1}{\Gamma(1-\alpha)} \qquad \text{and} \qquad c_0(\alpha) = -\frac{\zeta(\alpha)}{\Gamma(1-\alpha)}.$$

Let $r \geq 1$. Then using Lemma 2.5 and the induction hypothesis we get

$$\begin{aligned}
&\operatorname{Li}_{\alpha,r}(z) \\
&= \Gamma(1-\alpha)(1-z)^{\alpha-1}L^r(z) \\
&\quad + \sum_{k=1}^{r} \lambda_k^{(\alpha,r)} \left[\sum_{l=0}^{r-k} \mu_l^{(\alpha,r-k)} \operatorname{Li}_{\alpha,r-k-l}(z) + O(|1-z|^{\alpha-\epsilon}) + c_{r-k}(\alpha)[\alpha > 0]\right] \\
&\quad + O(|1-z|^{\alpha-\epsilon}) + (-1)^r \zeta^{(r)}(\alpha)[\alpha > 0] \\
&= \Gamma(1-\alpha)(1-z)^{\alpha-1}L^r(z) + \sum_{k=1}^{r} \lambda_k^{(\alpha,r)} \sum_{s=0}^{r-k} \mu_{r-k-s}^{(\alpha,r-k)} \operatorname{Li}_{\alpha,s}(z) \\
&\quad + O(|1-z|^{\alpha-\epsilon}) + \left(\sum_{k=1}^{r} \lambda_k^{(\alpha,r)} c_{r-k}(\alpha) + (-1)^r \zeta^{(r)}(\alpha)\right)[\alpha > 0] \\
&= \Gamma(1-\alpha)(1-z)^{\alpha-1}L^r(z) + \sum_{s=0}^{r-1} \nu_s^{(\alpha,r)} \operatorname{Li}_{\alpha,s}(z) \\
&\quad + O(|1-z|^{\alpha-\epsilon}) + \gamma_r(\alpha)[\alpha > 0],
\end{aligned}$$

where, for $0 \leq s \leq r-1$,

$$\nu_s^{(\alpha,r)} := \sum_{k=1}^{r-s} \lambda_k^{(\alpha,r)} \mu_{r-s-k}^{(\alpha,r-k)},$$

and where

$$\gamma_r(\alpha) := \sum_{k=1}^{r} \lambda_k^{(\alpha,r)} c_{r-k}(\alpha) + (-1)^r \zeta^{(r)}(\alpha).$$

Setting

$$\mu_0^{(\alpha,r)} = \frac{1}{\Gamma(1-\alpha)}, \qquad \mu_k^{(\alpha,r)} = -\frac{\nu_{r-k}^{(\alpha,r)}}{\Gamma(1-\alpha)}, \quad 1 \leq k \leq r,$$

and

$$c_r(\alpha) = -\frac{\gamma_r(\alpha)}{\Gamma(1-\alpha)},$$

the result follows. □

For the calculation of the mean, the following refinement of a special case of Lemma 2.5 is required. It is a simple consequence of Theorem 1 of [8].

Lemma 2.7. *When $\alpha < 0$, we have the singular expansion*

$$\operatorname{Li}_{\alpha,0}(z) = \Gamma(1-\alpha)(1-z)^{\alpha-1} - \Gamma(1-\alpha)\frac{1-\alpha}{2}(1-z)^{\alpha} + O(|1-z|^{\alpha+1}) + \zeta(\alpha)[\alpha > -1].$$

For the sake of completeness, we state a result of particular relevance from [6].

Theorem 2.8. *If f and g are amenable to singularity analysis and*

$$f(z) = O(|1-z|^a) \qquad \text{and} \qquad g(z) = O(|1-z|^b)$$

as $z \to 1$, then $f \odot g$ is also amenable to singularity analysis. Furthermore

(a) If $a + b + 1 < 0$ then

$$f(z) \odot g(z) = O(|1-z|^{a+b+1}).$$

(b) If $k < a+b+1 < k+1$ for some integer $-1 \le k < \infty$, then

$$f(z) \odot g(z) = \sum_{j=0}^{k} \frac{(-1)^j}{j!} (f \odot g)^{(j)}(1)(1-z)^j + O(|1-z|^{a+b+1}).$$

(c) If $a+b+1$ is a nonnegative integer then

$$f(z) \odot g(z) = \sum_{j=0}^{a+b} \frac{(-1)^j}{j!} (f \odot g)^{(j)}(1)(1-z)^j + O(|1-z|^{a+b+1}|L(z)|).$$

3. The toll sequence (n^α)

In this section we consider additive functionals when the toll function b_n is n^α with $\alpha > 0$.

3.1. **Asympotics of the mean.** The main result of this Section 3.1 is a singular expansion for $A(z) \odot \mathrm{CAT}(z/4)$. The result is (3.1), (3.4), or (3.5) according as $\alpha < 1/2$, $\alpha = 1/2$, or $\alpha > 1/2$.

Since $b_n = n^\alpha$, by definition $B = \mathrm{Li}_{-\alpha,0}$. Thus, by Lemma 2.7,

$$B(z) = \Gamma(1+\alpha)(1-z)^{-\alpha-1} - \Gamma(1+\alpha)\frac{\alpha+1}{2}(1-z)^{-\alpha} + O(|1-z|^{-\alpha+1}) + \zeta(-\alpha)[\alpha < 1].$$

We will now use (2.3) to obtain the asymptotics of the mean.

First we treat the case $\alpha < 1/2$. From the singular expansion $\mathrm{CAT}(z/4) = 2 + O(|1-z|^{1/2})$ as $z \to 1$, we have, by part (b) of Theorem 2.8,

$$B(z) \odot \mathrm{CAT}(z/4) = C_0 + O(|1-z|^{-\alpha+\frac{1}{2}}),$$

where

$$C_0 := B(z) \odot \mathrm{CAT}(z/4)\Big|_{z=1} = \sum_{n=1}^{\infty} n^\alpha \frac{\beta_n}{4^n}.$$

We now already know the constant term in the singular expansion of $B(z) \odot \mathrm{CAT}(z/4)$ at $z = 1$ and henceforth we need only compute lower-order terms. The constant $\bar{c}$ is used in the sequel to denote an unspecified (possibly 0) constant, possibly different at each appearance.

Let's write $B(z) = L_1(z) + R_1(z)$, and $\mathrm{CAT}(z/4) = L_2(z) + R_2(z)$, where

$$\begin{aligned}
L_1(z) &:= \Gamma(1+\alpha)(1-z)^{-\alpha-1} - \Gamma(1+\alpha)\frac{\alpha+1}{2}(1-z)^{-\alpha} + \zeta(-\alpha),\\
R_1(z) &:= B(z) - L_1(z) = O(|1-z|^{1-\alpha}),\\
L_2(z) &:= 2(1-(1-z)^{1/2}),\\
R_2(z) &:= \mathrm{CAT}(z/4) - L_2(z) = O(|1-z|).
\end{aligned}$$

We will analyze each of the four Hadamard products separately. First,

$$\begin{aligned}
L_1(z) \odot L_2(z) = &-2\Gamma(1+\alpha)\left[(1-z)^{-\alpha-1} \odot (1-z)^{1/2}\right]\\
&+ 2\Gamma(1+\alpha)\frac{\alpha+1}{2}\left[(1-z)^{-\alpha} \odot (1-z)^{1/2}\right] + \bar{c}.
\end{aligned}$$

By Theorem 4.1 of [6],

$$(1-z)^{-\alpha-1} \odot (1-z)^{1/2} = \bar{c} + \frac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha+1)\Gamma(-1/2)}(1-z)^{-\alpha+\frac{1}{2}} + O(|1-z|),$$

and

$$(1-z)^{-\alpha} \odot (1-z)^{1/2} = \bar{c} + O(|1-z|)$$

by another application of part (b) of Theorem 2.8, this time with $k=1$. Hence

$$L_1(z) \odot L_2(z) = \big[L_1(z) \odot L_2(z)\big]\Big|_{z=1} + \frac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{\pi}}(1-z)^{-\alpha+\frac{1}{2}} + O(|1-z|).$$

The other three Hadamard products are easily handled as

$$\begin{aligned} L_1(z) \odot R_2(z) &= \big[L_1(z) \odot R_2(z)\big]\Big|_{z=1} + O(|1-z|^{-\alpha+1}), \\ L_2(z) \odot R_1(z) &= \big[L_2(z) \odot R_1(z)\big]\Big|_{z=1} + O(|1-z|), \\ R_1(z) \odot R_2(z) &= \big[R_1(z) \odot R_2(z)\big]\Big|_{z=1} + O(|1-z|). \end{aligned}$$

Putting everything together, we get

$$B(z) \odot \mathrm{CAT}(z/4) = C_0 + \frac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{\pi}}(1-z)^{-\alpha+\frac{1}{2}} + O(|1-z|^{-\alpha+1}).$$

Using this in (2.3), we get

$$A(z) \odot \mathrm{CAT}(z/4) = C_0(1-z)^{-1/2} + \frac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{\pi}}(1-z)^{-\alpha} + O(|1-z|^{-\alpha+\frac{1}{2}}). \tag{3.1}$$

To treat the case $\alpha \geq 1/2$ we make use of the estimate

$$(1-z)^{1/2} = \frac{1}{\Gamma(-1/2)}[\mathrm{Li}_{3/2,0}(z) - \zeta(3/2)] + O(|1-z|), \tag{3.2}$$

a consequence of Theorem 1 of [8], so that

$$B(z) \odot (1-z)^{1/2} = \mathrm{Li}_{-\alpha,0}(z) \odot (1-z)^{1/2} = \frac{1}{\Gamma(-1/2)} \mathrm{Li}_{\frac{3}{2}-\alpha,0}(z) + R(z),$$

where

$$R(z) = \begin{cases} \bar{c} + O(|1-z|^{1-\alpha}) & 1/2 \leq \alpha < 1 \\ O(|L(z)|) & \alpha = 1 \\ O(|1-z|^{1-\alpha}) & \alpha > 1. \end{cases} \tag{3.3}$$

Hence

$$B(z) \odot \mathrm{CAT}(z/4) = -\frac{2}{\Gamma(-1/2)} \mathrm{Li}_{\frac{3}{2}-\alpha,0}(z) + \widetilde{R}(z),$$

where $\widetilde{R}$, like R, satisfies (3.3) (with a possibly different $\bar{c}$). When $\alpha = 1/2$, this gives us

$$B(z) \odot \mathrm{CAT}(z/4) = -\frac{2}{\Gamma(-1/2)} L(z) + \bar{c} + O(|1-z|^{1/2}),$$

so that

$$A(z) \odot \mathrm{CAT}(z/4) = \frac{1}{\sqrt{\pi}}(1-z)^{-1/2} L(z) + \bar{c}(1-z)^{-1/2} + O(1). \tag{3.4}$$

For $\alpha > 1/2$ another singular expansion leads to the conclusion that

$$A(z) \odot \mathrm{CAT}(z/4) = \frac{\Gamma(\alpha - \frac{1}{2})}{\sqrt{\pi}}(1-z)^{-\alpha} + \widehat{R}(z), \tag{3.5}$$

where

$$\widehat{R}(z) = \begin{cases} O(|1-z|^{-\frac{1}{2}}) & 1/2 < \alpha < 1 \\ O(|1-z|^{-\frac{1}{2}}|L(z)|) & \alpha = 1 \\ O(|1-z|^{-\alpha+\frac{1}{2}}) & \alpha > 1. \end{cases}$$

We defer deriving the asymptotics of a_n until Sections 3.2–3.3.

