LEFT AND RIGHT PATHLENGHTS IN RANDOM BINARY TREES

SVANTE JANSON

Abstract. We study the difference between the left and right total pathlengths in a random binary tree. The results include exact and asymptotic formulas for moments and an asymptotic distribution that can be expressed in terms of either the Brownian snake or ISE. The proofs are based on computing expectations for a subcritical binary Galton–Watson tree, and studying asymptotics as the Galton–Watson process approaches a critical one.

1. Introduction

A *binary tree* consists of either only a root or a root with two subtrees, one left and one right, both being binary trees. Each node thus has 0 or 2 children; nodes without children (i.e. leaves) are called *external* and nodes with 2 children are called *internal*. A binary tree with $n \geq 0$ internal nodes has $n+1$ external nodes and thus $2n+1$ nodes in total. (See e.g. [12, Section 2.3], but note that there and in many other papers, including several of the present author, binary trees are defined slightly differently, corresponding to the internal nodes only in the definition used in this paper; cf. [12, Exercise 2.3.20]. The trees considered in this paper are sometimes called *full binary trees.*)

The depth of a node is the length of the path from the root to it, and the *total pathlength* $P(T)$ of a tree T is the sum of the depths of all nodes. Summing over internal or external nodes only, we get the *internal pathlength* $P_i(T)$ and the *external pathlength* $P_e(T)$. Clearly,

$$P(T) = P_i(T) + P_e(T).$$

It is further well-known that, for a binary tree with n internal nodes,

$$P(T) = 2P_i(T) + 2n, \tag{1.1}$$

because every internal node has exactly two children, and thus

$$P_e(T) = P_i(T) + 2n. \tag{1.2}$$

The quantities P_i, P_e and P are thus equivalent when the number of nodes is fixed.

In applications of binary trees, left and right links often have different meanings. A typical example is the correspondence between binary trees and ordered trees [12, Section 2.3.2], where a left link corresponds to a child

Date: December 2, 2004; revised September 22, 2005.

in the ordered tree but a right link corresponds to a sibling. It is therefore of interest to study also the *left depth* and *right depth* of a node, defined as the number of left [right] links in the path from the root to the node; summing over all nodes we obtain the *left total pathlength* L and *right total pathlength* R. We similarly define the left and right internal and external pathlengths L_i, R_i, L_e, R_e by summing over internal or external nodes only. The simple argument yielding (1.2) shows that if T has n internal nodes, then

$$L_e(T) = L_i(T) + n \qquad\qquad R_e(T) = R_i(T) + n,$$

so again it is equivalent to study any of these versions.

We define the imbalance $D(T)$ by

$$D(T) := L_e(T) - R_e(T) = L_i(T) - R_i(T) = \frac{1}{2}\big(L(T) - R(T)\big). \tag{1.3}$$

Let T_n denote a random binary tree with n internal nodes (and thus $2n+1$ nodes in total); as usual with the uniform distribution over all such trees. It is well-known that the pathlengths $P(T_n)$, $P_i(T_n)$, $P_e(T_n)$ are of order $n^{3/2}$, more precisely, by Aldous [1], [2] and (1.2), (1.1),

$$n^{-3/2}P(T_n) \xrightarrow{\mathrm{d}} 2^{3/2}\xi, \quad n^{-3/2}P_i(T_n) \xrightarrow{\mathrm{d}} 2^{1/2}\xi, \quad n^{-3/2}P_e(T_n) \xrightarrow{\mathrm{d}} 2^{1/2}\xi,$$

where the random variable ξ is twice the area $\int_0^1 B^{\mathrm{ex}}(t)\,dt$ under a normalized Brownian excursion B^{ex}. (This is sometimes called the Airy distribution, see e.g. [8] for much more information.)

It has been shown by Marckert [15] that the imbalance $D(T_n)$ is of the smaller order $n^{5/4}$. More precisely, it is implicit in [15, Theorem 5] that $n^{-5/4}D(T_n) \xrightarrow{\mathrm{d}} 2^{1/4}S$, where S can be described as the center of mass of the integrated superbrownian excursion defined by Aldous [3], or, equivalently, as the integral of the head of the Brownian snake; see Section 4 for definitions and details. An immediate corollary is that $n^{-3/2}L(T_n) \xrightarrow{\mathrm{d}} 2^{1/2}\xi$, and similarly for L_i, L_e, and R, R_i, R_e.

