Binary Trees, Left and Right Paths, WKB Expansions, and Painlevé Transcendents*

Charles Knessl† Wojciech Szpankowski‡

Abstract

During the *10th Seminar on Analysis of Algorithms*, MSRI, Berkeley, June 2004, Knuth posed the problem of analyzing the left and the right path length in a random binary trees. In particular, Knuth asked about properties of the generating function of the joint distribution of the left and the right path lengths. In this paper, we mostly focus on the asymptotic properties of the distribution of the difference between the left and the right path lengths. Among other things, we show that the Laplace transform of the appropriately normalized moment generating function of the path difference satisfies the *first Painlevé transcendent*. This is a nonlinear differential equation that has appeared in many modern applications, from nonlinear waves to random matrices. Surprisingly, we find out that the difference between path lengths is of the order $n^{5/4}$ where n is the number of nodes in the binary tree. This was also recently observed by Marckert and Janson. We present precise asymptotics of the distribution's tails and moments. We shall also discuss the joint distribution of the left and right path lengths. Throughout, we use methods of analytic algorithmics such as generating functions and complex asymptotics, as well as methods of applied mathematics such as the WKB method.

1 Introduction

Trees are the most important nonlinear structures that arise in computer science. Applications are in abundance; here we discuss binary unlabeled ordered trees (further called binary trees) and study their asymptotic properties when the number of nodes, n, becomes large. While various interesting questions concerning statistics of randomly generated binary trees were investigated since Euler and Cayley [8, 17, 18, 25, 27, 28], recently novel applications have been surfacing. In 2003 Seroussi [22], when studying universal types for sequences and Lempel-Ziv'78 parsings, asked for the number of binary trees of *given* path length (sum of all paths from the root to all nodes). This was an open problem; partial solutions are reported in [15, 23].

During the *10th Seminar on Analysis of Algorithms*, MSRI, Berkeley, June 2004, Knuth asked to analyze the joint distribution of the left and the right path lengths in random binary trees. This problem received a lot of attention in the community (cf. related papers [11, 20]) and leads to an interesting analysis, that encompasses several other problems studied recently [11, 15, 19, 20, 22, 23, 27]. Here, we mostly focus on the asymptotic properties of the distribution of the difference between the left and the right path lengths. However, we also obtain some results for the joint distribution of the left and the right path lengths in a random binary tree.

In the standard model, that we adopt here, one selects uniformly a tree among all binary unlabeled ordered trees built on n nodes, $\mathcal{T}_n$ (where $|\mathcal{T}_n| = \binom{2n}{n}\frac{1}{n+1}$ =Catalan number). Many deep and interesting results concerning the behavior of binary trees in the standard model were uncovered. For example, Flajolet and Odlyzko [6] and Takacs [27] established the average and the limiting distribution for the height (longest path), while Louchard [19] and Takacs [26, 27, 28] derive the limiting distribution for the path length. As we indicate below, these limiting distributions are expressible in terms of the Airy's function (cf. [2, 3]). Recently, Seroussi [22, 23], and Knessl and Szpankowski [15] analyzed properties of random binary trees when selected uniformly from the set $\mathcal{T}_t$ of all binary trees of given path length t. Among other results, they enumerated the number of trees in $\mathcal{T}_t$ and analyze the number of nodes in a randomly selected tree from $\mathcal{T}_t$.

We now summarize our main results and put them into a bigger perspective. Let $N_n(p,q)$ be the number of binary trees built on n nodes with the right path length equal to p and the left path length equal to q. It is easy

*The work was supported by NSF Grants CCR-0208709 and DMS-02-02815, NIH Grant R01 GM068959-01, and NSA Grant MDA 904-03-1-0036

†Dept. Math., Stat. & Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607-7045, U.S.A

‡Department of Computer Science, Purdue University, West Lafayette, IN 47907-2066 U.S.A.

to see that its generating function $G_n(w,v)$ satisfies

$$G_{n+1}(w,v) = \sum_{i=0}^{n} w^i v^{n-i} G_i(w,v) G_{n-i}(w,v) \tag{1.1}$$

with $G_0(w,v) = 1$. Summing over n we obtain the triple transform $C(w,v,z)$ (cf. also (2.12) below) that satisfies

$$C(w,v,z) = 1 + zC(w,v,wz)C(w,v,vz). \tag{1.2}$$

This is exactly the equation that Knuth asked to analyze.

