

Left and right length of paths in binary trees or on a question of Knuth

Alois Panholzer^{1†}

¹ Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, A-1040 Wien, Austria. {Alois.Panholzer}@tuwien.ac.at

We consider extended binary trees and study the common right and left depth of leaf j , where the leaves are labelled from left to right by $0, 1, \dots, n$, and the common right and left external pathlength of binary trees of size n . Under the random tree model, i.e., the Catalan model, we characterize the common limiting distribution of the suitably scaled left depth and the difference between the right and the left depth of leaf j in a random size- n binary tree when $j \sim \rho n$ with $0 < \rho < 1$, as well as the common limiting distribution of the suitably scaled left external pathlength and the difference between the right and the left external pathlength of a random size- n binary tree.

Keywords: binary trees, imbalance, node depth

1 Introduction

Several recent papers are treating questions concerning the left and right length of paths from the root to nodes in binary trees under the so-called Catalan model, i.e., when all $B_n := \frac{1}{n+1} \binom{2n}{n}$ binary trees of size n are assumed to appear equally likely. In particular the difference between the right and the left depth of nodes and the difference between the right and the left pathlength (also called imbalance) of binary trees are examined. Here the right (left) depth of a certain node in a binary tree is the number of right (left) edges on the path from the root to the node considered, whereas the right (left) pathlength of a binary tree is the sum of the right (left) depths over all nodes in the tree. In Bousquet-Mélou (2005), amongst many others, the random variable $X_n(j)$ was studied, which counts the number of nodes in a random binary tree of size n with difference j between the right and the left depth. As one of the main results of Bousquet-Mélou (2005) the limiting distribution of $X_n(\lambda n^{\frac{1}{4}})$, $\lambda \geq 0$, was characterized and thus the so called “vertical profile” of binary trees well described. In Janson (2004) all moments $\mathbb{E}(D_n^r)$ of the difference D_n between the right and the left pathlength of random binary trees of size n are computed asymptotically and as a consequence the limiting distribution of the suitably scaled random variable $n^{-\frac{5}{4}} D_n$ was characterized. This result was also obtained in Knessl and Szpankowski (2006), but the main focus of that paper is lying in revealing (analytic) properties of the limiting distribution function of D_n . Furthermore in Chassaing and Janson (2004); Marckert (2004) close relations between the left-right difference of paths in binary trees and the so-called Integrated Superbrownian Excursion are given.

Here we give another contribution to this topic, where we are mainly interested in the behaviour of the *common* right and left depth of nodes and the *common* right and left pathlength of binary trees, not only their difference. We will restrict ourselves in presenting only results for the external nodes (or leaves) of so called extended binary trees (see Knuth (1997)), but we mention that similar results can also be obtained for the internal nodes of binary trees. In particular we treat in this paper questions about properties of the common right depth $H_{n,j}^{[R]}$ and left depth $H_{n,j}^{[L]}$ of the leaf labelled by j (not only a randomly chosen leaf) in a random size- n binary tree, where the $n+1$ leaves are assumed to be labelled from left to right with $0, 1, \dots, n$, and the common right external pathlength R_n and left external pathlength L_n , i.e., the sum of the right depths and the left depths over all nodes in the tree, of a random size- n binary tree. As an answer to the first question we can characterize the common distribution of the suitably scaled left depth $n^{-\frac{1}{2}} H_{n,j}^{[L]}$ and the scaled difference $n^{-\frac{1}{4}} (H_{n,j}^{[R]} - H_{n,j}^{[L]})$ between the right and the left depth of leaf j , for $j \sim \rho n$, with $0 < \rho < 1$. As an answer to the second question we can characterize the common distribution of the suitably scaled left external pathlength $n^{-\frac{3}{2}} L_n$ and the scaled difference $n^{-\frac{5}{4}} (R_n - L_n)$ between the right

[†]This work is supported by the Austrian Science Foundation FWF, grant S9608-N13.

and the left external pathlength. This work was inspired by a question of Donald Knuth during the workshop *Analysis of Algorithms* at MSRI, Berkeley, USA, in June 2004 asking about distributional properties of the common left and right pathlengths of binary trees and the results presented here might give an answer.

2 Results

Theorem 1 *The common probability that for a random binary tree of size n the left depth $H_{n,j}^{[L]}$ of leaf j is m_1 , whereas the right depth $H_{n,j}^{[R]}$ of leaf j is m_2 , is given by the following explicit formula:*

$$\mathbb{P}\{H_{n,j}^{[L]} = m_1 \wedge H_{n,j}^{[R]} = m_2\} = \begin{cases} \frac{m_1 m_2 (n+1)}{(2n-2j-m_1)(2j-m_2)} \frac{\binom{m_1+m_2}{m_1} \binom{2n-2j-m_1}{n-j} \binom{2j-m_2}{j}}{\binom{2n}{n}}, & \text{for } n \geq 1, 1 \leq j \leq n-1, 0 \leq m_1 \leq n-j, 0 \leq m_2 \leq j, \\ \frac{m_1(n+1)}{2n-m_1} \frac{\binom{2n-m_1}{n}}{\binom{2n}{n}}, & \text{for } n \geq 1, j = 0, 0 \leq m_1 \leq n, m_2 = 0, \\ \frac{m_2(n+1)}{2n-m_2} \frac{\binom{2n-m_2}{n}}{\binom{2n}{n}}, & \text{for } n \geq 1, j = 0, m_1 = 0, 0 \leq m_2 \leq n, \\ 1, & \text{for } n = 0, j = 0, m_1 = 0, m_2 = 0, \end{cases}$$

and $\mathbb{P}\{H_{n,j}^{[L]} = m_1 \wedge H_{n,j}^{[R]} = m_2\} = 0$, otherwise.