The purpose of the present paper is to present an alternative approach, where we study the moments of $D(T_n)$. Our method is based on studying expectations for a subcritical Galton–Watson tree. More precisely, we consider the family tree of a binary Galton–Watson process with expected number of offspring $1-\delta$, and study this as a function of $\delta > 0$. This is algebraically equivalent to the study of generating functions, but the version here seems to be convenient in order to obtain recursion formulas. Our method yields both exact and asymptotic formulas for moments of $D(T_n)$. (The exact formulas are obtained recursively, and quickly become complicated. The asymptotic formulas are obtained by studing asymptotics as $\delta \to 0$, i.e. when the process approaches the critical binary Galton–Watson process.) See Section 2 for statements and Section 3 for proofs. The asymptotic formulas give, by the method of moments, another proof that $n^{-5/4}D(T_n)$ converges in distribution; the limit variable is now characterized by its moments which are given by a recursion formula. A comparison with [6], where the moments of S above are computed, shows that this limit indeed equals $2^{1/4}S$, thus yielding another proof of Marckert's result.

In related work, the difference between left and right depths of individual nodes has been studied from a different point of view by Bousquet-Mélou [5]. She considers the generating function of the number of nodes with a given difference and obtains limit results on the distribution of this number.

In another related work, Jim Fill [7] has found a limiting distribution for the imbalance $D(T)$ for binary search trees.

Personal remark. I often work with Brownian limits of discrete structures and I have sometimes given new "Brownian" proofs of results first proved using generating functions. This time I do the opposite. I regard this as an illustration of the fact that many different methods are useful and valuable, even for the same type of problems, and should be employed without prejudice.

Acknowledgement. This research was mainly done during the workshop *Analysis of Algorithms* at MSRI, Berkeley, CA, USA, in June 2004, where Donald Knuth asked about properties of the left and right pathlengths of binary trees. (His question was about properties of their joint generating function, so the present work represents only an indirect and partial answer.)

I further thank Mireille Bousquet-Mélou and Jean-François Marckert for helpful discussions.

2. Results

We begin with exact formulas for the second and fourth moments of $D(T_n)$. Note that all odd moments $\mathbb{E}\, D(T_n)^{2k+1}$ vanish by symmetry. Proofs are given in Section 3.

Theorem 2.1. *Let*

$$\gamma_n := \sqrt{\pi}\frac{\Gamma(n+1)}{\Gamma(n+1/2)} = \frac{2^n n!}{(2n-1)!!} = 2^{2n}\binom{2n}{n}^{-1}.$$

Then

$$\begin{aligned}
\mathbb{E}\, D(T_n)^2 &= \frac{\Gamma(1/2)\Gamma(n+3)}{2\,\Gamma(n+1/2)} - (2n+1)(n+1) \\
&= \frac{(n+1)(n+2)}{2}\,\gamma_n - (2n+1)(n+1); \\
\mathbb{E}\, D(T_n)^4 &= \frac{14}{5}\,n^5 + \frac{171}{5}\,n^4 + \frac{468}{5}\,n^3 + \frac{539}{5}\,n^2 + \frac{283}{5}\,n + 11 \\
&\qquad - \frac{\sqrt{\pi}\left(121\,n^2 + 199\,n + 88\right)\Gamma\left(n+3\right)}{16\,\Gamma\left(n+1/2\right)} \\
&= \frac{\left(2\,n+1\right)\left(n+1\right)\left(7\,n^3 + 75\,n^2 + 118\,n + 55\right)}{5} \\
&\qquad - \frac{\left(n+1\right)\left(n+2\right)\left(121\,n^2 + 199\,n + 88\right)}{16}\,\gamma_n.
\end{aligned}$$

As $n \to \infty$, $\gamma_n \sim \sqrt{\pi} n^{1/2}$, and thus $\mathbb{E}\, D(T_n)^2 \sim \frac{\sqrt{\pi}}{2} n^{5/2}$ and $\mathbb{E}\, D(T_n)^4 \sim \frac{14}{5} n^5$.

We can recursively find exact expressions for higher moments too by the same method, see Section 3, but no simple or general formula is apparent, and we see no reason to give further exact formulas here. The asymptotical results generalize more easily.

Theorem 2.2. *The odd moments $\mathbb{E}\, D(T_n)^{2k+1}$, $k \geq 0$, vanish, while the even moments have asymptotics, as $n \to \infty$,*

$$\mathbb{E}\, D(T_n)^{2k} \sim \frac{2\sqrt{\pi}}{\Gamma\big((5k-1)/2\big)} c_k n^{5k/2},$$

where c_k is defined recursively by $c_1 = \frac{1}{4}$ and, for $k \geq 2$,

$$c_k = \frac{2k(2k-1)}{8}(5k-6)(5k-4)c_{k-1} + \frac{1}{2}\sum_{i=1}^{k-1}\binom{2k}{2i} c_i c_{k-i}. \tag{2.1}$$

Corollary 2.3. *As $n \to \infty$, $n^{-5/4} D(T_n) \xrightarrow{\mathrm{d}} Z$, where Z is a symmetric random variable Z with moments $\mathbb{E}\, Z^{2k+1} = 0$, $k \geq 0$, and*

$$\mathbb{E}\, Z^{2k} = \frac{2\sqrt{\pi}}{\Gamma\big((5k-1)/2\big)}\, c_k, \qquad k \geq 1, \tag{2.2}$$

with c_k as in Theorem 2.2.