The above functional equation encompasses many properties of binary trees. Let us first set $w = v$ and define $C(w,z) = C(w,w,z)$. Recurrence (1.2) then becomes

$$C(w,z) = 1 + zC^2(w,zw). \tag{1.3}$$

Observe that this equation is asymmetric with respect to z and w. When enumerating trees in $\mathcal{T}_n$, we set $w = 1$ to get the well known algebraic equation $C(1,z) = 1 + zC^2(1,z)$ that can be explicitly solved as $C(1,z) = \left(1 - \sqrt{1-4z}\right)/(2z)$, leading to the Catalan number. A randomly (uniformly) selected tree from $\mathcal{T}_n$ has path length T_n that is asymptotically distributed as Airy's distribution [26, 27], that is,

$$\mathsf{Pr}\{T_n/\sqrt{2n^3} \le x\} \to W(x)$$

where $W(x)$ is the Airy distribution function defined by its moments [7]. The Airy distribution arises in surprisingly many contexts, such as parking allocations, hashing tables, trees, discrete random walks, area under a Brownian bridge, etc. [7, 19, 26, 27, 28].

Setting $z = 1$ in (1.3) we arrive at

$$C(w,1) = 1 + C^2(w,w)$$

which does not seem to be explicitly solvable, e.g., by using a Taylor expansion in w. Observe that the coefficient of w^t in $C(w,1)$ enumerates binary trees with a *given* path length t. In [15, 23] it was shown that

$$[w^t]C(w,1) = |\mathcal{T}_t| =$$

$$= \frac{1}{(\log_2 t)\sqrt{\pi t}} 2^{\frac{2t}{\log_2 t}\left(1 + c_1 \log^{-2/3} t + c_2 \log^{-1} t + O(\log^{-4/3} t)\right)}$$

for large t, where c_1 and c_2 are explicitly computable constants.

Let us now set $v = 1$ in (1.2). Then

$$C(w,1,z) = 1 + zC(w,1,wz)C(w,1,z)$$

while $G_n(w,1) = [z^n]C(w,1,z)$ satisfies

$$G_{n+1}(w,1) = \sum_{i=1}^{n} w^i G_i(w,1) G_{n-i}(w,1). \tag{1.4}$$

Observe that $G_n(w,1)$ is the generating function of the right path length. Actually, recurrence (1.4) was studied by Takacs in [26] when analyzing the area under a Bernoulli random walk. Also, it appears in the Kleitman-Winston conjecture [13, 29].

It turns out that there is a close link between the path length difference in trees and the center of mass S of the Integrated SuperBrownian Excursion (ISE), which was introduced by Aldous in [1]. The ISE is also related to the Brownian snake, which has various applications in branching processes, random maps and lattice trees (see the survey in [24]). Recent work on this problem, and its connection to the path difference in trees, is due to Marckert [20], Janson [10, 11], and Janson and Chassaing [12].

In particular moments of the path length difference were recently analyzed by Janson [11] using a Galton-Watson branching process approach, and the limiting distribution is implicit in Marckert [20], who applied Brownian analysis. As pointed out by Janson [11] "many different methods are useful and valuable, even for the same types of problems, and should be employed without prejudice".

Finally, we turn attention to results presented in this paper. We first analyze the limiting distribution of the difference $\mathcal{D}_n = \mathcal{R}_n - \mathcal{L}_n$ where $\mathcal{R}_n, \mathcal{L}_n$ are the right and left path lengths, respectively. We observe that the difference $\mathcal{D}_n$ is of order $n^{5/4}$. This was also recently observed by Marckert [20] and Janson [11]. Among other things, we show that the tail of the distribution is thicker than that of the Gaussian distribution.