3.2. **Higher moments.** We will analyze separately the cases $0 < \alpha < 1/2$, $\alpha = 1/2$, and $\alpha > 1/2$. The reason for this will become evident soon; though the technique used to derive the asymptotics is induction in each case, the induction hypothesis is different for each of these cases.

3.2.1. *Small toll functions (*$0 < \alpha < 1/2$*).* We start by restricting ourselves to tolls of the form n^α where $0 < \alpha < 1/2$. In this case we observe that by singularity analysis applied to (3.1),

$$\frac{a_n \beta_n}{4^n} = \frac{C_0}{\sqrt{\pi}} n^{-1/2} + O(n^{-3/2}) + O(n^{\alpha-1}) = \frac{C_0}{\sqrt{\pi}} n^{-1/2} + O(n^{\alpha-1}),$$

so

$$a_n = n^{\frac{3}{2}}[1 + O(n^{-1})][C_0 n^{-\frac{1}{2}} + O(n^{\alpha-1})] = C_0 n + O(n^{\alpha+\frac{1}{2}}) = (C_0 + o(1))(n+1).$$

The lead-order term of the mean $a_n = \mathbf{E}\, X_n$ is thus linear, irrespective of the value of $0 < \alpha < 1/2$ (though the coefficient C_0 does depend on α). We next perform an approximate centering to get to further dependence on α.

Define $\widetilde{X}_n := X_n - C_0(n+1)$, with $X_0 := 0$; $\tilde{\mu}_n(k) := \mathbf{E}\, \widetilde{X}_n^k$, with $\tilde{\mu}_n(0) = 1$ for all $n \geq 0$; and $\hat{\mu}_n(k) := \beta_n \tilde{\mu}_n(k)/4^n$. Let $\widehat{M}_k(z)$ denote the ordinary generating function of $\hat{\mu}_n(k)$ in the argument n.

By an argument similar to the one that led to (2.2), we get, for $k \geq 2$,

$$\hat{\mu}_n(k) = \frac{1}{2}\sum_{j=1}^{n} \frac{\beta_{n-j}}{4^{n-j}} \hat{\mu}_{j-1}(k) + \hat{r}_n(k), \qquad n \geq 1,$$

where

$$\begin{aligned} \hat{r}_n(k) &:= \frac{1}{4}\sum_{j=1}^{n} \sum_{\substack{k_1+k_2+k_3=k \\ k_1,k_2<k}} \binom{k}{k_1, k_2, k_3} \hat{\mu}_{j-1}(k_1)\hat{\mu}_{n-j}(k_2) b_n^{k_3} \\ &= \frac{1}{4} \sum_{\substack{k_1+k_2+k_3=k \\ k_1,k_2<k}} \binom{k}{k_1, k_2, k_3} b_n^{k_3} \sum_{j=1}^{n} \hat{\mu}_{j-1}(k_1)\hat{\mu}_{n-j}(k_2), \end{aligned}$$

for $n \geq 1$ and $\hat{r}_0(k) := \hat{\mu}_0(k) = \tilde{\mu}_0(k) = (-1)^k C_0^k$. Let $\widehat{R}_k(z)$ denote the ordinary generating function of $\hat{r}_n(k)$ in the argument n. Then, mimicking (2.3),

$$\widehat{M}_k(z) = \frac{\widehat{R}_k(z)}{\sqrt{1-z}} \tag{3.6}$$

with

$$\widehat{R}_k(z) = (-1)^k C_0^k + \sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}} \binom{k}{k_1,k_2,k_3} \left(B(z)^{\odot k_3}\right) \odot \left[\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z)\right], \tag{3.7}$$

where for k a nonnegative integer

$$B(z)^{\odot k} := \underbrace{B(z) \odot \cdots \odot B(z)}_{k}.$$

Note that $\widehat{M}_0(z) = \mathrm{CAT}(z/4)$.

Proposition 3.1. *Let $\epsilon > 0$ be arbitrary, and define*

$$c := \begin{cases} 2\alpha - \epsilon & 0 < \alpha \le 1/4 \\ 1/2 & 1/4 < \alpha < 1/2. \end{cases}$$

Then we have the singular expansion

$$\widehat{M}_k(z) = C_k(1-z)^{-k(\alpha+\frac{1}{2})+\frac{1}{2}} + O(|1-z|^{-k(\alpha+\frac{1}{2})+\frac{1}{2}+c}),$$

The C_k's here are defined by the recurrence

$$C_k = \frac{1}{4}\sum_{j=1}^{k-1}\binom{k}{j}C_jC_{k-j} + kC_{k-1}\frac{\Gamma(k\alpha+\frac{k}{2}-1)}{\Gamma((k-1)\alpha+\frac{k}{2}-1)}, \quad k \ge 2; \quad C_1 = \frac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{\pi}}. \tag{3.8}$$

Proof. For $k = 1$ the claim is true as shown in (3.1) with C_1 as defined in (3.8). We will now analyze each term in (3.7) for $k \ge 2$.

One can analyze separately the cases $0 < \alpha \le 1/4$ and $1/4 < \alpha < 1/2$. The proof technique in either case is induction. We shall treat here the case $0 < \alpha \le 1/4$; the details in the other case can be found in [7].

For notational convenience, define $\alpha' := \alpha + \frac{1}{2}$. Also, observe that

$$B(z)^{\odot k} = \mathrm{Li}_{-k\alpha,0}(z) = \Gamma(1+k\alpha)(1-z)^{-k\alpha-1} + O(|1-z|^{-k\alpha-\epsilon})$$

by Lemma 2.5. We shall find that the dominant terms in the sum in (3.7) are those with (i) $k_3 = 0$, (ii) $(k_1,k_2,k_3) = (k-1,1,0)$, and (iii) $(k_1,k_2,k_3) = (0,k-1,1)$.

For this paragraph, consider the case that k_1 and k_2 are both nonzero. It follows from the induction hypothesis that

$$\begin{aligned} \frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) &= \frac{1}{4}(1-(1-z))\left[C_{k_1}(1-z)^{-k_1\alpha'+\frac{1}{2}} + O(|1-z|^{-k_1\alpha'+\frac{1}{2}+(2\alpha-\epsilon)})\right] \\ &\quad \times \left[C_{k_2}(1-z)^{-k_2\alpha'+\frac{1}{2}} + O(|1-z|^{-k_2\alpha'+\frac{1}{2}+(2\alpha-\epsilon)})\right] \\ &= \frac{1}{4}C_{k_1}C_{k_2}(1-z)^{-(k_1+k_2)\alpha'+1} + O(|1-z|^{-(k_1+k_2)\alpha'+1+(2\alpha-\epsilon)}). \end{aligned}$$

If $k_3 = 0$ then the corresponding contribution to $\widehat{R}_k(z)$ is

$$\frac{1}{4}\binom{k}{k_1}C_{k_1}C_{k_2}(1-z)^{-k\alpha'+1} + O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)}).$$

If $k_3 \neq 0$ we use Lemma 2.6 to express

$$\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) = \frac{C_{k_1}C_{k_2}}{4\Gamma((k_1+k_2)\alpha'-1)}\operatorname{Li}_{-(k_1+k_2)\alpha'+2,0}(z)$$
$$+ O(|1-z|^{-(k_1+k_2)\alpha'+1+(2\alpha-\epsilon)}) - \frac{C_{k_1}C_{k_2}}{4}[(k_1+k_2)\alpha' < 2]\frac{\zeta(-(k_1+k_2)\alpha'+2)}{\Gamma((k_1+k_2)\alpha'-1)}.$$

The corresponding contribution to $\widehat{R}_k(z)$ is then $\binom{k}{k_1,k_2,k_3}$ times:

$$\frac{C_{k_1}C_{k_2}}{4\Gamma((k_1+k_2)\alpha'-1)}\operatorname{Li}_{-k\alpha'+\frac{k_3}{2}+2,0}(z) + \operatorname{Li}_{-k_3\alpha,0}(z) \odot O(|1-z|^{-(k_1+k_2)\alpha'+1+(2\alpha-\epsilon)}).$$

Now $k_3 \leq k-2$ so $-k\alpha' + \frac{k_3}{2} + 2 < 1$. Hence the contribution when $k_3 \neq 0$ is

$$O(|1-z|^{-k\alpha'+\frac{k_3}{2}+1}) = O(|1-z|^{-k\alpha'+\frac{3}{2}}) = O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)}).$$

Next we consider the case when k_1 is nonzero but $k_2 = 0$. In this case using the induction hypothesis we see that

$$\begin{aligned}\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) &= \frac{z}{4}\operatorname{CAT}(z/4)\widehat{M}_{k_1}(z)\\ &= \frac{1-(1-z)^{1/2}}{2}\left[C_{k_1}(1-z)^{-k_1\alpha'+\frac{1}{2}}\right] + O(|1-z|^{-k_1\alpha'+\frac{1}{2}+(2\alpha-\epsilon)})\\ &= \frac{C_{k_1}}{2}(1-z)^{-k_1\alpha'+\frac{1}{2}} + O(|1-z|^{-k_1\alpha'+\frac{1}{2}+(2\alpha-\epsilon)}).\end{aligned}$$