Theorem 2 *The pair of the suitably scaled left depth $n^{-\frac{1}{2}} H_{n,j}^{[L]}$ and the scaled difference $n^{-\frac{1}{4}} (H_{n,j}^{[R]} - H_{n,j}^{[L]})$ between the right and the left depth of leaf j in a random size- n binary tree converges for $j \sim \rho n$, with $0 < \rho < 1$, in distribution to a pair of random variables (X_ρ, Y_ρ) :*

$$(n^{-\frac{1}{2}} H_{n,j}^{[L]}, n^{-\frac{1}{4}} (H_{n,j}^{[R]} - H_{n,j}^{[L]})) \xrightarrow{(d)} (X_\rho, Y_\rho),$$

where (X_ρ, Y_ρ) have the common density $f_\rho(x, y)$, which is for $y \in \mathbb{R}$ and $x > 0$ given as follows:

$$f_\rho(x, y) = \frac{x^{\frac{3}{2}}}{4\pi\rho^{\frac{3}{2}}(1-\rho)^{\frac{3}{2}}} e^{-\frac{x^2}{4\rho(1-\rho)} - \frac{y^2}{4x}}, \quad \text{for } x > 0, y \in \mathbb{R}, \tag{1}$$

and $f_\rho(x, y) = 0$, otherwise.

Theorem 3 *The pair of the suitably scaled left external pathlength $n^{-\frac{3}{2}} L_n$ and the scaled difference $n^{-\frac{5}{4}} (R_n - L_n)$ between the right and the left external pathlength of a random size- n binary tree converges in distribution to a pair of random variables $(\tilde{L}, \tilde{\Delta})$:*

$$(n^{-\frac{3}{2}} L_n, n^{-\frac{5}{4}} (R_n - L_n)) \xrightarrow{(d)} (\tilde{L}, \tilde{\Delta}),$$

where the common distribution of $(\tilde{L}, \tilde{\Delta})$ is fully characterized by their mixed moments, which are for all $k, l \geq 0$ given below:

$$\mathbb{E}(\tilde{L}^k \tilde{\Delta}^{2l}) = \frac{\sqrt{\pi} k! (2l)! 2^{-3k-4l+2}}{\Gamma(\frac{3k+5l-1}{2})} \tilde{w}_{k,l}, \quad \text{and} \quad \mathbb{E}(\tilde{L}^k \tilde{\Delta}^{2l+1}) = 0. \tag{2}$$

The coefficients $\tilde{w}_{k,l}$ appearing in (2) are for $k+l > 0$ given by the recurrence

$$\tilde{w}_{k,l} = 2(3k+5l-6)(3k+5l-4)\tilde{w}_{k,l-1} + 2(3k+5l-4)\tilde{w}_{k-1,l} + \sum_{0 < i+j < k+l} \tilde{w}_{i,j} \tilde{w}_{k-i,l-j}, \tag{3}$$

with $\tilde{w}_{0,0} = -\frac{1}{2}$ and $\tilde{w}_{k,l} = 0$ if $k < 0$ or $l < 0$.

Remark: The sequence of numbers $\tilde{w}_{k,l}$ already appeared in a different context in Janson (2003) when studying the mixed moments of random variables ξ and η as defined below during the analysis of the Wiener index of binary trees. If $e(t)$, $0 \leq t \leq 1$ denotes a normalized Brownian excursion then the random variables ξ and η are defined as follows:

$$\xi := 2 \int_0^1 e(t) dt, \quad \text{and} \quad \eta := 4 \iint_{0 < s < t < 1} \min_{s \leq u \leq t} e(u) ds dt. \tag{4}$$

Using the results for the mixed moments of ξ and η as given in Janson (2003) we obtain further

$$(\tilde{L}, \tilde{\Delta}) \stackrel{(d)}{=} (2^{-\frac{1}{2}} \xi, 2^{\frac{1}{4}} \sqrt{\eta} N), \tag{5}$$

where N denotes a standard normal distributed random variable, which is independent of η and ξ .

References

- M. Bousquet-Mélou. Limit laws for embedded trees. Applications to the Integrated Superbrowonian Excursion. *Random Structures and Algorithms*, 2005. to appear.
- P. Chassaing and S. Janson. The center of mass of the ISE and the Wiener index of trees. *Electronic Communications in Probability*, 9:178–187, 2004.
- S. Janson. The Wiener index of simply generated random trees. *Random Structures and Algorithms*, 22:337–358, 2003.
- S. Janson. Left and right pathlengths in random binary trees. *Algorithmica*, 2004. to appear.
- C. Knessl and W. Szpankowski. Binary trees, left and right paths, WKB expansions, and Painlevé transcendents. In *Proc. of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO06), Miami*, 2006.
- D. E. Knuth. *The art of computer programming, Vol. 1, Fundamental Algorithms*. Addison-Wesley, Reading, Massachusetts, third edition, 1997.
- J.-F. Marckert. The rotation correspondence is asymptotically a dilatation. *Random Structures and Algorithms*, 24:118–132, 2004.