Next, we analyze moments of $\mathcal{D}_n$. We first observe that odd moments vanish, while the *normalized* even moments satisfy (asymptotically) a certain *nonlinear* recurrence that occurs in various forms in many other problems, that are described by nonlinear functional equations similar to (1.1) (e.g., quicksort, linear hashing, enumeration of trees in $\mathcal{T}_t$). In these cases, usually the limiting distribution can be characterized only by moments. We conjecture that these problems constitute a new class of distributions determined by moments. More precisely, let Z be a (normalized) limiting distribution of such a process. Then for some $a_m \to \infty$ we have $\mathbf{E}[Z^m]/a_m = c_m$ such that in general c_m satisfies

$$c_{m+1} = \alpha_m + \beta_m c_m + \gamma_m \sum_{i=0}^{m} c_i c_{m-i} \tag{1.5}$$

with some initial conditions, and given α_m, β_m and γ_m. In our case (cf. also [11] and [12]) the even normalized moments of $\mathcal{D}_n$ converge as $\mathbf{E}[\mathcal{D}_n^{2m+2}]/n^{5(m+1)/2} \to (2m+2)!\sqrt{\pi}\Delta_m$ for any integer $m \ge 0$, where Δ_m satisfy the recurrence similar to (1.5) (cf. (2.23) and (2.24)

below). Similar recurrences appear in the quicksort [14], linear hashing [7], path length in binary trees [17, 19, 27, 28], area under Bernoulli walk [26], enumeration of trees with given path length [15], and many others [8, 18, 25].

Finally, we analyze the moment generating function of $\mathcal{D}_n$. We shall show that the Laplace transform of an appropriately normalized moment generating function satisfies the *first Painlevé transcendent* nonlinear differential equation [9]. This also appears in many modern applications, including nonlinear waves and random matrices. We shall also discuss the joint distribution of the left and the right path lengths, and this will be the starting point of our analysis.

Throughout, we use methods of analytic algorithmics such as generating functions and complex asymptotics, as well as methods of applied mathematics, such as the WKB method [4]. In this approach we make some assumptions about the forms of asymptotic expansions, for which we provide extensive numerical back-up.

2 Problem Statement and Summary of Results

Here we give a detailed summary of our main results; the technical derivations appear in [16].

Let $N(p,q;n)$ be the number of binary trees with n nodes that have a total right path length p and a total left path length q. We also set

$$N(p,q;n) = \bar{N}(p+q, p-q; n) \tag{2.1}$$

and note the obvious symmetry relation

$$N(p,q;n) = N(q,p;n). \tag{2.2}$$

We shall mostly focus on analyzing the difference between the right and left path length, and this we denote by

$$J = p - q. \tag{2.3}$$

It is well known [8] that the total number of trees with n nodes is the Catalan number

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{2.4}$$

Then we define the probability distribution of the path length difference, $\mathcal{D}_n$, by

$$\begin{aligned} P_-(J;n) &= \text{Prob}[\mathcal{D}_n = J] \\ &= \frac{1}{C_n}\sum_{i=0}^{\binom{n}{2}-|J|} N(i, i+|J|; n), \end{aligned} \tag{2.5}$$

for $J \in \left[-\binom{n}{2}, \binom{n}{2}\right]$. Here we used the fact that the left or right path in a tree with n nodes can be at most $\binom{n}{2}$. We can easily verify that

$$P_-\left(\binom{n}{2} - 1; n\right) = 0, \tag{2.6}$$

$$\begin{aligned} P_-\left(\binom{n}{2}; n\right) &= P_-\left(\binom{n}{2} - 2; n\right) \\ &= P_-\left(\binom{n}{2} - 3; n\right) = 1/C_n, \end{aligned} \tag{2.7}$$

since there are no trees where the path length difference is one below the maximum, and exactly one tree (out of C_n) has this difference either zero or two or three below the maximum value of $\binom{n}{2}$. In view of (2.2) we have $P_-(J;n) = P_-(-J;n)$ so it is sufficient to analyze (2.5) for $J \geq 0$.

The generating function of $N(p,q)$ in (2.1)

$$G_n(w,v) = \sum_p \sum_q N(p,q;n) w^p v^q \tag{2.8}$$

satisfies the recurrence

$$G_{n+1}(w,v) = \sum_{i=0}^{n} w^i v^{n-i} G_i(w,v) G_{n-i}(w,v), \ n \geq 0, \tag{2.9}$$

subject to the initial condition

$$G_0(w,v) = 1. \tag{2.10}$$

From (2.9) we also obtain the functional equation

$$C(w,v,z) = 1 + zC(w,v,wz)C(w,v,vz) \tag{2.11}$$

for the triple transform

$$C(w,v,z) = \sum_{n=0}^{\infty} G_n(w,v) z^n = \sum_n \sum_p \sum_q N(p,q;n) z^n w^p v^q. \tag{2.12}$$