Applying Lemma 2.6 to the last expression we get

$$\begin{aligned}\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) &= \frac{C_{k_1}}{2\Gamma(k_1\alpha'-\frac{1}{2})}\operatorname{Li}_{-k_1\alpha'+\frac{3}{2},0}(z)\\ &\quad + O(|1-z|^{-k_1\alpha'+\frac{1}{2}+(2\alpha-\epsilon)}) - \frac{C_{k_1}}{2}[k_1\alpha'-\tfrac{1}{2} < 1]\frac{\zeta(-k_1\alpha'+\frac{3}{2})}{\Gamma(k_1\alpha'-\frac{1}{2})}.\end{aligned}$$

The contribution to $\widehat{R}_k(z)$ is hence $\binom{k}{k_1}$ times:

$$\frac{C_{k_1}}{2\Gamma(k_1\alpha'-\frac{1}{2})}\operatorname{Li}_{-k\alpha'+\frac{k_3}{2}+\frac{3}{2},0}(z) + \operatorname{Li}_{-k_3\alpha,0}(z) \odot O(|1-z|^{-k_1\alpha'+\frac{1}{2}+(2\alpha-\epsilon)}).$$

Using the fact that $\alpha > 0$ and $k_3 \leq k-1$, we conclude that $-k\alpha' + \frac{k_3}{2} + \frac{3}{2} < 1$ so that, by Lemma 2.5 and part (a) of Theorem 2.8, the contribution is

$$O(|1-z|^{-k\alpha'+\frac{k_3}{2}+\frac{1}{2}}) = O(|1-z|^{-k\alpha'+\frac{3}{2}})$$

where the displayed equality holds unless $k_3 = 1$. When $k_3 = 1$ we get a corresponding contribution to $\widehat{R}_k(z)$ of $\binom{k}{k-1}$ times:

$$\frac{C_{k-1}\Gamma(k\alpha'-1)}{2\Gamma((k-1)\alpha'-\frac{1}{2})}(1-z)^{-k\alpha'+1} + O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)}),$$

since for $k \geq 2$ we have $k\alpha' > 1 + (2\alpha - \epsilon)$. The introduction of ϵ handles the case when $k\alpha' = 1 + 2\alpha$, which would have otherwise, according to part (c) of Thoerem 2.8, introduced a logarithmic remainder. In either case the remainder is $O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)})$. The case when k_2 is nonzero but $k_1 = 0$ is handled similarly by exchanging the roles of k_1 and k_2.

The final contribution comes from the single term where both k_1 and k_2 are zero. In this case the contribution to $\widehat{R}_k(z)$ is, recalling (2.1),

(3.9)

$$\mathrm{Li}_{-k\alpha,0}(z)\odot[\frac{z}{4}\,\mathrm{CAT}^2(z/4)] = \mathrm{Li}_{-k\alpha,0}(z)\odot(\mathrm{CAT}(z/4)-1) = \mathrm{Li}_{-k\alpha,0}(z)\odot\mathrm{CAT}(z/4).$$

Now, using Theorem 1 of [8],

$$\begin{aligned}\mathrm{CAT}(z/4) &= 2 - 2(1-z)^{1/2} + O(|1-z|)\\ &= 2 + 2\frac{\zeta(3/2)}{\Gamma(-1/2)} - \frac{2}{\Gamma(-1/2)}\,\mathrm{Li}_{3/2,0}(z) + O(|1-z|),\end{aligned}$$

so that (3.9) is

$$-\frac{2}{\Gamma(-1/2)}\,\mathrm{Li}_{\frac{3}{2}-k\alpha,0}(z) + O(|1-z|^{1-k\alpha}) + \begin{cases} 0 & 1-k\alpha<0,\\ O(|1-z|^{-\epsilon}) & 1-k\alpha=0,\\ O(1) & 1-k\alpha>0.\end{cases}$$

When $\frac{3}{2}-k\alpha<1$ this is $O(|1-z|^{-k\alpha+\frac{1}{2}})$; when $\frac{3}{2}-k\alpha\ge 1$, it is $O(1)$. In either case we get a contribution which is $O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)})$.

Hence

$$\begin{aligned}\widehat{R}_k(z) &= \Bigg[\sum_{\substack{k_1+k_2=k\\ k_1,k_2<k}}\binom{k}{k_1}\frac{C_{k_1}C_{k_2}}{4} + 2k\frac{C_{k-1}}{2}\,\frac{\Gamma(k\alpha+\frac{k}{2}-1)}{\Gamma((k-1)\alpha+\frac{k}{2}-1)}\Bigg](1-z)^{-k\alpha'+1}\\ &\qquad + O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)})\\ &= C_k(1-z)^{-k\alpha'+1} + O(|1-z|^{-k\alpha'+1+(2\alpha-\epsilon)}),\end{aligned}$$

with the C_k's defined by the recurrence (3.8). Now using (3.6), the claim follows. □

3.2.2. *Large toll functions ($\alpha \ge 1/2$).* When $\alpha \ge 1/2$ there is no need to apply the centering techiniques. Define $\mu_n(k) := \mathbf{E}\,X_n^k$ and $\bar{\mu}_n(k) := \beta_n\mu_n(k)/4^n$. Let $\overline{M}_k(z)$ denote the ordinary generating function of $\bar{\mu}_n(k)$ in n. Observe that $\overline{M}_0(z) = \mathrm{CAT}(z/4)$. As earlier, conditioning on the key stored at the root, we get, for $k\ge 2$,

$$\bar{\mu}_n(k) = \frac{1}{2}\sum_{j=1}^{n}\frac{\beta_{n-j}}{4^{n-j}}\,\bar{\mu}_{j-1}(k) + \bar{r}_n(k), \qquad n\ge 1,$$

where

$$\bar{r}_n(k) := \frac{1}{4}\sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}}\binom{k}{k_1,k_2,k_3} b_n^{k_3}\sum_{j=1}^{n}\bar{\mu}_{j-1}(k_1)\bar{\mu}_{n-j}(k_2),$$

for $n\ge 1$ and $\bar{r}_0(k) := \bar{\mu}_0(k) = \mu_0(k) = 0$. Let $\overline{R}_k(z)$ denote the ordinary generating function of $\bar{r}_n(k)$ in n. Then

$$\overline{M}_k(z) = \frac{\overline{R}_k(z)}{\sqrt{1-z}}$$

and

$$\overline{R}_k(z) = \sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}}\binom{k}{k_1,k_2,k_3}\big(B(z)^{\odot k_3}\big)\odot\big[\frac{z}{4}\overline{M}_{k_1}(z)\overline{M}_{k_2}(z)\big]. \tag{3.10}$$

We can now state the result about the asymptotics of the generating function $\overline{M}_k$ when $\alpha > 1/2$. The case $\alpha = 1/2$ will be handled subsequently, in Proposition 3.3.

Proposition 3.2. *Let $\epsilon > 0$ be arbitrary, and define*

$$c := \begin{cases} \alpha - \frac{1}{2} & \frac{1}{2} < \alpha < 1 \\ \frac{1}{2} - \epsilon & \alpha = 1 \\ \frac{1}{2} & \alpha > 1. \end{cases} \tag{3.11}$$

Then the generating function $\overline{M}_k(z)$ of $\bar{\mu}_n(k)$ has the singular expansion

$$\overline{M}_k(z) = C_k(1-z)^{-k(\alpha+\frac{1}{2})+\frac{1}{2}} + O(|1-z|^{-k(\alpha+\frac{1}{2})+\frac{1}{2}+c})$$

for $k \geq 1$, where the C_k's are defined by the recurrence (3.8).

Proof. The proof is very similar to that of Proposition 3.1. We present a sketch. The reader is invited to compare the cases enumerated below to those in the earlier proof.

When $k = 1$ the claim is true by (3.5). We analyze the various terms in (3.10) for $k \geq 2$, employing the notational convenience $\alpha' := \alpha + \frac{1}{2}$.

When both k_1 and k_2 are nonzero then the contribution to $\overline{R}_k(z)$ is

$$\frac{1}{4}\binom{k}{k_1} C_{k_1} C_{k_2} (1-z)^{-k\alpha'+1} + O(|1-z|^{-k\alpha'+c+1})$$

when $k_3 = 0$ and is $O(|1-z|^{-k\alpha'+c+1})$ otherwise.

When k_1 is nonzero and $k_2 = 0$ the contribution to $\overline{R}_k(z)$ is

$$k\frac{C_{k-1}\Gamma(k\alpha'-1)}{2\Gamma((k-1)\alpha'-\frac{1}{2})}(1-z)^{-k\alpha'+1} + O(|1-z|^{-k\alpha'+c+1})$$

when $k_3 = 1$ and $O(|1-z|^{-k\alpha'+c+1})$ otherwise. The case when k_2 is nonzero and $k_1 = 0$ is identical.

The final contribution comes from the single term when both k_1 and k_2 are zero. In this case we get a contribution of $O(|1-z|^{-k\alpha+\frac{1}{2}})$ which is $O(|1-z|^{-k\alpha'+c+1})$. Adding all these contributions yields the desired result. □

The result when $\alpha = 1/2$ is as follows. Recall that $L(z) := \log((1-z)^{-1})$.

