Combinatorics, Probability and Computing (2008) 17, 831–835. © 2008 Cambridge University Press
doi:10.1017/S0963548308009486 Printed in the United Kingdom

The Hitting Time for the Height of a Random Recursive Tree

THOMAS M. LEWIS

Department of Mathematics, Furman University, Greenville, SC 29613, USA
(e-mail: tom.lewis@furman.edu)

Received 8 January 2008; revised 7 August 2008; first published online 3 October 2008

In this paper we provide a simple formula for the expected time for a random recursive tree to grow to a given height.

1. Introduction

A tree is a connected graph that has no cycles. A tree on the vertex set $\{0, 1, \ldots, n\}$ is called a *recursive tree* (or an *increasing tree*) if it is rooted at vertex 0 and the vertices on the path from 0 to each vertex $i \neq 0$ form an increasing sequence. Let Ω_n denote the set of recursive trees on the vertex set $\{0, 1, \ldots, n\}$ and observe that $|\Omega_n| = n!$. We will study the random sequence of recursive trees $\{T_n : n \geqslant 0\}$ that evolves according to the following rules: let $T_0 \in \Omega_0$ be the trivial tree, consisting of the vertex 0; thereafter, let $T_i \in \Omega_i$ be obtained by selecting one of the vertices of T_{i-1} with equal probability and attaching i to it as its descendant. T_n, so produced, is uniformly distributed on Ω_n. In Figure 1 we present an element of Ω_5 which we can interpret as the result of selecting, in order, the vertices 0, 1, 0, 1, 3.

In this paper we are interested in capturing, if only roughly, the rate of growth of random recursive trees. As is customary, the *depth* of a vertex v in a rooted tree is the length of the path from the root to v. The *height* of a rooted tree is the maximum of the depths of its vertices. Given a sequence of random recursive trees $\{T_n : n \geqslant 0\}$ and a positive integer L, let

$$\tau_L = \min\{n : \text{the height of } T_n = L\}.$$

In other words, τ_L is the first time when the tree has height L. Let $h_1(t) = 1$ for $t \in \mathbb{R}$ and, for $k \geqslant 2$, let

$$h_k(t) = \exp\left(\int_0^t h_{k-1}(s)\, ds\right), \quad t \in \mathbb{R}. \tag{1.1}$$

https://doi.org/10.1017/S0963548308009486 Published online by Cambridge University Press

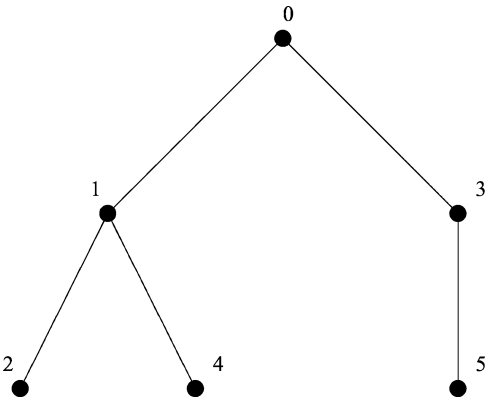


Figure 1. This is one of the 120 possible elements of Ω_5. The vertices were selected in this order: 0, 1, 0, 1, 3.

We will prove the following theorem.

Theorem 1.1. *For each* $k \geqslant 1$, $E(\tau_k) = h_k(1)$.

Recursive trees and related structures have been used by Gastwirth [5] and Gastwirth and Battacharya [6] to model pyramid schemes and chain letters, and by Najock and Heyde [15] to model the family trees of preserved copies in ancient manuscripts.

A number of diverse investigations of recursive trees have been carried out. Various aspects of the distance between pairs of vertices in recursive trees have been studied by Moon [14], Szymański [19], Mahmoud [10], Dobrow [3], and Neininger [16]. Kuba and Panholzer [9] have found the distribution of the number of descendants of a vertex. Su, Feng and Hu [18] have found the distribution of the number of branches of a given size and the distribution of the length of a random climb. The study of random climbs on recursive trees was initiated by Meir and Moon [12, 13]. Mahmoud and Smythe [11] have studied the distribution of leaves in rooted subtrees of recursive trees. Janson [8] has found, among other things, the distribution of the vertex degrees, establishing their asymptotic normality. Tetzlaff [20] has explored the effect of an avalanche on a recursive tree, that is, the removal of a subtree from a uniformly chosen vertex. Panholzer [17] has investigated the size of the ancestor tree and the size of the minimal spanning tree of p nodes selected uniformly from recursive trees. Devroye [2] has noted the connection between recursive trees and records. Recursive trees are related to the Stirling numbers, and probabilistic properties of the Stirling numbers have been studied by Feller [4], Harper [7], and Charalambides and Singh [1].