The random variable S described in the introduction and in Section 4 was studied in [6], where it was shown that

$$\mathbb{E}\, S^{2k} = \frac{(2k)!\, \sqrt{\pi}}{2^{(9k-4)/2}\Gamma\big((5k-1)/2\big)}\, a_k, \qquad k \geq 1, \tag{2.3}$$

where $a_1 = 1$, and, for $k \geq 2$,

$$a_k = 2(5k-4)(5k-6)a_{k-1} + \sum_{i=1}^{k-1} a_i a_{k-i}. \tag{2.4}$$

(Note that a_k equals ω^*_{0k} in [9, Theorem 3.3], see [6].)

By comparing (2.1) and (2.4), we see that $c_k = 2^{1-4k}(2k)!\, a_k$. It now follows from Corollary 2.3 and (2.3) that $\mathbb{E}\, Z^{2k} = 2^{k/2}\, \mathbb{E}\, S^{2k}$, $k \geq 1$. Consequently:

Corollary 2.4. *The limit Z in Corollary 2.3 equals $2^{1/4} S$.*

Remark 2.5. The moments of Z or S are thus described by the quadratic recurrence (2.1) or (2.4). It is well-known that the moments of the Brownian excursion area (Airy distribution) $\xi/2$ can be described by a similar formula and a similar quadratic recurrence, see e.g. [9, Theorem 3.3]; in that case there is also a linear recurrence, see e.g. [10, Section 8]; the c_r there are related to the moments of ξ by $\mathbb{E}\, \xi^k = 2^{1-k/2} 3\pi^{1/2} k!\, (k-1)c_{k-1}/\Gamma((3k-1)/2)$. We do not know any similar linear recurrence for the moments of S.

3. Proofs

We will use the method of [9, Section 5] and consider the random binary Galton–Watson tree T_δ defined by starting with the root and, recursively, letting each node be a leaf with probability $(1+\delta)/2$ and an internal node with two new nodes as offspring with probability $(1-\delta)/2$, where δ is a fixed number with $0 < \delta < 1$. We let $\mathbb{P}_\delta$ and $\mathbb{E}_\delta$ denote probability and expectation for this subcritical Galton–Watson tree T_δ. Note that the tree T_δ is finite a.s.

We begin with some general results on the expectations $\mathbb{E}_\delta$ as functions of δ; some of these results are stated in [9] and repeated here for completeness. We consider an arbitrary functional $Z = Z(T)$ of binary trees, and first observe that studying the expectations $\mathbb{E}_\delta Z$ can be seen as studying a generating function; note that the left hand side of (3.2) is the generating function $\sum Z(T)x^{N_i(T)}$, summing over all binary trees T, where N_i is the number of internal nodes.

Lemma 3.1. *Let b_n be the number of binary trees with n internal nodes. Assume either that $\mathbb{E}_\delta |Z| < \infty$ for $0 < \delta < 1$, or that $Z \geq 0$, and let $z_n := \mathbb{E}\, Z(T_n)$. Then*

$$\mathbb{E}_\delta Z = \frac{1+\delta}{2} \sum_{n=0}^{\infty} b_n z_n \Bigl(\frac{1-\delta^2}{4}\Bigr)^n. \tag{3.1}$$

Equivalently, if $0 < x < 1/4$ and $\delta = \sqrt{1-4x}$, then

$$\sum_{n=0}^{\infty} b_n z_n x^n = \frac{2}{1+\delta}\, \mathbb{E}_\delta Z. \tag{3.2}$$

Proof. If t is a binary tree with n internal nodes, and thus $n+1$ external nodes, then

$$\mathbb{P}(T_\delta = t) = \Bigl(\frac{1-\delta}{2}\Bigr)^n \Bigl(\frac{1+\delta}{2}\Bigr)^{n+1} = \frac{1+\delta}{2}\Bigl(\frac{1-\delta^2}{4}\Bigr)^n.$$

Consequently, if $\mathcal{B}_n$ is the set of the b_n binary trees with n internal nodes, then

$$\mathbb{E}_\delta Z = \sum_{n=0}^{\infty} \sum_{t \in \mathcal{B}_n} \mathbb{P}(T_\delta = t) Z(t) = \frac{1+\delta}{2} \sum_{n=0}^{\infty} \Bigl(\frac{1-\delta^2}{4}\Bigr)^n \sum_{t \in \mathcal{B}_n} Z(t),$$

and (3.1) follows because $\sum_{t\in\mathcal{B}_n} Z(t) = b_n\, \mathbb{E}\, Z(T_n) = b_n z_n$. $\square$

In particular, taking $Z = 1$, we obtain by (3.2), with $\delta = \sqrt{1-4x}$,

$$\sum_{n=0}^{\infty} b_n x^n = \frac{2}{1+\delta}\, \mathbb{E}_\delta 1 = \frac{2}{1+\sqrt{1-4x}} = \frac{1-\sqrt{1-4x}}{2x}, \qquad 0 < x < 1/4,$$

which yields the well-known generating function for b_n [12, 2.3.4.4-(13)]. This gives another proof of the well-known fact that b_n is the Catalan number $\frac{1}{n+1}\binom{2n}{n}$ [17, Exercise 6.19d].