We note that $G_n(1,1) = C_n$ and from (2.5) we obtain

$$P_-(J;n) = \frac{1}{C_n}[w^J] G_n\left(w, \frac{1}{w}\right). \tag{2.13}$$

We study the limit $n \to \infty$, with an appropriate scaling of p and q. First we consider the path length difference, with J scaled as

$$J = \beta n^{5/4} = O(n^{5/4}). \tag{2.14}$$

For a fixed β we shall obtain

$$P_-(J;n) \sim n^{-5/4} p_-(\beta) \tag{2.15}$$

where $p_-(\beta)$ is a probability density that can be represented as

$$\begin{aligned} p_-(\beta) &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{-\beta b}[1+\sqrt{\pi}\bar{H}(b)]db \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\beta i x}[1+\sqrt{\pi}\bar{H}(ix)]dx \\ &= \frac{1}{\pi}\int_0^{\infty} \cos(\beta x) S(x) dx, \end{aligned} \tag{2.16}$$

where $S(x) = 1 + \sqrt{\pi}\bar{H}(ix)$ and

$$1 + \sqrt{\pi}\bar{H}(b) = \int_{-\infty}^{\infty} e^{\beta b} p_-(\beta) d\beta. \tag{2.17}$$

Thus the left side of (2.17) is the moment generating function of $p_-(\beta)$, which is an *entire* function of b.

While we do not have an explicit formula for $p_-(\beta)$, we have the following asymptotic and numerical values:

$$p_-(\beta) \sim \sqrt{\frac{5}{6}}(5\beta)^{1/3} c_0 \exp\left(-\frac{3}{4}5^{1/3}\beta^{4/3}\right)[1+O(\beta^{-4/3})] \tag{2.18}$$

as $\beta \to \infty$, with

$$c_0 \approx .5513288; \tag{2.19}$$

$$p_-(0) \approx .45727\ ;\ p_-''(0) \approx -.71462. \tag{2.20}$$

We also find that the density has two inflection points, at $\beta = \pm\beta_c$, with

$$\beta_c \approx .75898. \tag{2.21}$$

This is our first main result. We comment that tails with exponent 4/3 are also observed for problems dealing with the Brownian snake (see [5, 12, 21]).

The function $p_-(\beta)$ is graphed in Figure 1 over the range $\beta \in [0,3]$, and the derivatives $p_-'(\beta)$ and $p_-''(\beta)$ are given over the same range in Figure 2. The graph of $p_-(\beta)$ somewhat resembles a Gaussian, but it differs from the Gaussian density in at least three important respects. First, the tail is clearly thicker in view of (2.18). For any Gaussian density, we would have $\beta_c p_-(0) = 1/\sqrt{2\pi} = .39894\ldots$ while for the present density (2.20) and (2.21) yield the value $\beta_c p_-(0) \approx .34705$. Finally, the Gaussian density would be an entire function of β, while we will show that if we view $p_-(\beta)$

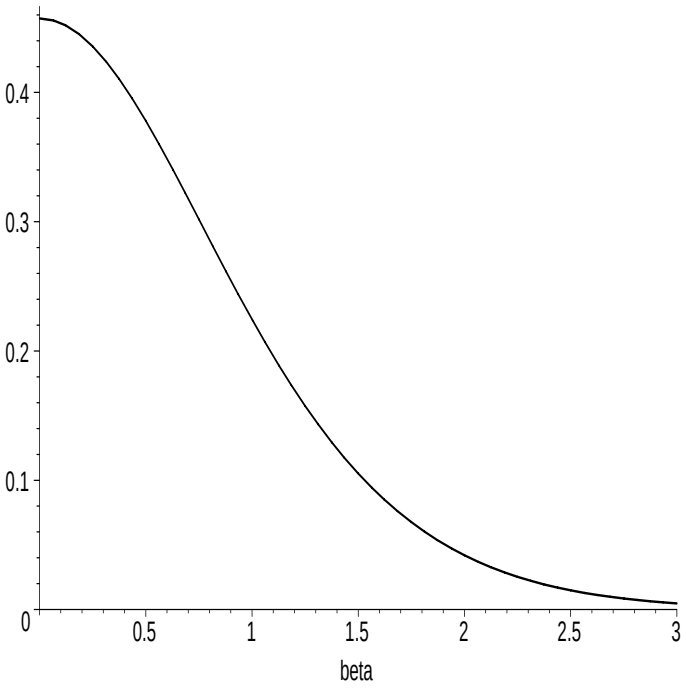


Figure 1: The density $p_-(\beta)$ for $\beta \in [0,3]$.