2. The proof of Theorem 1.1

Proof. Given an integer $k \geqslant 1$, let $c(i_1, i_2, \ldots, i_k)$ denote the number of recursive trees on with vertex set $\{0, 1, \ldots, i_1 + i_2 + \cdots + i_k\}$ with i_j vertices in generation j, $1 \leqslant j \leqslant k$. Some easy calculations show, for example, that $c(1, 2) = 1$ and $c(2, 1) = 3$. This function satisfies

https://doi.org/10.1017/S0963548308009486 Published online by Cambridge University Press

the recursive formula:

$$c(i_1, i_2, i_3, \ldots, i_k) = c(i_1 - 1, i_2, i_3, \ldots, i_k) + i_1 c(i_1, i_2 - 1, i_3, \ldots, i_k) + i_2 c(i_1, i_2, i_3 - 1, \ldots, i_k) + \cdots + i_{k-1} c(i_1, i_2, \ldots, i_k - 1). \tag{2.1}$$

Of course,

$$\begin{aligned} P(\tau_{k+1} > m) &= P(\text{depth of } T_m < k+1) \\ &= \frac{\text{number of trees in } \Omega_m \text{ with depth less than } k+1}{m!} \\ &= \frac{\sum_{i_1+\cdots+i_k=m} c(i_1, \ldots, i_k)}{m!}, \end{aligned}$$

from which it follows that

$$E(\tau_{k+1}) = \sum_{m=0}^{\infty} P(\tau_{k+1} > m) = \sum_{i_1,\ldots,i_k \geqslant 0} \frac{c(i_1, \ldots, i_k)}{(i_1 + \cdots + i_k)!}.$$

From the estimate

$$\frac{c(i_1, i_2, \ldots, i_k)}{(i_1 + i_2 + \cdots + i_k)!} \leqslant \frac{(i_1)^{i_2} (i_2)^{i_3} \cdots (i_{k-1})^{i_k}}{i_1! i_2! \cdots i_k!},$$

we conclude that the generating function

$$\sum_{i_1,i_2,\ldots,i_k=0}^{\infty} \frac{c(i_1, i_2, \ldots, i_k)}{(i_1 + i_2 + \cdots + i_k)!} \lambda_1^{i_1} \lambda_2^{i_2} \cdots \lambda_k^{i_k}$$

has infinite radius of convergence in each λ_i.

For $1 \leqslant j \leqslant k$, let $g_j(t) = h_{k-j+1}(1-t)$. By (1.1), $g_j(1) = 1$ and

$$g_j'(t) = -g_j(t) g_{j+1}(t). \tag{2.2}$$

Let

$$f(t) = \sum_{i_1,i_2,\ldots,i_k=0}^{\infty} \frac{c(i_1, i_2, \ldots, i_k)}{(i_1 + i_2 + \cdots + i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j}.$$

Since $f(1) = E(\tau_{k+1})$, it is enough to show that $f(1) = h_{k+1}(1)$.

We begin by noting that

$$\begin{aligned} f'(t) = &\sum_{i_1,i_2,\ldots,i_k=0}^{\infty} \frac{c(i_1, i_2, \ldots, i_k)}{(i_1 + i_2 + \cdots + i_k - 1)!} t^{i_1+i_2+\cdots+i_k-1} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &+ \sum_{i_1,i_2,\ldots,i_k=0}^{\infty} \frac{c(i_1, i_2, \ldots, i_k)}{(i_1 + i_2 + \cdots + i_k)!} t^{i_1+i_2+\cdots+i_k} \frac{d}{dt} \prod_{j=1}^{k} g_j(t)^{i_j} \\ = &\, A(t) + B(t). \end{aligned}$$

https://doi.org/10.1017/S0963548308009486 Published online by Cambridge University Press

By (2.1),

$$\begin{aligned}A(t) = &\sum_{i_1,i_2,\dots,i_k=0}^{\infty} \frac{c(i_1-1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k-1)!} t^{i_1+i_2+\cdots+i_k-1} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &+ \sum_{i_1,i_2,\dots,i_k=0}^{\infty} \frac{i_1 c(i_1,i_2-1,\dots,i_k)}{(i_1+i_2+\cdots+i_k-1)!} t^{i_1+i_2+\cdots+i_k-1} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &\vdots \\ &+ \sum_{i_1,i_2,\dots,i_k=0}^{\infty} \frac{i_{k-1} c(i_1,i_2,\dots,i_k-1)}{(i_1+i_2+\cdots+i_k-1)!} t^{i_1+i_2+\cdots+i_k-1} \prod_{j=1}^{k} g_j(t)^{i_j}.\end{aligned}$$

By a change of variables, this can be expressed as

$$\begin{aligned}A(t) = &\sum_{i_1,i_2,\dots,i_k=0}^{\infty} g_1(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &+ \sum_{i_1,i_2,\dots,i_k=0}^{\infty} i_1 g_2(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &\vdots \\ &+ \sum_{i_1,i_2,\dots,i_k=0}^{\infty} i_{k-1} g_k(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j}.\end{aligned}$$

