

PROFILES OF RANDOM TREES: CORRELATION AND WIDTH OF RANDOM RECURSIVE TREES AND BINARY SEARCH TREES

MICHAEL DRMOTA,* *Technische Universität Wien*

HSIEN-KUEI HWANG,** *Academia Sinica*

Abstract

In a tree, a level consists of all those nodes that are the same distance from the root. We derive asymptotic approximations to the correlation coefficients of two level sizes in random recursive trees and binary search trees. These coefficients undergo sharp sign-changes when one level is fixed and the other is varying. We also propose a new means of deriving an asymptotic estimate for the expected width, which is the number of nodes at the most abundant level. Crucial to our methods of proof is the uniformity achieved by singularity analysis.

Keywords: Correlation coefficient; random recursive tree; random binary search tree; width; total path length; sign-change

2000 Mathematics Subject Classification: Primary 60C05

Secondary 62J10; 05C05

1. Introduction

This paper is a sequel to Drmota and Hwang (2005) and Fuchs *et al.* (2005), in which the authors addressed the limit distributions of profiles (the numbers of nodes at each level) in random recursive trees and binary search trees. Further to the many intriguing phenomena unveiled there, we show in this paper that the correlation coefficients of two level sizes, in both classes of tree, exhibit sharp sign-changes. The method of proof used to derive the uniform estimates of covariances will be useful in obtaining the asymptotics of the expected widths, for which only almost-sure results, but no moment estimates, were previously known.

1.1. Random recursive trees

Recursive trees of n nodes are nonplane, rooted, labeled trees with labels $\{1, \dots, n\}$ (at nodes) such that the labels along any path from the root form a strictly increasing sequence. By a *random* recursive tree, we mean that all recursive trees of n nodes are equally likely. An alternative way of constructing a random recursive tree of n nodes is as follows. We start from a single node with label 1; then, at the i th insertion step, the new label i chooses any of the previous $i - 1$ nodes to be its parent with equal probability (and we link them by an edge), and the same procedure continues until the tree contains n nodes. This procedure implies that there are $(n - 1)!$ such trees.

Received 26 October 2004; revision received 21 January 2005.

* Postal address: Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstrasse 8-10/118, 1040 Wien, Austria.

** Postal address: Institute of Statistical Science, Academia Sinica, Taipei, 115, Taiwan.

Email address: hkhwang@stat.sinica.edu.tw

Recursive trees (following Meir and Moon (1974)) have also appeared in other fields under different names: ‘concave node-weighted trees’ in Tapia and Myers (1967), ‘growing trees’ in Na and Rapoport (1970), ‘pyramid scheme’ in Gastwirth (1977), ‘heap-ordered trees’ in Grossman and Larson (1989), and ‘random circuits with fanin one’ in Arya *et al.* (1999). They have been introduced as simple growing models for several real-life networks like social systems (Na and Rapoport (1970)), sales-distribution networks (Moon (1974)), and the Internet; see Fuchs *et al.* (2005) for more references. Their simple tree representations also found applications in many linear tree algorithms; see Mitchell *et al.* (1979).

1.2. Profile of random recursive trees

We consider the number of nodes, denoted by $Y_{n,k}$, at distance k from the root in a random recursive tree of n nodes. Many properties of $Y_{n,k}$ are known. We briefly summarize the interesting phenomena exhibited by $Y_{n,k}$, as follows; see Drmota and Hwang (2005) and Fuchs *et al.* (2005) for more information.

- For large, fixed n , the mean of $Y_{n,k}$ is asymptotically unimodal in k , but the variance is asymptotically *bimodal*.
- The normalized random variables $Y_{n,k}/E(Y_{n,k})$ converge in distribution to some limit law $Y(\alpha)$ when $k \geq 1$ and where $\alpha := \lim_{n \rightarrow \infty} k/\log n \in [0, e)$.
- Convergence of all moments of $Y_{n,k}/E(Y_{n,k})$ to $Y(\alpha)$ holds only for $\alpha \in [0, 1]$.
- If $\alpha = 0$ (and $k \geq 1$) then the sequence of the centered and normalized random variables $(Y_{n,k} - E(Y_{n,k}))/\text{var}(Y_{n,k})^{1/2}$ converges in distribution to the standard normal law.
- If $\alpha = 1$ and $|k - \log n| \rightarrow \infty$ then $(Y_{n,k} - E(Y_{n,k}))/\text{var}(Y_{n,k})^{1/2}$ converges in distribution (and with all moments) to $Y'(1) = (dY(\alpha)/d\alpha)|_{\alpha=1}$, the limit law converged to by the total path length $\sum_k kY_{n,k}$.
- If $k = \log n + O(1)$ then $(Y_{n,k} - E(Y_{n,k}))/\text{var}(Y_{n,k})^{1/2}$ does not converge to a fixed limit law.