We let $N = N(T) := |T|$, the total number of nodes in a binary tree T; thus T has $(N-1)/2$ internal and $(N+1)/2$ external nodes. (This is in accordance with the notation in [9]; note however that T_n here is denoted T_{2n+1} there.) In particular, $N(T_n) = 2n+1$.

Lemma 3.2. *Assume that $\mathbb{E}_\delta |Z| < \infty$ for $0 < \delta < 1$. Then $\mathbb{E}_\delta |NZ| < \infty$ for $0 < \delta < 1$, and if $f(\delta) = \mathbb{E}_\delta Z$, then*

$$\mathbb{E}_\delta(NZ) = -\Bigl(\frac{1}{\delta} - \delta\Bigr) f'(\delta) + \frac{1}{\delta} f(\delta).$$

Proof. Since $N(T_n) = 2n+1$, $\mathbb{E}\bigl(NZ(T_n)\bigr) = (2n+1)z_n$. By Lemma 3.1, $\sum_{n=0}^\infty b_n z_n x^n$ is absolutely convergent for $|x| < 1/4$, and thus analytic there; it follows that $\sum_{n=0}^\infty (2n+1) b_n z_n x^n$ too is absolutely convergent for $|x| < 1/4$, and thus $\mathbb{E}_\delta |NZ| < \infty$ for $0 < \delta < 1$. Moreover, by (3.2) again,

$$\frac{2}{1+\delta}\,\mathbb{E}_\delta(NZ) = \sum_{n=0}^\infty (2n+1) b_n z_n x^n = \Bigl(2x\frac{d}{dx} + 1\Bigr)\Bigl(\frac{2}{1+\delta} f(\delta)\Bigr)$$

and the result follows by simple calculations because

$$2x\frac{d}{dx} = 2x\frac{d\delta}{dx}\frac{d}{d\delta} = \frac{-4x}{\sqrt{1-4x}}\frac{d}{d\delta} = -\frac{1-\delta^2}{\delta}\frac{d}{d\delta}.$$

∎

Lemma 3.2 shows, by induction, that $\mathbb{E}_\delta N^k < \infty$ for every $k \ge 0$. (This well-known fact can be shown in many other ways too.) Explicitly, we obtain from $\mathbb{E}_\delta N^0 = 1$ in the first two steps

$$\mathbb{E}_\delta N = \delta^{-1}, \qquad \mathbb{E}_\delta N^2 = \delta^{-3} + \delta^{-2} - \delta^{-1}. \tag{3.3}$$

In general, $\mathbb{E}_\delta N^k$ is a polynomial in δ^{-1} of degree $2k-1$ for $k \ge 1$.

Conversely, if $\mathbb{E}_\delta Z$ is such a polynomial we can find $\mathbb{E} Z(T_n)$.

Lemma 3.3. *Assume that $\mathbb{E}_\delta Z = \sum_{j=0}^m a_j \delta^{-j}$ for $0 < \delta < 1$. Then, with $1/\Gamma(0) = 0$ and $a_{m+1} = 0$,*

$$\begin{aligned}\mathbb{E} Z(T_n) &= \sum_{j=0}^m a_j \Bigl(\frac{2\Gamma(1/2)\Gamma(n+(j+2)/2)}{\Gamma(j/2)\Gamma(n+1/2)} - \frac{2\Gamma(1/2)\Gamma(n+(j+1)/2)}{\Gamma((j-1)/2)\Gamma(n+1/2)}\Bigr) \\ &= \sum_{j=1}^m \frac{2\Gamma(1/2)}{\Gamma(j/2)}(a_j - a_{j+1})\frac{\Gamma(n+(j+2)/2)}{\Gamma(n+1/2)} + a_0.\end{aligned} \tag{3.4}$$

In particular, if $m \ge 1$,

$$\mathbb{E} Z(T_n) = 2\frac{\Gamma(1/2)}{\Gamma(m/2)} a_m n^{(m+1)/2} + O\bigl(n^{m/2}\bigr) \qquad \text{as } n \to \infty.$$