Proposition 3.3. *Let $\alpha = 1/2$. In the notation of Proposition 3.2,*

$$\overline{M}_k(z) = (1-z)^{-k+\frac{1}{2}} \sum_{l=0}^{k} C_{k,l} L^{k-l}(z) + O(|1-z|^{-k+1-\epsilon})$$

for $k \geq 1$ and any $\epsilon > 0$, where the $C_{k,l}$'s are constants. The constant multiplying the lead-order term is given by

$$C_{k,0} = \frac{(2k-2)!}{2^{2k-2}(k-1)!\pi^{k/2}}. \tag{3.12}$$

Proof. We omit the proof, referring the interested reader to [7]. □

3.3. **Asymptotics of moments.** For $0 < \alpha < 1/2$, we have seen in Proposition 3.1 that the generating function $\widehat{M}_k(z)$ of $\hat{\mu}_n(k) = \beta_n \tilde{\mu}_n(k)/4^n$ has the singular expansion

$$\widehat{M}_k(z) = C_k(1-z)^{-k(\alpha+\frac{1}{2})+\frac{1}{2}} + O(|1-z|^{-k(\alpha+\frac{1}{2})+\frac{1}{2}+c}),$$

where $c := \min\{2\alpha - \epsilon, 1/2\}$. By singularity analysis [10],

$$\frac{\beta_n \tilde{\mu}_n(k)}{4^n} = C_k \frac{n^{k(\alpha+\frac{1}{2})-\frac{3}{2}}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})} + O(n^{k(\alpha+\frac{1}{2})-\frac{3}{2}-c}).$$

Recall that

$$\beta_n = \frac{4^n}{\sqrt{\pi} n^{3/2}} \left(1 + O(\tfrac{1}{n})\right),$$

so that

$$\tilde{\mu}_n(k) = \frac{C_k\sqrt{\pi}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})} n^{k(\alpha+\frac{1}{2})} + O(n^{k(\alpha+\frac{1}{2})-c}). \tag{3.13}$$

For $\alpha > 1/2$ a similar analysis using Proposition 3.2 yields

$$\mu_n(k) = \frac{C_k\sqrt{\pi}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})} n^{k(\alpha+\frac{1}{2})} + O(n^{k(\alpha+\frac{1}{2})-c}), \tag{3.14}$$

with now c as defined at (3.11). Finally, when $\alpha = 1/2$ the asymptotics of the moments are given by

$$\mu_n(k) = \left(\frac{1}{\sqrt{\pi}}\right)^k (n \log n)^k + O(n^k (\log n)^{k-1}). \tag{3.15}$$

3.4. **The limiting distributions.** In Section 3.4.1 we will use our moment estimates (3.13) and (3.14) with the method of moments to derive limiting distributions for our additive functions. The case $\alpha = 1/2$ requires a somewhat delicate analysis, which we will present separately in Section 3.4.2.

3.4.1. $\alpha \neq 1/2$. We first handle the case $0 < \alpha < 1/2$. (We assume this restriction until just before Proposition 3.5.) We have

$$\tilde{\mu}_n(1) = \mathbf{E}\,\widetilde{X}_n = \mathbf{E}\,[X_n - C_0(n+1)] = \frac{C_1\sqrt{\pi}}{\Gamma(\alpha)} n^{\alpha+\frac{1}{2}} + O(n^{\alpha+\frac{1}{2}-c}) \tag{3.16}$$

with $c := \min\{2\alpha - \epsilon, 1/2\}$ and

$$\tilde{\mu}_n(2) = \mathbf{E}\,\widetilde{X}_n^2 = \frac{C_2\sqrt{\pi}}{\Gamma(2\alpha+\frac{1}{2})} n^{2\alpha+1} + O(n^{2\alpha+1-c}).$$

So

$$\mathbf{Var}\,X_n = \mathbf{Var}\,\widetilde{X}_n = \tilde{\mu}_n(2) - [\tilde{\mu}_n(1)]^2 = \sigma^2 n^{2\alpha+1} + O(n^{2\alpha+1-c}), \tag{3.17}$$

where

$$\sigma^2 := \frac{C_2\sqrt{\pi}}{\Gamma(2\alpha+\frac{1}{2})} - \frac{C_1^2\pi}{\Gamma^2(\alpha)}. \tag{3.18}$$

We also have, for $k \geq 1$,

$$\mathbf{E}\left[\frac{\widetilde{X}_n}{n^{\alpha+\frac{1}{2}}}\right]^k = \frac{\tilde{\mu}_n(k)}{n^{k(\alpha+\frac{1}{2})}} = \frac{C_k\sqrt{\pi}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})} + O(n^{-c}). \tag{3.19}$$

The following lemma provides a sufficient bound on the moments facilitating the use of the method of moments.

Lemma 3.4. *Define* $\alpha' := \alpha + \frac{1}{2}$. *There exists a constant* $A < \infty$ *depending only on* α *such that*

$$\left|\frac{C_k}{k!}\right| \leq A^k k^{\alpha' k}$$

for all $k \geq 1$.

Proof. The proof is fairly similar to those of Propositions 3.1, 3.2 and Proposition 4.1. We omit the details, referring the reader to [7]. ☐

It follows from Lemma 3.4 and Stirling's approximation that

$$\left|\frac{C_k\sqrt{\pi}}{k!\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})}\right| \leq B^k \tag{3.20}$$

for large enough B depending only on α. Using standard arguments [1, Theorem 30.1] it follows that X_n suitably normalized has a limiting distribution that is characterized by its moments. Before we state the result, we observe that the argument presented above can be adapted with minor modifications to treat the case $\alpha > 1/2$, with $\widetilde{X}_n$ replaced by X_n. We can now state a result for $\alpha \neq 1/2$. We will use the notation $\xrightarrow{\mathcal{L}}$ to denote convergence in law (or distribution).

Proposition 3.5. *Let* X_n *denote the additive functional on Catalan trees induced by the toll sequence* $(n^\alpha)_{n\geq 0}$. *Define the random variable* Y_n *as follows:*

$$Y_n := \begin{cases} \dfrac{X_n - C_0(n+1)}{n^{\alpha+\frac{1}{2}}} & 0 < \alpha < 1/2, \\ \dfrac{X_n}{n^{\alpha+\frac{1}{2}}} & \alpha > 1/2, \end{cases}$$

where

$$C_0 := \sum_{n=0}^{\infty} n^\alpha \frac{\beta_n}{4^n}, \qquad \beta_n = \frac{1}{n+1}\binom{2n}{n}.$$

Then

$$Y_n \xrightarrow{\mathcal{L}} Y;$$

here Y *is a random variable with the unique distribution whose moments are*

$$\mathbf{E}\,Y^k = \frac{C_k\sqrt{\pi}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})}, \tag{3.21}$$

where the C_k*'s satisfy the recurrence*

$$C_k = \frac{1}{4}\sum_{j=1}^{k-1}\binom{k}{j}C_jC_{k-j} + k\frac{\Gamma(k\alpha+\frac{k}{2}-1)}{\Gamma((k-1)\alpha+\frac{k}{2}-1)}C_{k-1}, k \geq 2; \quad C_1 = \frac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{\pi}}.$$

The case $\alpha = 1/2$ is handled in Section 3.4.2, leading to Proposition 3.8, and a unified result for all cases is stated as Theorem 3.10.

Remark 3.6. We now consider some properties of the limiting random variable $Y \equiv Y(\alpha)$ defined by its moments at (3.21) for $\alpha \neq 1/2$.

(a) When $\alpha = 1$, setting $\Omega_k := C_k/2$ we see immediately that

$$\mathbf{E}\,Y^k = \frac{-\Gamma(-1/2)}{\Gamma((3k-1)/2)}\Omega_k,$$

where

$$2\Omega_k = \sum_{j=1}^{k-1}\binom{k}{j}\Omega_j\Omega_{k-j} + k(3k-4)\Omega_{k-1}, \qquad \Omega_1 = \frac{1}{2}.$$

Thus Y has the ubiquitous Airy distribution and we have recovered the limiting distribution of path length in Catalan trees [23, 25]. The Airy distribution arises in many contexts including parking allocations, hashing tables, trees, discrete random walks, mergesorting, etc.—see, for example, the introduction of [9] which contains numerous references to the Airy distribution.

(b) When $\alpha = 2$, setting $\eta := Y/\sqrt{2}$ and $a_{0,l} := 2^{2l-1}C_l$, we see that

$$\mathbf{E}\,\eta^l = \frac{\sqrt{\pi}}{2^{(5l-2)/2}\Gamma((5l-1)/2)}a_{0,l},$$

where

$$a_{0,l} = \frac{1}{2}\sum_{j=1}^{l-1}\binom{l}{j}a_{0,j}a_{0,l-j} + l(5l-4)(5l-6), \qquad a_{0,1} = 1.$$

We have thus recovered the recurrence for the moments of the distribution $\mathcal{L}(\eta)$, which arises in the study of the Wiener index of Catalan trees [15, proof of Theorem 3.3 in Section 5].

(c) Consider the variance σ^2 defined at (3.18).
   (i) Figure 3.1, plotted using `Mathematica`, suggests that σ^2 is positive for all $\alpha > 0$. We will prove this fact in Theorem 3.10. There is also numerical

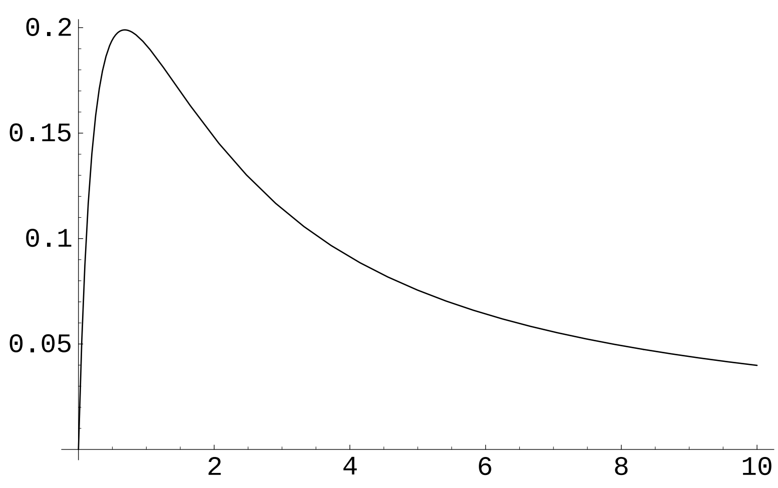

FIGURE 3.1. σ^2 of (3.18) as a function of α.

evidence that σ^2 is unimodal with $\max_\alpha \sigma^2(\alpha) \doteq 0.198946$ achieved at $\alpha \doteq 0.682607$. (Here $\doteq$ denotes approximate equality.)

(ii) As $\alpha \to \infty$, using Stirling's approximation one can show that $\sigma^2 \sim (\sqrt{2}-1)\alpha^{-1}$.

(iii) As $\alpha \downarrow 0$, using a Laurent series expansion of $\Gamma(\alpha)$ we see that $\sigma^2 \sim 4(1-\log 2)\alpha$.