1.3. Covariance of $Y_{n,k}$ and $Y_{n,h}$

The results derived in the authors’ previous papers dealt with stochastic behaviors of a *single* level size. In this paper, we examine the asymptotics of the correlation coefficient of *two* level sizes, which turns out to undergo a sharp sign-change at $\alpha = 1$ (when the other level is fixed and not near $\log n$).

To state our results, we first introduce some notation. Define

$$f(u, v) := \frac{1}{\Gamma(u + v)(u + v - uv)} - \frac{1}{\Gamma(u + 1)\Gamma(v + 1)}, \tag{1.1}$$

where Γ is the gamma function, and

$$p(s, t) := c_2st + c_1(s + t) + c_0, \tag{1.2}$$

where the coefficients are given by

$$\begin{aligned} c_2 &:= f''_{uv}(1, 1) = 2 - \frac{1}{6}\pi^2, \\ c_1 &:= -\frac{1}{2}f'''_{uv^2}(1, 1) = c_2(1 - \gamma) - \zeta(3) + 1, \\ c_0 &:= \frac{1}{4}f^{(4)}_{u^2v^2}(1, 1) = c_2(1 + 2\gamma - \gamma^2) + 2c_1(1 - \gamma) - \frac{1}{360}\pi^4. \end{aligned} \tag{1.3}$$

Here, a prime or a superscript in parentheses denotes partial differentiation with respect to the variables in subscript, γ denotes Euler’s constant, and $\zeta(3) := \sum_{j \geq 1} j^{-3}$. Also define

$$c_3 := f'_v(\alpha, 1) = -\frac{\psi(\alpha + 1) + \gamma - \alpha}{\Gamma(\alpha + 1)},$$

$$c_4 := -\frac{1}{2} f''_{v^2}(\alpha, 1)$$

$$= -\frac{(\psi(\alpha + 1) + 1 - \alpha)^2 + (\alpha - 1)^2 - (1 - \gamma)^2 - \psi'(\alpha + 1) - 1 + \frac{1}{6}\pi^2}{2\Gamma(\alpha + 1)},$$

where $\psi(x)$ denotes the derivative of $\log \Gamma(x)$. For $k, h \geq 1$, let $\alpha_{n,k} := k/\log n$ and $\beta_{n,h} := h/\log n$, and let α and β be their respective limit ratios, if they exist (when n tends to infinity).

Theorem 1.1. *If $\alpha, \beta \in [0, 2)$ then the correlation coefficient of $Y_{n,k}$ and $Y_{n,h}$ satisfies*

$$\rho(Y_{n,k}, Y_{n,h}) \sim \begin{cases} \frac{\sqrt{(2k-1)(2h-1)}}{k+h-1} & \text{if } \alpha = \beta = 0, \\ 0 & \text{if } \alpha = 0, \beta \neq 0, \\ \frac{f(\alpha, \beta)}{\sqrt{f(\alpha, \alpha)f(\beta, \beta)}} & \text{if } \alpha, \beta \neq 1, \\ \frac{c_3 t_{n,h} + c_4}{\sqrt{f(\alpha, \alpha)p(t_{n,h}, t_{n,h})}} & \text{if } \alpha \neq 1, \beta = 1, \\ \frac{p(s_{n,k}, t_{n,h})}{\sqrt{p(s_{n,k}, s_{n,k})p(t_{n,h}, t_{n,h})}} & \text{if } \alpha = \beta = 1, \end{cases} \tag{1.4}$$

where $s_{n,k} := k - \log n$ and $t_{n,h} := h - \log n$.

By symmetry, all cases with $\alpha, \beta \in [0, 2)$ are covered. In particular, the result here also implies the estimates the authors derived for $\text{var}(Y_{n,k})$ in previous papers. A comparison of the different approaches used so far for $\text{var}(Y_{n,k})$ is given in the final section.

1.4. Corollaries and intuitive interpretations

Corollary 1.1. *The correlation coefficient of $Y_{n,k}$ and $Y_{n,h}$ is asymptotic to 0 if $k = o(\log n)$ and $h = o(\log n)$, where $0 \leq \beta < 2$.*

Thus, the sizes at the first few levels ($k = o(\log n)$) are asymptotically independent of those at levels that are much greater than k .