Proof. Consider first the case $\mathbb{E}_\delta Z = \delta^{-j}$. In this case we have by Lemma 3.1, with $z_n := \mathbb{E} Z(T_n)$,

$$\sum_{n=0}^\infty 2^{-2n-1} b_n z_n (1-\delta^2)^{n+1} = (1-\delta)\,\mathbb{E}_\delta Z = \delta^{-j} - \delta^{-(j-1)}$$

and thus, taking $\delta = \sqrt{1-x}$,

$$\begin{aligned} 2^{-2n-1} b_n z_n &= [x^{n+1}]\bigl((1-x)^{-j/2} - (1-x)^{-(j-1)/2}\bigr) \\ &= \frac{\Gamma(j/2+n+1)}{\Gamma(j/2)(n+1)!} - \frac{\Gamma((j-1)/2+n+1)}{\Gamma((j-1)/2)(n+1)!}. \end{aligned} \tag{3.5}$$

In particular, the choice $Z = 1$ is of this type with $j = 0$; hence (3.5) yields, since $1/\Gamma(0) = 0$,

$$2^{-2n-1} b_n = -\frac{\Gamma(n+1/2)}{\Gamma(-1/2)(n+1)!} = \frac{\Gamma(n+1/2)}{2\Gamma(1/2)(n+1)!}. \tag{3.6}$$

(This is equivalent to the standard formula for Catalan numbers.) Dividing (3.5) by (3.6) we obtain

$$\mathbb{E}\, Z(T_n) = \frac{2\Gamma(1/2)\Gamma(n+(j+2)/2)}{\Gamma(j/2)\Gamma(n+1/2)} - \frac{2\Gamma(1/2)\Gamma(n+(j+1)/2)}{\Gamma((j-1)/2)\Gamma(n+1/2)},$$

which gives (3.4) in the case $\mathbb{E}_\delta Z = \delta^{-j}$. The general case follows by linearity. The asymptotic estimate now follows because $\Gamma(n+b)/\Gamma(n) \sim n^b$ as $n \to \infty$, for any fixed b. □

We now return to the imbalance $D(T)$. By definition, a binary tree T is either just a root, and then $D(T) = 0$, or it consists of a root with left and right subtrees T' and T''; in the latter case

$$\begin{aligned} L_i(T) &= L_i(T') + \tfrac{1}{2}\bigl(N(T') - 1\bigr) + L_i(T''), \\ R_i(T) &= R_i(T') + R_i(T'') + \tfrac{1}{2}\bigl(N(T'') - 1\bigr), \end{aligned}$$

and thus

$$D(T) = D(T') + D(T'') + \tfrac{1}{2}N(T') - \tfrac{1}{2}N(T'').$$

Now, let T be the random tree T_δ. Since the second case above occurs with probability $(1-\delta)/2$, and then T' and T'' are independent and have the same distribution as T, we have for every $k \geq 1$

$$\begin{aligned} \mathbb{E}_\delta D^k = \mathbb{E}_\delta D(T)^k &= \frac{1-\delta}{2}\, \mathbb{E}_\delta \Bigl(D(T') + \tfrac{1}{2}N(T') + D(T'') - \tfrac{1}{2}N(T'')\Bigr)^k \\ &= \frac{1-\delta}{2} \sum_{j=0}^{k} \binom{k}{j} \mathbb{E}_\delta \Bigl(D(T') + \tfrac{1}{2}N(T')\Bigr)^j \mathbb{E}_\delta \Bigl(D(T'') - \tfrac{1}{2}N(T'')\Bigr)^{k-j} \\ &= \frac{1-\delta}{2} \sum_{j=0}^{k} \binom{k}{j} \mathbb{E}_\delta \bigl(D + \tfrac{1}{2}N\bigr)^j \mathbb{E}_\delta \bigl(D - \tfrac{1}{2}N\bigr)^{k-j}. \end{aligned} \tag{3.7}$$

Note that all moments $\mathbb{E}_\delta D^k$ are finite because $|D(T)| \leq N(T)^2$.

Observing that all odd moments of $D(T_\delta)$ are 0 by symmetry, let us first consider $k=2$. Since $\mathbb{E}_\delta D = \mathbb{E}_\delta(DN) = 0$, (3.7) yields

$$\begin{aligned}
\mathbb{E}_\delta D^2 &= \frac{1-\delta}{2}\Bigl(\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)^2 + 2\,\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)\,\mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr) \\
&\qquad\qquad + \mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr)^2\Bigr) \\
&= \frac{1-\delta}{2}\Bigl(\mathbb{E}_\delta\bigl(D^2+\tfrac14 N^2\bigr) - \tfrac12\bigl(\mathbb{E}_\delta N\bigr)^2 + \mathbb{E}_\delta\bigl(D^2+\tfrac14 N^2\bigr)\Bigr) \\
&= (1-\delta)\,\mathbb{E}_\delta D^2 + \frac{1-\delta}{4}\bigl(\mathbb{E}_\delta N^2 - (\mathbb{E}_\delta N)^2\bigr)
\end{aligned}$$

and thus

$$\mathbb{E}_\delta D^2 = \frac{1-\delta}{4\delta}\bigl(\mathbb{E}_\delta N^2 - (\mathbb{E}_\delta N)^2\bigr). \tag{3.8}$$

Using (3.3), we find

$$\mathbb{E}_\delta D^2 = \frac{1-\delta}{4\delta}\bigl(\delta^{-3}-\delta^{-1}\bigr) = \tfrac14\delta^{-4} - \tfrac14\delta^{-3} - \tfrac14\delta^{-2} + \tfrac14\delta^{-1}. \tag{3.9}$$

The formula for $\mathbb{E}\, D(T_n)^2$ in Theorem 2.1 now follows from Lemma 3.3.