as a function of the complex variable β, this function has an essential singularity at $\beta = 0$. While for $|\beta|$ small and β real, $p_-(\beta)$ is locally Gaussian, its behavior for $|\beta|$ small and β imaginary is quite different.

The function $\bar{H}$ in (2.16) is an entire function satisfying $\bar{H}(b) = \bar{H}(-b)$ and $\bar{H}(0) = 0$. Denoting its Taylor series as

$$\bar{H}(b) = \sum_{m=0}^{\infty} b^{2m+2}\Delta_m \tag{2.22}$$

and setting

$$\Delta_m = \frac{1}{\Gamma\left(\frac{5}{2}m+2\right)}\tilde{\Delta}_m \tag{2.23}$$

we find that $\tilde{\Delta}_m$ satisfies the nonlinear recurrence

$$\tilde{\Delta}_{m+1} = \frac{(5m+6)(5m+4)}{8}\tilde{\Delta}_m + \frac{1}{4}\sum_{\ell=0}^{m}\tilde{\Delta}_\ell\tilde{\Delta}_{m-\ell},\ m \geq 0 \tag{2.24}$$

with

$$\Delta_0 = \tilde{\Delta}_0 = \frac{1}{4}. \tag{2.25}$$

This recurrence agrees with results in [12]. In view of (2.22) and (2.17) the variance of the limiting density $p_-(\beta)$ in (2.15) is

$$\int_{-\infty}^{\infty}\beta^2 p_-(\beta)d\beta = 2\sqrt{\pi}\Delta_0 = \frac{\sqrt{\pi}}{2}. \tag{2.26}$$

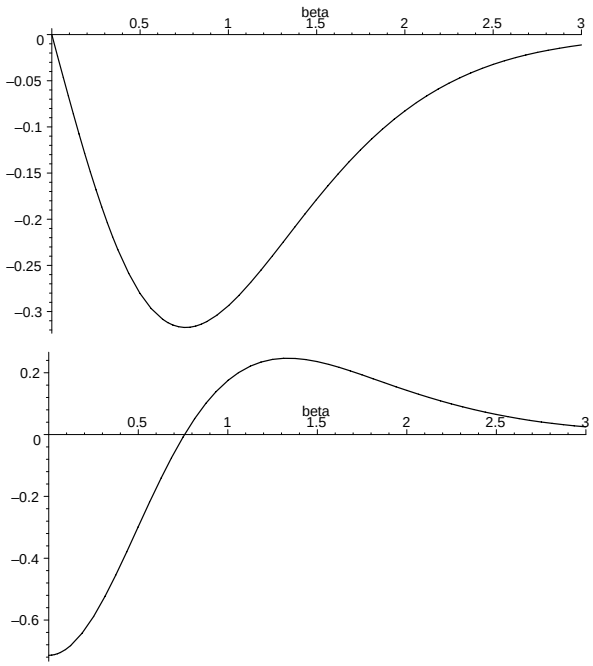


Figure 2: The first derivative $p'_-(\beta)$ and the second derivative $p''_-(\beta)$.

We thus observe that the even moments of the difference converge as follows

$$\frac{\mathbf{E}[\mathcal{D}_n^{2m+2}]}{n^{5(m+1)/2}} \to (2m+2)!\sqrt{\pi}\Delta_m.$$

These results agree with Janson [11].