(iv) Though the random variable $Y(\alpha)$ has been defined only for $\alpha \neq 1/2$, the variance σ^2 has a limit at $\alpha = 1/2$:

$$\lim_{\alpha \to 1/2} \sigma^2(\alpha) = \frac{8\log 2}{\pi} - \frac{\pi}{2}. \tag{3.22}$$

(d) Figure 3.2 shows the third central moment $\mathbf{E}[Y - \mathbf{E}\,Y]^3$ as a function of α. The plot suggests that the third central moment is positive for each $\alpha > 0$, which would also establish that $Y(\alpha)$ is not normal for any $\alpha > 0$. However we do not know a proof of this positive skewness. [Of course, the law of $Y(\alpha)$ is not normal for any $\alpha > 1/2$, since its support is a subset of $[0, \infty)$.]

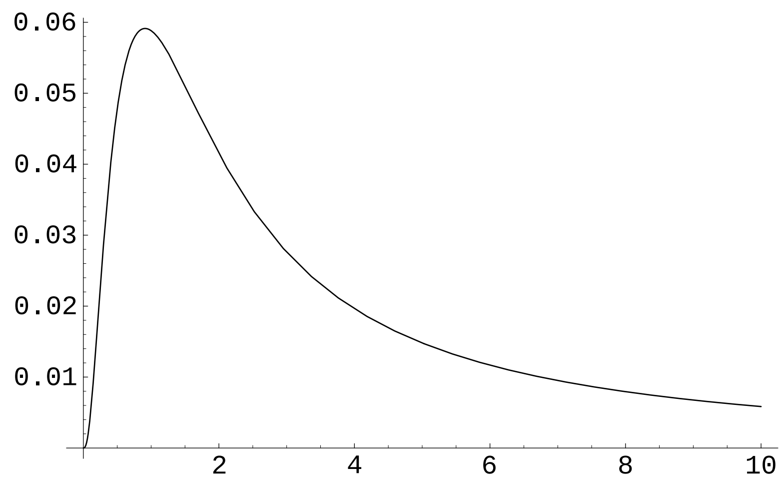

FIGURE 3.2. $\mathbf{E}[Y - \mathbf{E}\,Y]^3$ of Proposition 3.5 as a function of α.

(e) When $\alpha = 0$, the additive functional with toll sequence $(n^\alpha = 1)_{n \geq 1}$ is n for all trees with n nodes. However, if one considers the random variable $\alpha^{-1/2}Y(\alpha)$ as $\alpha \downarrow 0$, using (3.21) and induction one can show that $\alpha^{-1/2}Y(\alpha)$ converges in distribution to the normal distribution with mean 0 and variance $4(1-\log 2)$.

(f) Finally, if one considers the random variable $\alpha^{1/2}Y(\alpha)$ as $\alpha \to \infty$, again using (3.21) and induction we find that $\alpha^{1/2}Y(\alpha)$ converges in distribution to the unique distribution with kth moment $\sqrt{k!}$ for $k = 1, 2, \ldots$. In Remark 3.7 next, we will show that the limiting distribution has a bounded, infinitely smooth density on $(0, \infty)$.

Remark 3.7. Let Y be the unique distribution whose kth moment is $\sqrt{k!}$ for $k = 1, 2, \ldots$. Taking Y^* to be an independent copy of Y and defining $X := YY^*$, we see immediately that X is Exponential with unit mean. It follows by taking logarithms that the distribution of $\log Y$ is a convolution square root of the distribution of $\log X$. In particular, the characteristic function ϕ of $\log Y$ has square equal to $\Gamma(1+it)$ at $t \in (-\infty, \infty)$; we note in passing that $\Gamma(1+it)$ is the characteristic function of $-G$, where G has the Gumbel distribution. By exponential decay of $\Gamma(1+it)$ as $t \to \pm\infty$ and standard theory (see, e.g., [4, Chapter XV]), $\log Y$ has an infinitely smooth density on $(-\infty, \infty)$, and the density and each of its derivatives are bounded.

So Y has an infinitely smooth density on $(0, \infty)$. By change of variables, the density f_Y of Y satisfies

$$f_Y(y) = \frac{f_{\log Y}(\log y)}{y}.$$

Clearly $f_Y(y)$ is bounded for y *not* near 0. (We shall drop further consideration of derivatives.) To determine the behavior near 0, we need to know the behavior of $f_{\log Y}(\log y)/y$ as $y \to 0$. Using the Fourier inversion formula, we may equivalently study

$$e^x f_{\log Y}(-x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{(1+it)x} \phi(t)\, dt,$$

as $x \to \infty$. By an application of the method of steepest descents [(7.2.11) in [2], with $g_0 = 1$, $\beta = 1/2$, w the identity map, $z_0 = 0$, and $\alpha = 0$], we get

$$f_Y(y) \sim \frac{1}{\sqrt{\pi \log(1/y)}} \quad \text{as } y \downarrow 0.$$

Hence f_Y is bounded everywhere.

Using the Cauchy integral formula and simple estimates, it is easy to show that

$$f_Y(y) = o(e^{-My}) \quad \text{as } y \to \infty$$

for any $M < \infty$. Computations using the WKB method [12] suggest

$$f_Y(y) \sim (2/\pi)^{1/4} y^{1/2} \exp(-y^2/2) \quad \text{as } y \to \infty, \tag{3.23}$$

in agreement with numerical calculations using `Mathematica`. [In fact, the right-side of (3.23) appears to be a highly accurate approximation to $f_Y(y)$ for all $y \geq 1$.] Figure 3.3 depicts the salient features of f_Y. In particular, note the steep descent of $f_Y(y)$ to 0 as $y \downarrow 0$ and the quasi-Gaussian tail.

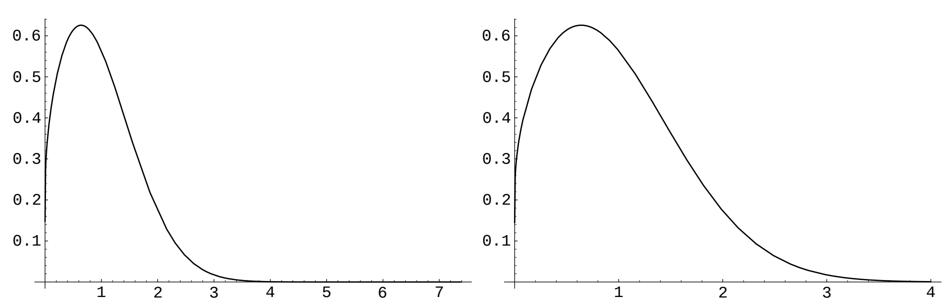

FIGURE 3.3. f_Y of Remark 3.7.

3.4.2. $\alpha = 1/2$. For $\alpha = 1/2$, from (3.15) we see immediately that

$$\mathbf{E}\left[\frac{X_n}{n\log n}\right]^k = \left(\frac{1}{\sqrt{\pi}}\right)^k + O\left(\frac{1}{\log n}\right).$$

Thus the random variable $X_n/(n\log n)$ converges in distribution to the degenerate random variable $1/\sqrt{\pi}$. To get a nondegenerate distribution, we carry out an analysis similar to the one that led to (3.4), getting more precise asymptotics for the mean of X_n. The refinement of (3.4) that we need is the following, whose proof we omit:

$$A(z)\odot \mathrm{CAT}(z/4) = \frac{1}{\sqrt{\pi}}(1-z)^{-1/2}L(z) + D_0(1-z)^{-1/2} + O(|1-z|^{\frac{1}{2}-\epsilon}),$$

where

$$D_0 = \sum_{n=1}^{\infty} n^{1/2}[4^{-n}\beta_n - \frac{1}{\sqrt{\pi}}n^{-3/2}]. \tag{3.24}$$

By singularity analysis this leads to

$$\mathbf{E}\,X_n = \frac{1}{\sqrt{\pi}}n\log n + D_1 n + O(n^{\epsilon}), \tag{3.25}$$

where

$$D_1 = \frac{1}{\sqrt{\pi}}(2\log 2 + \gamma + \sqrt{\pi}D_0). \tag{3.26}$$

Now analyzing the random variable $X_n - \pi^{-1/2}n\log n$ in a manner similar to that of Section 3.2.1 we obtain

$$\mathbf{Var}\,[X_n - \pi^{-1/2}n\log n] = \left(\frac{8}{\pi}\log 2 - \frac{\pi}{2}\right)n^2 + O(n^{\frac{3}{2}+\epsilon}). \tag{3.27}$$

Using (3.25) and (3.27) we conclude that

$$\mathbf{E}\left[\frac{X_n - \pi^{-1/2}n\log n - D_1 n}{n}\right] = o(1)$$

and

$$\mathbf{Var}\left[\frac{X_n - \pi^{-1/2}n\log n - D_1 n}{n}\right] \longrightarrow \frac{8}{\pi}\log 2 - \frac{\pi}{2} = \lim_{\alpha\to 1/2}\sigma^2(\alpha), \tag{3.28}$$

where $\sigma^2 \equiv \sigma^2(\alpha)$ is defined at (3.18) for $\alpha \neq 1/2$. [Recall (3.22) of Remark 3.6.]

It is possible to carry out a program similar to that of Section 3.2 to derive asymptotics of higher order moments using singularity analysis. However we choose to sidestep this arduous, albeit mechanical, computation. Instead we will derive the asymptotics of higher moments using a somewhat more direct approach akin to the one employed in [5]. The approach involves approximation of sums by Riemann integrals. To that end, define

(3.29)

$$\widetilde{X}_n := X_n - \pi^{-1/2}(n+1)\log(n+1) - D_1(n+1), \quad \text{and} \quad \hat{\mu}_n(k) := \frac{\beta_n}{4^{n+1}}\mathbf{E}\,\widetilde{X}_n^k.$$

Note that $\widetilde{X}_0 = -D_1$, $\hat{\mu}_n(0) = \beta_n/4^{n+1}$, and $\hat{\mu}_0(k) = (-D_1)^k/4$. Then, in a now familiar manner, for $n \geq 1$ we find

$$\hat{\mu}_n(k) = 2\sum_{j=1}^{n} \frac{\beta_{j-1}}{4^j}\hat{\mu}_{n-j}(k) + \hat{r}_n(k),$$

where now we define

$$\hat{r}_n(k) := \sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}} \binom{k}{k_1,k_2,k_3} \sum_{j=1}^{n} \hat{\mu}_{j-1}(k_1)\hat{\mu}_{n-j}(k_2)$$
$$\times \left[\frac{1}{\sqrt{\pi}}(j\log j + (n+1-j)\log(n+1-j) - (n+1)\log(n+1) + \sqrt{\pi}n^{1/2})\right]^{k_3}$$

Passing to generating functions and then back to sequences one gets, for $n \geq 0$,

$$\hat{\mu}_n(k) = \sum_{j=0}^{n}(j+1)\frac{\beta_j}{4^j}\hat{r}_{n-j}(k).$$

Using induction on k, we can approximate $\hat{r}_n(k)$ and $\hat{\mu}_n(k)$ above by integrals and obtain the following result. We omit the proof, leaving it as an exercise for the ambitious reader.