Corollary 1.2. *The correlation coefficient of $Y_{n,k}$ and $Y_{n,h}$ is asymptotic to 1 if (i) $\alpha = \beta \neq 1$ ($0 \leq \alpha, \beta < 2$), or (ii) both $|s_{n,k}| \rightarrow \infty$ and $|t_{n,h}| \rightarrow \infty$ (not necessarily at the same rate) when $\alpha = \beta = 1$.*

The first case is intuitively clear, but the second case is less transparent.

Corollary 1.3. *Asymptotically, the correlation coefficient $\rho(Y_{n,k}, Y_{n,h})$ exhibits a sharp sign-change at $\beta = 1$ when $\alpha \in (0, 2)$ is fixed and β is varied from 0 to 2.*

Plots of the asymptotic correlation coefficient are given in Figures 1, 2, and 3, highlighting, in particular, the discontinuous sign-change at $\beta = 1$.

Intuitively, the sizes of neighboring levels are expected to have positive correlation. The sharp sign-change at $\beta = 1$ is roughly because of the property that almost all nodes in a random

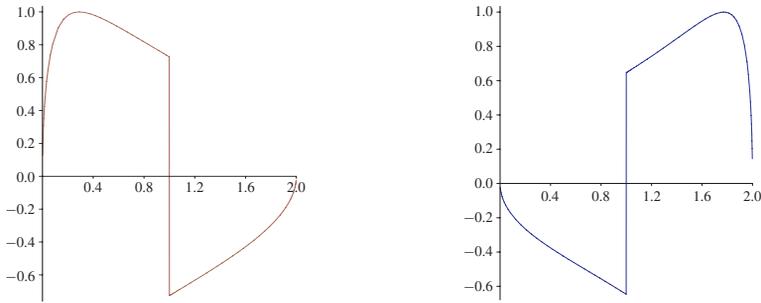


FIGURE 1: The asymptotic correlation coefficient of the number of nodes at two levels, plotted against β . The discontinuity in sign at $\beta = 1$ is visible in both figures. Here $\alpha = \frac{1}{2}\gamma \approx 0.28$ (left) and $\alpha = \pi^{1/2} \approx 1.77$ (right), and $\beta \in (0, 2)$.

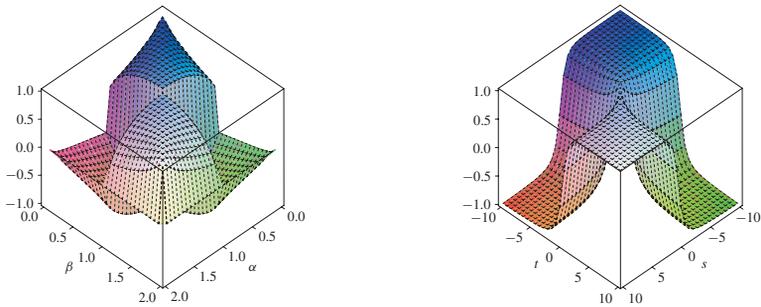


FIGURE 2: Three-dimensional renderings of the limiting correlation coefficients $f(\alpha, \beta)/[f(\alpha, \alpha) f(\beta, \beta)]^{1/2}$ (left) and $p(s, t)/[p(s, s)p(t, t)]^{1/2}$ (right).

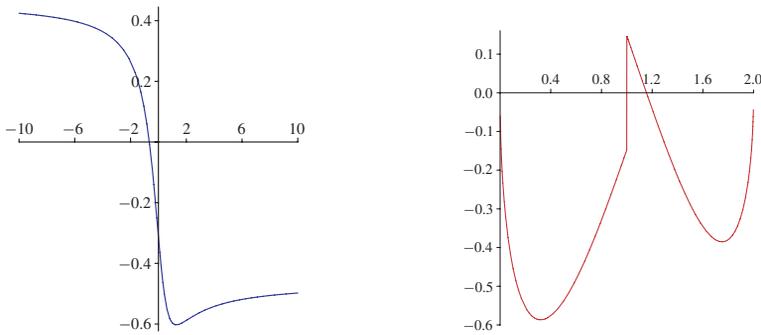


FIGURE 3: Asymptotic correlation coefficient when $\beta = 1$, plotted against t with $\alpha = 0.1$ (left), and against α with $t = \frac{1}{2}\gamma$ (right).

tree lie at the levels $k = \log n + O((\log n)^{1/2})$, each having about $n/(\log n)^{1/2}$ nodes (by the estimate

$$E(Y_{n,k}) \sim \frac{(\log n)^k}{k! \Gamma(\alpha_{n,k} + 1)} \quad (1 \leq k = O(\log n)),$$

and the bimodal behavior of the variance near these levels; see Drmota and Hwang (2005)). This implies that if one level k with, say, $k/\log n < 1$ has more nodes, then (i) levels near $\log n$ are likely to have more nodes, and (ii) levels with $h/\log n > 1$ have fewer nodes available; this also roughly explains why $Y_{n,k}$ and $Y_{n,h}$ are negatively correlated (see Figure 1).