Similarly, one finds by taking $k=4$ in (3.7), using Lemma 3.2, that

$$\begin{aligned}
\mathbb{E}_\delta D^4 = \frac{147}{16}\,\delta^{-9} - \frac{27}{2}\,\delta^{-8} - \frac{57}{4}\,y^{7} + \frac{105}{4}\,\delta^{-6} + \frac{27}{8}\,\delta^{-5} - \frac{59}{4}\,\delta^{-4} \\
+ 2\,\delta^{-3} + 2\,\delta^{-2} - \frac{5}{16}\,\delta^{-1},
\end{aligned}$$

and the formula for $\mathbb{E}\, D(T_n)^4$ in Theorem 2.1 follows from Lemma 3.3. (We used `Maple` for these calculations.) The asymptotic formulas in Theorem 2.1 follow immediately, which completes the proof of the theorem.

We can recursively find exact expressions for $\mathbb{E}_\delta D^6 = \mathbb{E}\, D(T_\delta)^6$ and $\mathbb{E}\, D(T_n)^6$ and higher moments too by the same method, but as said in the introduction, we do not give these expressions here; they can easily be found by the reader and a computer. Instead we turn to asymptotics.

Lemma 3.4. *For each $k \ge 1$, $\mathbb{E}_\delta D^{2k}$ is a polynomial in δ^{-1} of degree $5k-1$ with leading term $c_k\delta^{-5k+1}$, where c_k is as in Theorem 2.2.*

Proof. We have already proved the case $k = 1$ in (3.9) and proceed by induction. We rewrite (3.7) as

$$\begin{aligned}\mathbb{E}_\delta D^{2k} &= \frac{1-\delta}{2}\Bigl(\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)^{2k} + \mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr)^{2k} \\ &\qquad\qquad + \sum_{j=1}^{2k-1}\binom{2k}{j}\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)^{j}\,\mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr)^{2k-j}\Bigr) \\ &= (1-\delta)\Bigl(\mathbb{E}_\delta D^{2k} + \sum_{j=1}^{k}\binom{2k}{2j}2^{-2j}\,\mathbb{E}_\delta\bigl(D^{2k-2j}N^{2j}\bigr) \\ &\qquad\qquad + \frac12\sum_{j=1}^{2k-1}\binom{2k}{j}\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)^{j}\,\mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr)^{2k-j}\Bigr)\end{aligned}$$

and thus

$$\begin{aligned}\delta\,\mathbb{E}_\delta D^{2k} &= (1-\delta)\Bigl(\sum_{j=1}^{k}\binom{2k}{2j}2^{-2j}\,\mathbb{E}_\delta\bigl(D^{2k-2j}N^{2j}\bigr) \\ &\qquad + \frac12\sum_{j=1}^{2k-1}\binom{2k}{j}\mathbb{E}_\delta\bigl(D+\tfrac12 N\bigr)^{j}\,\mathbb{E}_\delta\bigl(D-\tfrac12 N\bigr)^{2k-j}\Bigr). \qquad (3.10)\end{aligned}$$

Let $\mathcal{P}$ denote the set of polynomials in δ^{-1}, and note that Lemma 3.2 implies that if $\mathbb{E}_\delta Z \in \mathcal{P}$, then $\mathbb{E}_\delta(NZ) \in \mathcal{P}$, and thus $\mathbb{E}_\delta(N^i Z) \in \mathcal{P}$ for every $i \geq 0$. By induction, and binomial expansion of $\mathbb{E}_\delta\bigl(D \pm \frac12 N\bigr)^\nu$, we thus see that the right hand side of (3.10) is $(1-\delta)$ times a polynomial in $\mathcal{P}$, and thus $\mathbb{E}_\delta D^{2k} \in \mathcal{P}$.