Proposition 3.8. *Let X_n be the additive functional induced by the toll sequence $(n^{1/2})_{n\geq 1}$ on Catalan trees. Define $\widetilde{X}_n$ as in* (3.29)*, with D_1 defined at* (3.26) *and D_0 at* (3.24)*. Then*

$$\mathbf{E}\,[\widetilde{X}_n/n]^k = m_k + o(1) \text{ as } n \to \infty,$$

where $m_0 = 1$, $m_1 = 0$, and, for $k \geq 2$,

$$m_k = \frac{1}{4\sqrt{\pi}}\frac{\Gamma(k-1)}{\Gamma(k-\frac{1}{2})} \times \left[\sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}} \binom{k}{k_1,k_2,k_3} m_{k_1} m_{k_2} \left(\frac{1}{\sqrt{\pi}}\right)^{k_3} J_{k_1,k_2,k_3} + 4\sqrt{\pi}k m_{k-1}\right], \tag{3.30}$$

where

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

Furthermore $\widetilde{X}_n/(n+1) \xrightarrow{\mathcal{L}} Y$, where Y is a random variable with the unique distribution whose moments are $\mathbf{E}\,Y^k = m_k$, $k \geq 0$.

3.4.3. *A unified result.* The approach outlined in the preceding section can also be used for the case $\alpha \neq 1/2$. For completeness, we state the result for that case here (without proof).

Proposition 3.9. *Let X_n be the additive functional induced by the toll sequence $(n^\alpha)_{n\ge1}$ on Catalan trees. Let $\alpha' := \alpha + \frac{1}{2}$. Define $\widetilde{X}_n$ as*

$$\widetilde{X}_n := \begin{cases} X_n - C_0(n+1) - \dfrac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)}(n+1)^{\alpha'} & 0<\alpha<1/2, \\ X_n - \dfrac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)}(n+1)^{\alpha'} & \alpha > 1/2, \end{cases} \tag{3.31}$$

where

$$C_0 := \sum_{n=1}^{\infty} n^\alpha \frac{\beta_n}{4^n}.$$

Then, for $k = 0, 1, 2, \ldots,$

$$\mathbf{E}\left[\widetilde{X}_n/n^{\alpha'}\right]^k = m_k + o(1) \text{ as } n\to\infty,$$

where $m_0 = 1$, $m_1 = 0$, and, for $k \ge 2$,

$$m_k = \frac{1}{4\sqrt{\pi}}\frac{\Gamma(k\alpha'-1)}{\Gamma(k\alpha'-\frac{1}{2})} \times \left[\sum_{\substack{k_1+k_2+k_3=k\\ k_1,k_2<k}} \binom{k}{k_1,k_2,k_3} m_{k_1} m_{k_2} \left(\frac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)}\right)^{k_3} J_{k_1,k_2,k_3} + 4\sqrt{\pi}k m_{k-1}\right], \tag{3.32}$$

with

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

Furthermore, $\widetilde{X}_n/n^{\alpha'} \xrightarrow{\mathcal{L}} Y$, where Y is a random variable with the unique distribution whose moments are $\mathbf{E}\,Y^k = m_k$.

[The reader may wonder as to why we have chosen to state Proposition 3.9 using several instances of $n+1$, rather than n, in (3.31). The reason is that use of $n+1$ is somewhat more natural in the calculations that establish the proposition.]

In light of Propositions 3.5, 3.8, and 3.9, there are a variety of ways to state a unified result. We state one such version here.

Theorem 3.10. *Let X_n denote the additive functional induced by the toll sequence $(n^\alpha)_{n\ge1}$ on Catalan trees. Then*

$$\frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} W,$$

where the distribution of W is described as follows:

(a) For $\alpha \ne 1/2$,

$$W = \frac{1}{\sigma}\left(Y - \frac{C_1\sqrt{\pi}}{\Gamma(\alpha)}\right), \quad \text{with} \quad \sigma^2 := \frac{C_2\sqrt{\pi}}{\Gamma(2\alpha+\frac{1}{2})} - \frac{C_1^2\pi}{\Gamma^2(\alpha)} > 0,$$

where Y is a random variable with the unique distribution whose moments are

$$\mathbf{E}\,Y^k = \frac{C_k\sqrt{\pi}}{\Gamma(k(\alpha+\frac{1}{2})-\frac{1}{2})},$$

and the C_k's satisfy the recurrence (3.8).

(b) For $\alpha = 1/2$,

$$W = \frac{Y}{\sigma}, \qquad \text{with} \qquad \sigma^2 := \frac{8}{\pi}\log 2 - \frac{\pi}{2},$$

where Y is a random variable with the unique distribution whose moments $m_k = \mathbf{E}\, Y^k$ are given by (3.30).

Proof. Define

$$W_n := \frac{X_n - \mathbf{E}\, X_n}{\sqrt{\mathbf{Var}\, X_n}}$$

(a) Consider first the case $\alpha < 1/2$ and let $\alpha' := \alpha + \frac{1}{2}$. By (3.16),

$$\mathbf{E}\, X_n = C_0(n+1) + \frac{C_1\sqrt{\pi}}{\Gamma(\alpha)} n^{\alpha'} + o(n^{\alpha'}). \tag{3.33}$$

Since $\widetilde{X}_n$ defined at (3.31) and X_n differ by a deterministic amount, $\mathbf{Var}\, X_n = \mathbf{Var}\, \widetilde{X}_n$. Now by Proposition 3.9,

$$\mathbf{Var}\, \widetilde{X}_n = \mathbf{E}\, \widetilde{X}_n^2 - (\mathbf{E}\, \widetilde{X}_n)^2 = (m_2 + o(1))n^{2\alpha'} - (m_1^2 + o(1))n^{2\alpha'} = (m_2 + o(1))n^{2\alpha'}. \tag{3.34}$$

So σ^2 equals m_2 defined at (3.32), namely,

$$\frac{1}{4\sqrt{\pi}} \frac{\Gamma(2\alpha' - 1)}{\Gamma(2\alpha' - \frac{1}{2})} \left(\frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} \right)^2 J_{0,0,2}.$$

Thus to show $\sigma^2 > 0$ it is enough to show that $J_{0,0,2} > 0$. But

$$J_{0,0,2} = \int_0^1 x^{-3/2}(1-x)^{-3/2}[x^{\alpha'} + (1-x)^{\alpha'} - 1]^2\, dx,$$

which is clearly positive. Using (3.33) and (3.34),

$$W_n = \frac{X_n - C_0(n+1) - \frac{C_1\sqrt{\pi}}{\Gamma(\alpha)} n^{\alpha'} + o(n^{\alpha'})}{(1 + o(1))\sigma n^{\alpha'}},$$

so, by Proposition 3.5 and Slutsky's theorem [1, Theorem 25.4], the claim follows.

The case $\alpha > 1/2$ follows similarly.

(b) When $\alpha = 1/2$,

$$\mathbf{E}\, X_n = \frac{1}{\sqrt{\pi}} n \log n + D_1 n + o(n)$$

by (3.25) and

$$\mathbf{Var}\, X_n = \left(\frac{8}{\pi}\log 2 - \frac{\pi}{2} + o(1) \right) n^2$$

by (3.28). The claim then follows easily from Proposition 3.8 and Slutsky's theorem. □

4. The shape functional

We now turn our attention to the shape functional for Catalan trees. The shape functional is the cost induced by the toll function $b_n \equiv \log n$, $n \geq 1$. For background and results on the shape functional, we refer the reader to [5] and [19].

In the sequel we will improve on the mean and variance estimates obtained in [5] and derive a central limit theorem for the shape functional for Catalan trees. The technique employed is singularity analysis followed by the method of moments.

4.1. **Mean.** We use the notation and techniques of Section 3.1 again. Observe that now $B(z) = \mathrm{Li}_{0,1}(z)$ and (3.2) gives the singular expansion

$$\mathrm{CAT}(z/4) = 2 - \frac{2}{\Gamma(-1/2)}[\mathrm{Li}_{3/2,0}(z) - \zeta(3/2)] + 2\left(1 - \frac{\zeta(1/2)}{\Gamma(-1/2)}\right)(1-z) + O(|1-z|^{3/2}).$$

So

$$B(z) \odot \mathrm{CAT}(z/4) = -\frac{2}{\Gamma(-1/2)}\,\mathrm{Li}_{3/2,1}(z) + \bar{c} + \bar{\bar{c}}(1-z) + O(|1-z|^{\frac{3}{2}-\epsilon}),$$

where $\bar{c}$ and $\bar{\bar{c}}$ denote unspecified (possibly 0) constants. The constant term in the singular expansion of $B(z) \odot \mathrm{CAT}(z/4)$ is already known to be

$$C_0 = B(z) \odot \mathrm{CAT}(z/4)\Big|_{z=1} = \sum_{n=1}^{\infty} (\log n) \frac{\beta_n}{4^n}.$$

Now using the singular expansion of $\mathrm{Li}_{3/2,1}(z)$, we get

$$B(z) \odot \mathrm{CAT}(z/4) = C_0 - 2(1-z)^{1/2} L(z) - 2(2(1-\log(2)) - \gamma)(1-z)^{1/2} + O(|1-z|),$$

so that

$$A(z) \odot \mathrm{CAT}(z/4) = C_0 (1-z)^{-1/2} - 2L(z) - 2(2(1-\log 2) - \gamma) + O(|1-z|^{1/2}). \tag{4.1}$$

Using singularity analysis and the asymptotics of the Catalan numbers we get that the mean a_n of the shape functional is given by

$$a_n = C_0(n+1) - 2\sqrt{\pi} n^{1/2} + O(1), \tag{4.2}$$

which agrees with the estimate in Theorem 3.1 of [5] and improves the remainder estimate.