Our method of proof starts from the relation

$$\sum_{k,h} E(Y_{n,k} Y_{n,h}) u^k v^h = \frac{n}{u + v - uv} \left(\binom{n + u + v - 1}{n} - \binom{n + uv - 1}{n} \right); \quad (1.5)$$

see below for a self-contained proof or see van der Hofstad *et al.* (2002). Then, (1.4) is derived from a uniform estimate for the function on the right-hand side in the (u, v) -plane (by applying the singularity analysis of Flajolet and Odlyzko (1990)) and then by extending the saddle-point method used in Hwang (1995).

1.4.1. *Width.* Our analytic approach is also useful in deriving a uniform estimate for

$$E((Y_{n,k} - Y_{n,h})^2),$$

which turns out to be the crucial step in proving an asymptotic approximation to the expected width, defined to be $W_n := \max_k Y_{n,k}$.

Theorem 1.2. *The width W_n satisfies*

$$\frac{W_n}{n/\sqrt{2\pi \log n}} \rightarrow 1 \quad (1.6)$$

almost surely, and

$$E(W_n) = \frac{n}{\sqrt{2\pi \log n}} (1 + O((\log n)^{-1/4} \log \log n)). \quad (1.7)$$

The almost-sure convergence is proved by modifying the martingale arguments used in Chauvin *et al.* (2001) for random binary search trees. Such arguments, based on considering the random polynomial $\sum_k Y_{n,k} z^k$, also provide a natural interpretation of the result (see Fuchs *et al.* (2005)) that the sequence of random variables $(Y_{n,k} - E(Y_{n,k}))/\text{var}(Y_{n,k})^{1/2}$ converges to the same limit law as the total path length $T_n := \sum_k k X_{n,k}$ when $k \sim \log n$ and $|k - \log n| \rightarrow \infty$; see Section 3 for more details.

1.5. Binary search trees

Binary search trees (BSTs) are rooted, labeled binary trees with the *search property*: all labels in the left or right subtree of any node x are respectively smaller or larger than the label of x . Given a sequence of numbers, we can construct a BST by placing the first element at the root, and then by successively directing all smaller or larger numbers to the left or the right branch, respectively. Both subtrees, if nonempty, are recursively constructed by the same procedure and are themselves BSTs; see Figure 4.

BSTs were first introduced in the early 1960s by Windley (1960), Booth and Colin (1960), and Hibbard (1962), and are one of the simplest prototypical data structures; see Knuth (1997) and Mahmoud (1992).

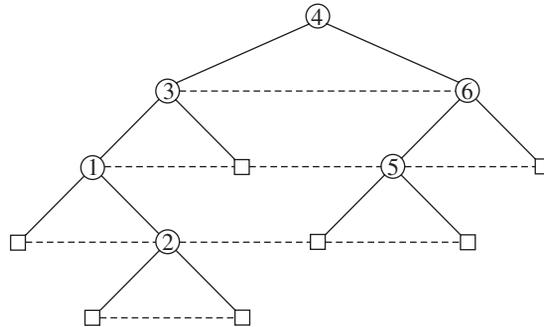


FIGURE 4: The binary search tree constructed from the sequence {4, 3, 1, 6, 5, 2}. Internal nodes are marked by circles and external nodes by squares.

1.6. Random binary search trees

Assume that all $n!$ permutations of n elements are equally likely. Given a random permutation, we call the BST constructed from the permutation a *random BST*. We distinguish between two types of node: *internal nodes* are nodes holding labels and *external nodes* are virtual nodes added so that all internal nodes are of outdegree 2; see Figure 4.

We denote by $X_{n,k}$ and $I_{n,k}$ the numbers of external and internal nodes at level k in a random BST of n internal nodes, respectively, the root being at level 0. The distributional properties of both types of profile ($X_{n,k}$ and $I_{n,k}$) are similar to those of $Y_{n,k}$; see Fuchs *et al.* (2005) for details.

Here, an interesting property for the covariance of two level sizes is that, while the limiting correlation coefficients of $I_{n,k}$ and $I_{n,h}$ exhibit a sharp sign-change at $\alpha = 2$, the limiting correlation coefficients of $X_{n,k}$ and $X_{n,h}$ exhibit two sharp sign-changes, one at $\alpha = 1$ and the other at $\alpha = 2$. An intuitive interpretation of this will be given in Section 4.