To find the leading term, note that Lemma 3.2 further shows that if $\mathbb{E}_\delta Z$ is a polynomial of degree $m \geq 1$ and leading term $a_m\delta^{-m}$, then $\mathbb{E}_\delta(NZ)$ has degree $m+2$ and leading term $ma_m\delta^{-(m+2)}$. Thus, by induction, the degree of $\mathbb{E}_\delta\bigl(D^{2k-2j}N^{2j}\bigr)$ is $5(k-j)-1+4j = 5k-1-j$ for $1 \leq j < k$; using (3.3), this holds for $j = k \geq 1$ too. Similarly, a binomial expansion yields, for $1 \leq \nu < 2k$,

$$\mathbb{E}_\delta\bigl(D \pm \tfrac12 N\bigr)^\nu = \mathbb{E}_\delta D^\nu + O\bigl(\delta^{-(5\nu/2-3/2)}\bigr).$$

Hence, the right hand side of (3.10) is $1-\delta$ times

$$\begin{aligned}&\binom{2k}{2}2^{-2}\,\mathbb{E}_\delta\bigl(D^{2k-2}N^2\bigr) + \frac12\sum_{j=1}^{2k-1}\binom{2k}{j}\mathbb{E}_\delta D^j\,\mathbb{E}_\delta D^{2k-j} + O\bigl(\delta^{-(5k-3)}\bigr) \\ &\qquad = \frac{2k(2k-1)}{8}(5k-6)(5k-4)c_{k-1}\delta^{-(5k-2)} \\ &\qquad\qquad + \frac12\sum_{i=1}^{k-1}\binom{2k}{2i}c_i c_{k-i}\delta^{-(5k-2)} + O\bigl(\delta^{-(5k-3)}\bigr).\end{aligned}$$

The result follows by (3.10). ∎

Remark 3.5. The argument further shows that the polynomial $\mathbb{E}_\delta D^{2k}$ in δ^{-1} has rational coefficients and constant term 0.

Theorem 2.2 follows from Lemmas 3.4 and 3.3.

Finally, by Theorem 2.2, the moments of $n^{-5/4}D(T_n)$ converge to the values in (2.2) for even moments (and 0 for odd), which by the argument in Section 2 equal the moments of $2^{1/4}S$. Since the distribution of S is uniquely determined by its moments, as was shown in [6], this implies $n^{-5/4}D(T_n) \xrightarrow{\mathrm{d}} 2^{1/4}S$, proving both corollaries.

4. The Brownian connection

4.1. **The Brownian snake.**

We begin by recalling the definition of the Brownian snake, see [13, Chapter IV] or [14] for further details. In general, let ζ, the *lifetime*, be a non-negative stochastic process on some interval I, and let for $s, t \in I$

$$m(s,t;\zeta) := \min\{\zeta(u) : u \in [s,t]\}$$

when $s \le t$, and $m(s,t;\zeta) := m(t,s;\zeta)$ when $s > t$. The Brownian snake with lifetime ζ then can be defined as the stochastic process $W(s,t)$ on $I \times [0,\infty)$ such that, conditioned on ζ, W is Gaussian with mean 0 and covariances

$$\operatorname{Cov}\bigl(W(s_1,t_1)W(s_2,t_2) \mid \zeta\bigr) = \min\bigl(t_1,t_2,m(s_1,s_2;\zeta)\bigr).$$

It is easily verified that such a process exists; moreover, if ζ is (locally) Hölder continuous with some positive exponent α, then the Kolmogorov–Chentsov criterion [11, Theorem 3.23] shows that W has a continuous version, and thus we may and will assume W to be continuous. The following properties follow easily:

(i) $W(s,t) = W(s,\zeta(s))$ when $t \ge \zeta(s)$, so we only need to consider $t \le \zeta(s)$.
(ii) For fixed $s \in I$, $t \mapsto W(s,t)$ is conditioned on ζ a Brownian motion stopped at $\zeta(s)$.
(iii) $W(s_1,t) = W(s_2,t)$ when $0 \le t \le m(s_1,s_2;\zeta)$.
(iv) Let $s_1, s_2 \in I$. Conditioned on ζ, the two stopped Brownian motions $t \mapsto W(s_1,t)$ and $t \mapsto W(s_2,t)$ coincide up to $t = m(s_1,s_2;\zeta)$ and then evolve independently.

The stochastic process $\widehat{W}(s) := W(s,\zeta(s))$ is called the *head* of the Brownian snake. It is easily seen from (iii) that the pair $(\zeta,\widehat{W})$ determines W; see further [16]. Conditioned on ζ, $\widehat{W}$ is a Gaussian process on I with mean 0 and covariances $\mathbb{E}\bigl(\widehat{W}(s)\widehat{W}(t) \mid \zeta\bigr) = m(s,t;\zeta)$.

From now on, we will take the lifetime $\zeta = 2B^{\mathrm{ex}}$, a Brownian excursion on $I = [0,1]$. (In other contexts, the lifetime ζ is often taken to be reflected Brownian motion on $[0,\infty)$ [13].) The factor 2 is a normalization factor only, of no great importance.