4.2. **Second moment and variance.** We now derive the asymptotics of the approximately centered second moment and the variance of the shape functional. These estimates will serve as the basis for the induction to follow. We will use the notation of Section 3.2.1, centering the cost function as before by $C_0(n+1)$.

It is clear from (4.1) that

$$\widehat{M}_1(z) = -2L(z) - 2(2(1-\log 2) - \gamma) + O(|1-z|^{1/2}), \tag{4.3}$$

and (3.7) with $k = 2$ gives us, recalling (2.1),

(4.4)

$$\widehat{R}_2(z) = C_0^2 + \mathrm{CAT}(z/4) \odot \mathrm{Li}_{0,2}(z) + 4\,\mathrm{Li}_{0,1}(z) \odot [\frac{z}{4}\,\mathrm{CAT}(z/4)\widehat{M}_1(z)] + \frac{z}{2}\widehat{M}_1^2(z).$$

We analyze each of the terms in this sum. For the last term, observe that $z/2 \to 1/2$ as $z \to 1$, so that

$$\frac{z}{2}\widehat{M}_1^2(z) = 2L^2(z) + 4(2(1-\log 2)-\gamma)L(z) + 2(2(1-\log 2)-\gamma)^2 + O(|1-z|^{\frac{1}{2}-\epsilon}),$$

the ϵ introduced to avoid logarithmic remainders. The first term is easily seen to be

$$\mathrm{CAT}(z/4) \odot \mathrm{Li}_{0,2}(z) = K + O(|1-z|^{\frac{1}{2}-\epsilon}),$$

where

$$K := \sum_{n=1}^{\infty} (\log n)^2 \frac{\beta_n}{4^n}.$$

For the middle term, first observe that

$$\frac{z}{4}\mathrm{CAT}(z/4)\widehat{M}_1(z) = -L(z) - (2(1-\log 2)-\gamma) + (1-z)^{1/2}L(z) + O(|1-z|^{1/2})$$

and that $L(z) = \mathrm{Li}_{1,0}(z)$. Thus the third term on the right in (4.4) is 4 times:

$$-\mathrm{Li}_{1,1}(z) + \bar{c} + O(|1-z|^{\frac{1}{2}-2\epsilon}) = -\frac{1}{2}L^2(z) + \gamma L(z) + \bar{c} + O(|1-z|^{\frac{1}{2}-\epsilon}).$$

[The singular expansion for $\mathrm{Li}_{1,1}(z)$ was obtained using the results at the bottom of p. 379 in [8]. We state it here for the reader's convenience:

$$\mathrm{Li}_{1,1}(z) = \frac{1}{2}L^2(z) - \gamma L(z) + \bar{c} + O(|1-z|),$$

where $\bar{c}$ is again an unspecified constant.] Hence

$$\widehat{R}_2(z) = 8(1-\log 2)L(z) + \bar{c} + O(|1-z|^{\frac{1}{2}-\epsilon}),$$

which leads to

$$\widehat{M}_2(z) = 8(1-\log 2)(1-z)^{-1/2}L(z) + \bar{c}(1-z)^{-1/2} + O(|1-z|^{-\epsilon}). \tag{4.5}$$

We draw the attention of the reader to the cancellation of the ostensible lead-order term $L^2(z)$. This kind of cancellation will appear again in the next section when we deal with higher moments.

Now using singularity analysis and estimates for the Catalan numbers we get

$$\tilde{\mu}_n(2) = 8(1-\log 2)n\log n + \bar{c}n + O(n^{\frac{1}{2}+\epsilon}). \tag{4.6}$$

Using (4.2),

$$\mathbf{Var}\, X_n = \tilde{\mu}_n(2) - \tilde{\mu}_n(1)^2 = 8(1-\log 2)n\log n + \bar{c}n + O(n^{\frac{1}{2}+\epsilon}),$$

which agrees with Theorem 3.1 of [5] (after a correction pointed out in [19]) and improves the remainder estimate. In our subsequent analysis we will not need to evaluate the unspecified constant $\bar{c}$.

4.3. **Higher moments.** We now turn our attention to deriving the asymptotics of higher moments of the shape functional. The main result is as follows.

Proposition 4.1. *Define* $\widetilde{X}_n := X_n - C_0(n+1)$*, with* $X_0 := 0$*;* $\tilde{\mu}_n(k) := \mathbf{E}\,\widetilde{X}_n^k$*, with* $\tilde{\mu}_n(0) = 1$ *for all* $n \geq 0$*; and* $\hat{\mu}_n(k) := \beta_n \tilde{\mu}_n(k)/4^n$*. Let* $\widehat{M}_k(z)$ *denote the ordinary generating function of* $\hat{\mu}_n(k)$ *in the argument* n*. For* $k \geq 2$*,* $\widehat{M}_k(z)$ *has the singular expansion*

$$\widehat{M}_k(z) = (1-z)^{-\frac{k-1}{2}} \sum_{j=0}^{\lfloor k/2 \rfloor} C_{k,j} L^{\lfloor k/2 \rfloor - j}(z) + O(|1-z|^{-\frac{k}{2}+1-\epsilon}),$$

with

$$C_{2l,0} = \frac{1}{4} \sum_{j=1}^{l-1} \binom{2l}{2j} C_{2j,0} C_{2l-2j,0}, \qquad C_{2,0} = 8(1-\log 2).$$

Proof. The proof is by induction. For $k = 2$ the claim is true by (4.5). We note that the claim is *not* true for $k = 1$. Instead, recalling (4.3),

$$\widehat{M}_1(z) = -2L(z) - 2(2(1-\log 2) - \gamma) + O(|1-z|^{1/2}). \tag{4.7}$$

For the induction step, let $k \geq 3$. We will first get the asymptotics of $\widehat{R}_k(z)$ defined at (3.7) with $B(z) = \mathrm{Li}_{0,1}(z)$. In order to do that we will obtain the asymptotics of each term in the defining sum. We remind the reader that we are only interested in the form of the asymptotic expansion of $\widehat{R}_k(z)$ and the coefficient of the lead-order term when k is even. This allows us to "define away" all other constants, their determination delayed to the time when the need arises.

For this paragraph suppose that $k_1 \geq 2$ and $k_2 \geq 2$. Then by the induction hypothesis

$$\frac{z}{4} \widehat{M}_{k_1}(z) \widehat{M}_{k_2}(z) = \frac{1}{4} (1-z)^{-\frac{k_1+k_2}{2}+1} \sum_{l=0}^{\lfloor k_1/2 \rfloor + \lfloor k_2/2 \rfloor} A_{k_1,k_2,l} L^{\lfloor k_1/2 \rfloor + \lfloor k_2/2 \rfloor - l}(z) + O(|1-z|^{-\frac{k_1+k_2}{2}+\frac{3}{2}-\epsilon}), \tag{4.8}$$

where $A_{k_1,k_2,0} = C_{k_1,0} C_{k_2,0}$. (a) If $k_3 = 0$ then $k_1 + k_2 = k$ and the corresponding contribution to $\widehat{R}_k(z)$ is given by

$$\frac{1}{4} \binom{k}{k_1} (1-z)^{-\frac{k}{2}+1} \times \sum_{l=0}^{\lfloor k_1/2 \rfloor + \lfloor (k-k_1)/2 \rfloor} A_{k_1,k-k_1,l} L^{\lfloor k_1/2 \rfloor + \lfloor (k-k_1)/2 \rfloor - l}(z) + O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-\epsilon}). \tag{4.9}$$

Observe that if k is even and k_1 is odd the highest power of $L(z)$ in (4.9) is $\lfloor k/2 \rfloor - 1$. In all other cases the the highest power of $L(z)$ in (4.9) is $\lfloor k/2 \rfloor$. (b) If $k_3 \neq 0$ then we use Lemma 2.6 to express (4.8) as a linear combination of

$$\left\{ \mathrm{Li}_{-\frac{k_1+k_2}{2}+2,l}(z) \right\}_{l=0}^{\lfloor k_1/2 \rfloor + \lfloor k_2/2 \rfloor}$$

with a remainder that is $O(|1-z|^{-\frac{k_1+k_2}{2}+\frac{3}{2}-\epsilon})$. When we take the Hadamard product of such a term with $\mathrm{Li}_{0,k_3}(z)$ we will get a linear combination of

$$\left\{\mathrm{Li}_{-\frac{k_1+k_2}{2}+2,l+k_3}(z)\right\}_{l=0}^{\lfloor k_1/2\rfloor+\lfloor k_2/2\rfloor}$$

and a smaller remainder. Such terms are all $O(|1-z|^{-\frac{k_1+k_2}{2}+1-\epsilon})$, so that the contribution is $O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-\epsilon})$.

Next, consider the case when $k_1=1$ and $k_2\geq 2$. Using the induction hypothesis and (4.7) we get

$$\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) = -\frac{1}{2}(1-z)^{-\frac{k_2-1}{2}}\sum_{j=0}^{\lfloor k_2/2\rfloor+1} B_{k_2,j}L^{\left\lfloor\frac{k_2}{2}\right\rfloor+1-j}(z) + O(|1-z|^{-\frac{k_2}{2}+1-2\epsilon}), \tag{4.10}$$

with $B_{k_2,0}=C_{k_2,0}$. (a) If $k_3=0$ then $k_2=k-1$ and the corresponding contribution to $\widehat{R}_k(z)$ is given by

$$-\frac{k}{2}(1-z)^{-\frac{k}{2}+1}\sum_{j=0}^{\lfloor (k-1)/2\rfloor+1} B_{k-1,j}L^{\left\lfloor\frac{k-1}{2}\right\rfloor+1-j}(z) + O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-2\epsilon}). \tag{4.11}$$

(b) If $k_3\neq 0$ then Lemma 2.6 can be used once again to express (4.10) in terms of generalized polylogarithms, whence an argument similar to that at the end of the preceding paragraph yields that the contributions to $\widehat{R}(z)$ from such terms is $O(|1-z|^{-\frac{k_2-1}{2}-\epsilon})$, which is $O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-\epsilon})$. The case when $k_1\geq 2$ and $k_2=1$ is handled symmetrically.

When $k_1=k_2=1$ then $(z/4)\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z)$ is $O(|1-z|^{-\epsilon})$ and when one takes the Hadamard product of this term with $\mathrm{Li}_{0,k_3}(z)$ the contribution will be $O(|1-z|^{-2\epsilon})$.