Furthermore, setting

$$\bar{H}(b) = b^{6/5}\Delta(b^{4/5}) = B^{3/2}\Delta(B), \quad b = B^{5/4} \tag{2.27}$$

we shall show that for $b,\ B > 0$ the function $\Delta(B)$ satisfies the nonlinear integral equation

$$\begin{aligned} 0 &= \int_0^B \Delta(\xi)\Delta(B-\xi)d\xi + 2B^2\Delta(B) \\ &\quad +2\frac{\sqrt{B}}{\sqrt{\pi}} - \frac{4}{\sqrt{\pi}}\int_0^B \frac{\Delta'(\xi)}{\sqrt{B-\xi}}d\xi. \end{aligned} \tag{2.28}$$

We also have $\Delta(B) \sim B/4$ as $B \to 0^+$ and, in view of (2.22),

$$\Delta(B) = \sum_{m=0}^{\infty} B^{1+\frac{5}{2}m}\Delta_m. \tag{2.29}$$

The following asymptotic properties hold:

$$\begin{aligned} \Delta_m &= k'\frac{e^{m/2}}{\sqrt{m}}m^{-m/2}10^{-m/2}\left[1 - \frac{4}{15m} + O(m^{-2})\right],\ m \to \infty, \\ \Delta(B) &= c_0 B e^{B^5/20}[1 - B^{-5} + O(B^{-10})],\ B \to \infty, \\ \bar{H}(b) &= c_0 b^2 e^{b^4/20}[1 - b^{-4} + O(b^{-8})],\ b \to \infty. \end{aligned}$$

Here k' and c_0 are related by

$$c_0 = 2\sqrt{\pi}k', \tag{2.30}$$

so that the numerical value of k' can be obtained from (2.19).

We can also infer the behavior of $\bar{H}(b)$ for purely imaginary values of b. Letting $b = ix$ with $x > 0$ and using (2.22) yields

$$1+\sqrt{\pi}\bar{H}(ix) \equiv S(x) = \sum_{m=0}^{\infty} \Delta_{m-1}(-1)^m\sqrt{\pi}x^{2m} \tag{2.31}$$

where

$$\Delta_{-1} \equiv \frac{1}{\sqrt{\pi}}. \tag{2.32}$$

Then if

$$\bar{H}(ix) = -y^{3/2}\Lambda(y) = -x^{6/5}\Lambda(x^{4/5}), \quad y = x^{4/5} \tag{2.33}$$

we find that $\Lambda(y)$ satisfies

$$0 = \int_0^y \Lambda(\xi)\Lambda(y-\xi)d\xi + 2y^2\Lambda(y) - 2\frac{\sqrt{y}}{\sqrt{\pi}} + \frac{4}{\sqrt{\pi}}\int_0^y \frac{\Delta'(\xi)}{\sqrt{y-\xi}}d\xi. \tag{2.34}$$

Note that this equation differs from (2.28) only slightly, by the signs of the last two terms. However, (2.34) can be analyzed by a Laplace transform whereas (2.28) cannot. Indeed, setting

$$U(\phi) = \int_0^\infty e^{-y\phi}\Lambda(y)dy \tag{2.35}$$

we obtain from (2.34)

$$0 = 2U''(\phi) + U^2(\phi) + 4\sqrt{\phi}U(\phi) - \phi^{-3/2}. \tag{2.36}$$

Also, since we know the behavior of $\Lambda(y)$ as $y \to 0^+$ we must have

$$U(\phi) \sim \frac{1}{4\phi^2},\ \phi \to +\infty. \tag{2.37}$$

Setting

$$U(\phi) = -2\sqrt{\phi} + U_1(\phi) \tag{2.38}$$

we obtain from (2.36) and (2.37)

$$0 = U_1^2(\phi) + 2U_1''(\phi) - 4\phi, \tag{2.39}$$

with

$$U_1(\phi) = 2\sqrt{\phi} + \frac{1}{4}\phi^{-2}[1 + o(1)], \qquad \text{for} \quad \phi \to \infty.$$

The second order nonlinear ODE in (2.39) is (after a slight rescaling) the *first Painlevè transcendent* [9]. This classic problem has been studied for over 100 years, and modern applications in nonlinear waves and random matrices have been found in recent years. It is well known that each singularity of $U_1(\phi)$ is a double pole, and the Laurent expansion near any singularity at $\phi = -\nu$ has the form

$$U_1(\phi) = \frac{-12}{(\phi + \nu)^2} + O(1), \ \phi \to -\nu. \tag{2.40}$$

Let us denote by ν_* the singularity with the largest real part. Note that to uniquely fix this we need the second term in the expansion of $U_1(\phi)$ as $\phi \to \infty$, as given below (2.39). In view of (2.40), (2.35), and (2.38) we then obtain

$$\Lambda(y) - \frac{1}{\sqrt{\pi}} y^{-3/2} \sim -12 y e^{-\nu_* y}, \ y \to \infty \tag{2.41}$$

and (2.33) then yields

$$\sqrt{\pi}\bar{H}(ix) + 1 = S(x) \sim 12\sqrt{\pi} x^2 \exp(-\nu_* x^{4/5}), \ x \to \infty. \tag{2.42}$$