The integrated superbrownian excursion μ_{ISE} introduced by Aldous [3] is a random probability measure, and can be defined as the occupation measure of the process $\widehat{W}$, see [13], [6], [14]. The random variable S mentioned in the introduction thus has the two equivalent representations [6]:

$$S := \int x \, d\mu_{\mathrm{ISE}} = \int_0^1 \widehat{W}(s) \, ds.$$

4.2. **Random trees.** If v is a node (internal or external) in a binary tree T with n internal nodes, let $\Delta(v)$ be the difference between the left and right depths of v and let $\widetilde{\Delta}(v) := \Delta(v)/(2n)^{1/4}$. Further, let $\widetilde{M}(T) := \max_v |\widetilde{\Delta}(v)|$ and $\tilde{\mu}_T := (2n+1)^{-1} \sum_v \delta(\widetilde{\Delta}(v))$, where $\delta(a)$ is the Dirac measure at a; thus $\tilde{\mu}_T$ is the probability measure giving the distribution of $\widetilde{\Delta}$ for a random node in T. Note that, by (1.3),

$$2D(T) = \sum_{v \in T} \Delta(v) = (2n)^{1/4} \sum_{v \in T} \widetilde{\Delta}(v) = (2n)^{1/4}(2n+1) \int x \, d\tilde{\mu}_T. \tag{4.1}$$

We consider the random tree T_n; thus $\widetilde{M}(T_n)$ is a random variable and $\tilde{\mu}_{T_n}$ is a random probability measure. Marckert [15, Theorem 5] has proved that these converge in distribution: $\widetilde{M}(T_n) \xrightarrow{\mathrm{d}} W^* := \max_s |\widehat{W}(s)| = \max_{s,t} |W(s,t)|$ and $\tilde{\mu}_{T_n} \xrightarrow{\mathrm{d}} \mu_{\mathrm{ISE}}$. As said in the introduction, these results implies easily that $n^{-5/4} D(T_n) \xrightarrow{\mathrm{d}} 2^{1/4} S$. For completeness, we give the details:

The fact that $\tilde{\mu}_{T_n} \xrightarrow{\mathrm{d}} \mu_{\mathrm{ISE}}$ implies by the continuous mapping theorem [11, Theorem 4.27] that if h is a bounded continuous function on $\mathbb{R}$, then

$$\int h \, d\tilde{\mu}_{T_n} \xrightarrow{\mathrm{d}} \int h \, d\mu_{\mathrm{ISE}}. \tag{4.2}$$

If h is unbounded, we can apply (4.2) to $h_N := (h \wedge N) \vee (-N)$, h truncated at $\pm N$, and then let $N \to \infty$. It is easily seen by [4, Proposition 4.2] and the fact that $\widetilde{M}(T_n)$ is tight (because $\widetilde{M}(T_n) \xrightarrow{\mathrm{d}} W^*$), that this yields that (4.2) holds for any continuous h.

Taking $h(x) = x$, (4.1) and (4.2) show that, as asserted,

$$n^{-5/4} D(T_n) = 2^{1/4}(1 + 1/2n) \int x \, d\tilde{\mu}_{T_n} \xrightarrow{\mathrm{d}} 2^{1/4} \int x \, d\mu_{\mathrm{ISE}} = 2^{1/4} S.$$

References

[1] D. Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Proc., Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[2] D. Aldous, The continuum random tree III. *Ann. Probab.* **21** (1993), no. 1, 248–289.

[3] D. Aldous, Tree-based models for random distribution of mass. *J. Statist. Phys.* **73** (1993), 625–641.

[4] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[5] M. Bousquet-Mélou, Limit laws for embedded trees. Applications to the integrated superBrownian excursion. arXiv:math.CO/0501266. To appear in *Random Struct. Alg.*

[6] P. Chassaing & S. Janson, The center of mass of the ISE and the Wiener index of trees. *Electronic Comm. Probab.* **9** (2004), paper 20, 178–187.

[7] J. Fill, in preparation. (Personal communication, 2004.)

[8] P. Flajolet & G. Louchard, Analytic Variations on the Airy Distribution. *Algorithmica* **31** (2001), 361–377.

[9] S. Janson, The Wiener index of simply generated random trees. *Random Struct. Alg.* **22** (2003), no. 4, 337–358.

[10] S. Janson, D.E. Knuth, T. Łuczak & B. Pittel, The birth of the giant component. *Random Struct. Alg.* **3** (1993), 233–358.

[11] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.

[12] D. E. Knuth, *The Art of Computer Programming. Vol. 1: Fundamental Algorithms.* 3rd ed., Addison-Wesley, Reading, Mass., 1997.

[13] J.-F. Le Gall, *Spatial Branching Processes, Random Snakes and Partial Differential Equations.* Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1999.

[14] J.-F. Le Gall & M. Weill, Conditioned Brownian trees. Preprint, 2004.

[15] J.-F. Marckert, The rotation correspondence is asymptotically a dilatation. *Random Struct. Alg.* **24** (2004), no. 2, 118–132.

[16] J.-F. Marckert and A. Mokkadem. States spaces of the snake and its tour—convergence of the discrete snake. *J. Theoret. Probab.* **16** (2003), no. 4, 1015–1046.

[17] R. Stanley, *Enumerative Combinatorics*, Volume 2, Cambridge Univ. Press, Cambridge, 1999.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`