Now consider the case when $k_1=0$ and $k_2\geq 2$. Since $\widehat{M_0}(z)=\mathrm{CAT}(z/4)$, we have

$$\frac{z}{4}\widehat{M}_{k_1}(z)\widehat{M}_{k_2}(z) = \frac{1}{2}(1-z)^{-\frac{k_2-1}{2}}\sum_{j=0}^{\lfloor k_2/2\rfloor} C_{k_2,j}L^{\lfloor k_2/2\rfloor-j}(z) + O(|1-z|^{-\frac{k_2}{2}+1-\epsilon}). \tag{4.12}$$

By Lemma 2.6 this can be expressed as a linear combination of

$$\left\{\mathrm{Li}_{-\frac{k_2-1}{2}+1,j}(z)\right\}_{j=0}^{\lfloor k_2/2\rfloor}$$

with a $O(|1-z|^{-\frac{k_2}{2}+1-\epsilon})$ remainder. When we take the Hadamard product of such a term with $\mathrm{Li}_{0,k_3}(z)$ we will get a linear combination, call it $S(z)$, of

$$\left\{\mathrm{Li}_{-\frac{k_2-1}{2}+1,j+k_3}(z)\right\}_{j=0}^{\lfloor k_2/2\rfloor}$$

with a remainder of $O(|1-z|^{-\frac{k_2}{2}+1-2\epsilon})$, which is $O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-2\epsilon})$ unless $k_2=k-1$. When $k_2=k-1$, by Lemma 2.6 the constant multiplying the lead-order term

$\mathrm{Li}_{-\frac{k}{2}+2,\lfloor\frac{k-1}{2}\rfloor+1}(z)$ in $S(z)$ is $\frac{C_{k-1,0}}{2}\mu_0^{(-\frac{k}{2}+2,\lfloor\frac{k-1}{2}\rfloor)}$. When we take the Hadamard product of this term with $\mathrm{Li}_{0,k_3}(z)$ we get a lead-order term of

$$\frac{C_{k-1,0}}{2}\mu_0^{(-\frac{k}{2}+2,\lfloor\frac{k-1}{2}\rfloor)}\,\mathrm{Li}_{-\frac{k}{2}+2,\lfloor\frac{k-1}{2}\rfloor+1}(z).$$

Now we use Lemma 2.5 and the observation that $\lambda_0^{(\alpha,r)}\mu_0^{(\alpha,s)}=1$ to conclude that the contribution to $\widehat{R}_k(z)$ from the term with $k_1=0$ and $k_2=k-1$ is

$$\frac{k}{2}(1-z)^{-\frac{k}{2}+1}\sum_{j=0}^{\lfloor\frac{k-1}{2}\rfloor+1} D_{k,j}L^{\lfloor\frac{k-1}{2}\rfloor+1-j}(z)+O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-\epsilon}), \tag{4.13}$$

with $D_{k,0}=C_{k-1,0}$. Notice that the lead order from this contribution is precisely that from (4.11) but with opposite sign; thus the two contributions cancel each other to lead order. The case $k_2=0$ and $k_1\ge 2$ is handled symmetrically.

The last two cases are $k_1=0$, $k_2=1$ (or vice-versa) and $k_1=k_2=0$. The contribution from these cases can be easily seen to be $O(|1-z|^{-\frac{k}{2}+\frac{3}{2}-2\epsilon})$.

We can now deduce the asymptotic behavior of $\widehat{R}_k(z)$. The three contributions are (4.9), (4.11), and (4.13), with only (4.9) (in net) contributing a term of the form $(1-z)^{-\frac{k}{2}+1}L^{\lfloor k/2\rfloor}(z)$ when k is even. The coefficient of this term when k is even is given by

$$\frac{1}{4}\sum_{\substack{0<k_1<k\\ k_1\text{ even}}}\binom{k}{k_1}C_{k_1,0}C_{k_2,0}.$$

Finally we can sum up the rest of the contribution, define $C_{k,j}$ appropriately and use (3.6) to claim the result. □

4.4. **A central limit theorem.** Proposition 4.1 and singularity analysis allows us to get the asymptotics of the moments of the "approximately centered" shape functional. Using arguments identical to those in Section 3.3 it is clear that for $k\ge 2$

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

This and the asymptotics of the mean derived in Section 4.1 give us, for $k\ge 1$,

$$E\left[\frac{\tilde{X}_n}{\sqrt{n\log n}}\right]^{2k}\to\frac{C_{2k,0}\sqrt{\pi}}{\Gamma(k-\frac{1}{2})},\qquad E\left[\frac{\tilde{X}_n}{\sqrt{n\log n}}\right]^{2k-1}=o(1)$$

as $n\to\infty$. The recurrence for $C_{2k,0}$ can be solved easily to yield, for $k\ge 1$,

$$C_{2k,0}=\frac{(2k)!(2k-2)!}{2^k2^{2k-2}k!(k-1)!}\sigma^{2k},$$

where $\sigma^2:=8(1-\log 2)$. Then using the identity

$$\frac{\Gamma(k-\frac{1}{2})}{\sqrt{\pi}}=\left[2^{2k-2}\frac{(k-1)!}{(2k-2)!}\right]^{-1}$$

we get

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

It is clear now that both the "approximately centered" and the normalized shape functional are asymptotically normal.

Theorem 4.2. *Let X_n denote the shape functional, induced by the toll sequence $(\log n)_{n\geq 1}$, for Catalan trees. Then*

$$\frac{X_n - C_0(n+1)}{\sqrt{n\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2) \quad \text{and} \quad \frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where

$$C_0 := \sum_{n=1}^{\infty}(\log n)\frac{\beta_n}{4^n}, \qquad \beta_n = \frac{1}{n+1}\binom{2n}{n},$$

and $\sigma^2 := 8(1-\log 2)$.

Concerning numerical evaluation of the constant C_0, see the end of Section 5.2 in [6].

5. Sufficient conditions for asymptotic normality

In this speculative final section we briefly examine the behavior of a general additive functional X_n induced by a given "small" toll sequence (b_n). We have seen evidence [Remark 3.6(d)] that if (b_n) is the "large" toll sequence n^α for any fixed $\alpha > 0$, then the limiting behavior is non-normal. When $b_n = \log n$ (or $b_n = n^\alpha$ and $\alpha \downarrow 0$), the (limiting) random variable is normal. Where is the interface between normal and non-normal asymptotics? We have carried out arguments similar to those leading to Propositions 3.8 and 3.9 (see also [5]) that suggest a sufficient condition for asymptotic normality, but our "proof" is somewhat heuristic, and further technical conditions on (b_n) may be required. Nevertheless, to inspire further work, we present our preliminary indications.

We assume that $b_n \equiv b(n)$, where $b(\cdot)$ is a function of a nonnegative real argument. Suppose that $x^{-3/2}b(x)$ is (ultimately) nonincreasing and that $xb'(x)$ is slowly varying at infinity. Then

$$\mathbf{E}\,X_n = C_0(n+1) - (1+o(1))2\sqrt{\pi}n^{3/2}b'(n),$$

where

$$C_0 = \sum_{n=1}^{\infty} b_n \frac{\beta_n}{4^n}.$$

Furthermore,

$$\mathbf{Var}\,X_n \sim 8(1-\log 2)[nb'(n)]^2 n\log n,$$

and

$$\frac{X_n - C_0(n+1)}{nb'(n)\sqrt{n\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2), \text{ where } \sigma^2 = 8(1-\log 2).$$

This asymptotic normality can also be stated in the form

$$\frac{X_n - \mathbf{E}\,X_n}{\sqrt{\mathbf{Var}\,X_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Acknowledgments. We thank two anonymous referees for helpful comments.

References

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

[2] N. Bleistein and R. A. Handelsman. *Asymptotic expansions of integrals*. Dover Publications Inc., New York, second edition, 1986.

[3] 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).

[4] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[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, P. Flajolet, and N. Kapur. Singularity analysis, Hadamard products, and Tree recurrences, 2003, arXiv:math.CO/0306225. Submitted for publication.

[7] J. A. Fill and N. Kapur. Limiting distributions of addititive functionals on Catalan trees, 2003, arXiv:math.PR/0306226. Version 1 of the present paper.

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

[9] P. Flajolet and G. Louchard. Analytic variations on the Airy distribution. *Algorithmica*, 31(3):361–377, 2001. Mathematical analysis of algorithms.

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

[11] P. Flajolet and J.-M. Steyaert. A complexity calculus for recursive tree algorithms. *Math. Systems Theory*, 19(4):301–331, 1987.

[12] N. Fröman and P. O. Fröman. *JWKB approximation. Contributions to the theory*. North-Holland Publishing Co., Amsterdam, 1965.

[13] H.-K. Hwang and R. Neininger. Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.*, 31(6):1687–1722, 2002.

[14] S. Janson. Ideals in a forest, one-way infinite binary trees and the contraction method. In *Mathematics and computer science, II (Versailles, 2002)*, Trends Math., pages 393–414. Birkhäuser, Basel, 2002.

[15] S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

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

[17] Y. Koda and F. Ruskey. A Gray code for the ideals of a forest poset. *J. Algorithms*, 15(2):324–340, 1993.

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

[19] A. Meir and J. W. Moon. On the log-product of the subtree-sizes of random trees. *Random Structures Algorithms*, 12(2):197–212, 1998.

[20] R. Neininger. On binary search tree recursions with monomials as toll functions. *J. Comput. Appl. Math.*, 142(1):185–196, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[21] U. Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

[22] R. P. Stanley. *Enumerative combinatorics. Vol. 2*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[23] L. Takács. A Bernoulli excursion and its various applications. *Adv. in Appl. Probab.*, 23(3):557–585, 1991.

[24] L. Takács. Conditional limit theorems for branching processes. *J. Appl. Math. Stochastic Anal.*, 4(4):263–292, 1991.

[25] L. Takács. On a probability problem connected with railway traffic. *J. Appl. Math. Stochastic Anal.*, 4(1):1–27, 1991.

E-mail address, James Allen Fill: jimfill@jhu.edu
URL, James Allen Fill: http://www.ams.jhu.edu/~fill/

E-mail address, Nevin Kapur: nkapur@cs.caltech.edu
URL, Nevin Kapur: http://www.cs.caltech.edu/~nkapur/

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

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