Recall that $g_k(t) = h_1(t) = 1$ for all $t \in \mathbb{R}$; consequently, by (2.2),

$$\frac{d}{dt} \prod_{j=1}^{k} g_j(t)^{i_j} = -(i_1 g_2(t) + i_2 g_3(t) \cdots + i_{k-1} g_k(t)) \prod_{j=1}^{k} g_j(t)^{i_j},$$

from which it follows that

$$\begin{aligned}B(t) = &- \sum_{i_1,i_2,\dots,i_k=0}^{\infty} i_1 g_2(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &- \sum_{i_1,i_2,\dots,i_k=0}^{\infty} i_2 g_3(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j} \\ &\vdots \\ &- \sum_{i_1,i_2,\dots,i_k=0}^{\infty} i_{k-1} g_k(t) \frac{c(i_1,i_2,\dots,i_k)}{(i_1+i_2+\cdots+i_k)!} t^{i_1+i_2+\cdots+i_k} \prod_{j=1}^{k} g_j(t)^{i_j}.\end{aligned}$$

Consequently $f'(t) = A(t) + B(t) = g_1(t) f(t)$. Since $f(0) = 1$,

$$f(t) = \exp\left(\int_0^t g_1(s)\, ds \right)$$

https://doi.org/10.1017/S0963548308009486 Published online by Cambridge University Press

and

$$f(1) = \exp\left(\int_0^1 g_1(s)\,ds\right) = \exp\left(\int_0^1 h_k(s)\,ds\right) = h_{k+1}(1),$$

as was to be shown. □

References

[1] Charalambides, C. A. and Singh, J. (1988) A review of the Stirling numbers, their generalizations and statistical applications. *Comm. Statist. Theory Methods* **17** 2533–2595.

[2] Devroye, L. (1988) Applications of the theory of records in the study of random trees. *Acta Inform.* **26** 123–130.

[3] Dobrow, R. P. (1996) On the distribution of distances in recursive trees. *J. Appl. Probab.* **33** 749–757.

[4] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edn, Wiley, New York.

[5] Gastwirth, J. L. (1977) A probability model of a pyramid scheme. *Amer. Statist.* **31** 79–82.

[6] Gastwirth, J. L. and Bhattacharya, P. K. (1984) Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.* **32** 527–536.

[7] Harper, L. H. (1967) Stirling behavior is asymptotically normal. *Ann. Math. Statist.* **38** 410–414.

[8] Janson, S. (2005) Asymptotic degree distribution in random recursive trees. *Random Struct. Alg.* **26** 69–83.

[9] Kuba, M. and Panholzer, A. (2006) Descendants in increasing trees. *Electron. J. Combin.* **13** #8 (electronic).

[10] Mahmoud, H. M. (1994) A strong law for the height of random binary pyramids. *Ann. Appl. Probab.* **4** 923–932.

[11] Mahmoud, H. M. and Smythe, R. T. (1991) On the distribution of leaves in rooted subtrees of recursive trees. *Ann. Appl. Probab.* **1** 406–418.

[12] Meir, A. and Moon, J. W. (1976) Climbing certain types of rooted trees I. In *Proc. Fifth British Combinatorial Conference* (University of Aberdeen 1975), Vol. XV of *Congressus Numerantium*, pp. 461–469.

[13] Meir, A. and Moon, J. W. (1978) Climbing certain types of rooted trees II. *Acta Math. Acad. Sci. Hungar.* **31** 43–54.

[14] Moon, J. W. (1974) The distance between nodes in recursive trees. In *Combinatorics: Proc. British Combinatorial Conference* (University College Wales, Aberystwyth, 1973), Vol. 13 of *London Mathematical Society Lecture Notes*, Cambridge University Press, pp. 125–132.

[15] Najock, D. and Heyde, C. C. (1982) On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Probab.* **19** 675–680.

[16] Neininger, R. (2002) The Wiener index of random trees. *Combin. Probab. Comput.* **11** 587–597.

[17] Panholzer, A. (2004) The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Struct. Alg.* **25** 179–207.

[18] Su, C., Feng, Q. and Hu, Z. (2006) Uniform recursive trees: Branching structure and simple random downward walk. *J. Math. Anal. Appl.* **315** 225–243.

[19] Szymański, J. (1990) On the maximum degree and the height of a random recursive tree. In *Random Graphs '87* (Poznań 1987), Wiley, pp. 313–324.

[20] Tetzlaff, G. T. (2002) Breakage and restoration in recursive trees. *J. Appl. Probab.* **39** 383–390.

https://doi.org/10.1017/S0963548308009486 Published online by Cambridge University Press