This yields the behavior of the moment generating function of the density $p_-(\beta)$ along the imaginary axis. The constant ν_* is found numerically as

$$\nu_* = 3.41167\ldots\ . \tag{2.43}$$

Finally, we discuss the joint distribution for the total left and right path lengths. This problem we formulate, but do not analyze. Introducing the scaling

$$p + q = \alpha n^{3/2}, \ p - q = \beta n^{5/4} \tag{2.44}$$

we obtain

$$\begin{aligned} P(p,q;n) &= \\ &= \text{Prob [right path} = p, \text{ left path} = q \mid \#\text{ nodes} = n] \\ &= N(p,q;n)/C_n \\ &\sim 2n^{-11/4} \frac{1}{(2\pi i)^2} \int_{-i\infty}^{i\infty} \int_{-i\infty}^{i\infty} e^{-a\alpha} e^{-b\beta} [1 + \sqrt{\pi} H(a,b)] da \\ &\equiv 2n^{-11/4} p(\alpha, \beta). \end{aligned}$$

Thus

$$1 + \sqrt{\pi} H(a,b) = \int_0^\infty e^{a\alpha} \int_{-\infty}^\infty e^{b\beta} p(\alpha,\beta) d\beta d\alpha \tag{2.45}$$

is the moment generating function of the two-dimensional density, which has support over the range $\alpha \geq 0$ and $\beta \in \mathbb{R}$. We have $H(0,0) = 0$ and $p(\alpha, -\beta) = p(\alpha, \beta)$.

Setting $\alpha = 0$ with $\bar{H}(b) = H(0,b)$ we obtain the marginal distribution of the path length difference, with

$$p_-(\beta) = \int_0^\infty p(\alpha,\beta) d\alpha.$$

The marginal distribution of the total path length (without distinguishing between right and left paths) is given by

$$p_+(\alpha) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} [1 + \sqrt{\pi} H(a,0)] e^{-a\alpha} da$$

so that

$$1 + \sqrt{\pi} H(a,0) = \int_0^\infty e^{a\alpha} p_+(\alpha) d\alpha.$$

This has been previously shown to follow an Airy distribution [15].

The function $H(a,b)$ satisfies the integral integration

$$\begin{aligned} 0 &= (2\sqrt{\pi}) \int_0^1 \frac{H(x^{3/2}a, x^{5/4}b) H((1-x)^{3/2}a, (1-x)^{5/4}b)}{[x(1-x)]^{3/2}} dx \\ &+ \frac{2}{\sqrt{\pi}} \int_0^1 \left\{ \frac{H((1-x)^{3/2}a, (1-x)^{5/4}b)}{(1-x)^{3/2}} - H(a,b) \right\} \frac{dx}{x^{3/2}} \\ &- \frac{4}{\sqrt{\pi}} H(a,b) + (4a + 2b^2) \left[\frac{1}{\sqrt{\pi}} + H(a,b) \right]. \end{aligned}$$

In terms of the generating function in (2.8) the scaling (2.44) translates to

$$w = 1 + \frac{b}{n^{5/4}} + \frac{a}{n^{3/2}}, \ v = 1 - \frac{b}{n^{5/4}} + \frac{a}{n^{3/2}} \tag{2.47}$$

and then, for fixed a and b and $n \to \infty$,

$$G_n(w,v) \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}[1+\sqrt{\pi}H(a,b)]. \tag{2.48}$$

Here we used the asymptotic behavior of the Catalan numbers C_n. We have thus identified the scaling (2.44) and the problem (2.46) that must be analyzed to obtain the joint distribution of left and right paths in binary trees with large numbers of nodes n. While it does not seem feasible to solve (2.46) exactly, we believe that an asymptotic analysis for a and/or b large should be possible, and from this one can obtain asymptotic properties of the joint density $p(\alpha,\beta)$ for α and/or $|\beta|$ large.

References

[1] D. Aldous, Tree-based Models for Random Distribution of Mass, *Journal of Statistical Physics*, 73, 625–641, 1993.

[2] M. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1964.

[3] G. Andrews, R. Askey and R. Roy, *Special Functions*, Cambridge University Press, 1999.

[4] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, Mc-Graw Hill, 1978.

[5] A. Dembo and O. Zeitouni, Large Deviations for Random Distribution of Mass, *Random Discrete Structures (Minneapolis, MN, 1993)*, 45–53, IMA Vol. Math. Appl., 76, Springer, New York, 1996.

[6] P. Flajolet, and A. Odlyzko, The Average Height of Binary Trees and Other Simple Trees, *J. Computer and System Sciences*, 25, 171–213, 1982.

[7] P. Flajolet, P. Poblete, and A. Viola, On the Analysis of Linear Probing Hashing, *Algorithmica*, 22, 490–515, 1998.

[8] P. Flajolet and R. Sedgewick, *Analytical Combinatorics*, in preparation.

[9] E. L. Ince, *Ordinary Differential Equations*, Dover, New York, 1956.

[10] S. Janson, The Wiener Index of Simply Generated Random Trees, *Random Structures and Algorithms*, 22, 337–358, 2003.

[11] S. Janson, Left and Right Pathlengths in Random Binary Trees, preprint, 2004.

[12] S. Janson and P. Chassaing, The Center of Mass of the ISE and the Wiener Index of Trees, *Electronic Comm. Probab.*, 9, 2004.

[13] J. Kim and B. Pittel, Confirming the Kleitman–Watson Conjecture on the Largest Coefficient in a q-Catalan Number, *J. Comb. Theory*, Ser. A, 92, 197-206, 2000.

[14] C. Knessl and W. Szpankowski, Quicksort algorithm again revisited, *Discrete Math. Theor. Comput. Sci.*, 3 43–64, 1999.

[15] C. Knessl and W. Szpankowski, Enumeration of Binary Trees, Lempel-Ziv'78 Parsings, and Universal Types, *Proc. of the Second Workshop on Analytic Algorithmics and Combinatorics*, (ANALCO04), Vancouver, 2005.

[16] C. Knessl and W. Szpankowski, On the Joint Path Length Distribution in Random Binary Trees, preprint, 2005.

[17] D. E. Knuth, *Selected Papers on the Analysis of Algorithms*, Cambridge University Press, Cambridge, 2000.

[18] D. E. Knuth, *The Art of Computer Programming. Fundamental Algorithms. Combinatorial Algorithms*, Vol. 4, 2004; see `http://www-cs-faculty.stanford.edu/knuth`.

[19] G. Louchard, The Brownian Excursion Area: A Numerical Analysis, *Comp. & Maths. with Appls.*, 10, 413-417, 1984.

[20] J-F. Marckert, The Rotation Correspondence is Asymptotically a Dilatation, *Random Structures & Algorithms*, 24, 118-132, 2004.

[21] L. Serlet, A Large Deviation Principle for the Brownian Snake, *Stochastic Processes and Applications*, 67, 101-115, 1997.

[22] G. Seroussi, On Universal Types, *Proc. ISIT 2004*, pp. 223, Chicago, 2004.

[23] G. Seroussi, "On the Number of t-ary Trees with a Given Pathlength", preprint.

[24] G. Slade, Scaling Limits and Super-Brownian Motion, *Notices of the AMS*, 49, 1056-1067, 2002.

[25] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, John Wiley & Sons, New York, 2001.

[26] L. Takács, A Bernoulli Excursion and its Various Applications, *J. Appl. Probab.*, 23, 557-585, 1991.

[27] L. Takács, Conditional Limit Theorems for Branching Processes, *J. Applied Mathematics and Stochastic Analysis*, 4, 263-292, 1991.

[28] L. Takács, The Asymptotic Distribution of the Total Heights of Random Rooted Trees, *Acta Sci. Math. (Szegad)*, 57, 613-625, 1993.

[29] K. Watson and D. Kleitman, On the Asymptotic Number of Tournament Score Sequences, *J. Comb. Theory*, Ser. A, 35, 208-230, 1983.