1.7. Organization of the paper

In the next section, we prove Theorem 1.1, on the asymptotic estimates of the correlation coefficients of two level sizes in random recursive trees. The width and related quantities are addressed in Section 3. Results for random BSTs are given in Section 4, without proof. We then conclude the paper with a brief comparative discussion of the methods of proof used to derive asymptotic estimates for the variances.

2. Correlation of two level sizes

We prove Theorem 1.1 in this section. Note that the L_2 -convergence of $Y_{n,k}/\mu_{n,k}$ (see Fuchs *et al.* (2005)), where $\mu_{n,k} := E(X_{n,k})$, can also be applied to prove (1.4) in the case when $\alpha, \beta \notin \{0, 2\}$; here we give a direct and uniform approach applicable to all cases.

2.1. Recurrence of $Y_{n,k}$ and $E(Y_{n,k})$

All our results are based on the recurrence relation satisfied by $Y_{n,k}$:

$$Y_{n,k} \stackrel{D}{=} Y_{\text{uniform}[1,n-1],k-1} + Y_{n-\text{uniform}[1,n-1],k}^* \quad (n \geq 2, k \geq 1). \tag{2.1}$$

The initial values are $Y_{n,0} = \delta_{n,1}$, where we use the Kronecker symbol, the random variable $\text{uniform}[1, n - 1]$ takes each of the values $\{1, \dots, n - 1\}$ with equal probability, and $Y_{n,k}^*$ is an independent copy of $Y_{n,k}$; see Fuchs *et al.* (2005) or van der Hofstad *et al.* (2002).

From (2.1), it follows that the mean satisfies

$$\begin{aligned} \mu_{n,k} &= [u^k] \binom{n+u-1}{n-1} \\ &= \frac{(\log n)^k}{k! \Gamma(\alpha_{n,k} + 1)} (1 + O((\log n)^{-1})), \end{aligned} \tag{2.2}$$

where $[u^k]F(u)$ denotes the coefficient of u^k in the Taylor expansion of F and the O -term holds uniformly for $1 \leq k = O(\log n)$; see Hwang (1995).

Proof of (1.5). We now prove (1.5). By (2.1), we have the recurrence

$$\begin{aligned} E(Y_{n,k} Y_{n,h}) &= \frac{1}{n-1} \sum_{1 \leq j < n} (E(Y_{j,k-1} Y_{j,h-1}) + E(Y_{j,k}^* Y_{j,h}^*)) \\ &\quad + \frac{1}{n-1} \sum_{1 \leq j < n} (\mu_{j,k-1} \mu_{n-j,h} + \mu_{j,k} \mu_{n-j,h-1}). \end{aligned}$$

Let $F_n(u, v) := \sum_{k,h} E(X_{n,k} X_{n,h}) u^k v^h$. Then $F_1(u, v) = 1$ and

$$F_n(u, v) = \frac{1+uv}{n-1} \sum_{1 \leq j < n} F_j(u, v) + \frac{u+v}{n-1} \sum_{1 \leq j < n} \binom{j+u-1}{j-1} \binom{n-j+v-1}{n-j-1} \tag{2.3}$$

for $n \geq 2$. The final sum is equal to

$$\frac{u+v}{n-1} [z^n] z^2 (1-z)^{-u-v-2} = \frac{u+v}{n-1} \binom{n+u+v-1}{n-2}.$$

The recurrence equation (2.3) is then either solved by considering $nF_{n+1} - (n-1)F_n$ and iterating the resulting first-order difference equation, or by considering the differential equation satisfied by $\sum_n F_{n+1} z^n$. This proves (1.5).

2.2. An asymptotic expansion for the covariance

We now derive an asymptotic expansion for $\text{cov}(Y_{n,k}, Y_{n,h})$.

First, by singularity analysis (see Flajolet and Odlyzko (1990)), we have

$$n \binom{n+w-1}{n} = n [z^n] (1-z)^{-w} = \frac{n^w}{\Gamma(w)} (1 + O(|w|^{-2} n^{-1})),$$

the O -term holding uniformly for finite complex w . Thus, by (1.5) and (2.2),

$$\text{cov}(Y_{n,k}, Y_{n,h}) = C_{k,h}(n) (1 + O(n^{-1})) + O\left(\delta_{k,h} \frac{(\log n)^k}{k!}\right) \tag{2.4}$$

uniformly for $1 \leq k, h \leq K \log n$, where

$$C_{k,h}(n) := [u^k v^h] f(u, v) n^{u+v},$$

with f defined as in (1.1).

If $\alpha + \beta \neq 0$ then we apply the saddle-point method used in Hwang (1995) by first expanding f as

$$f(u, v) = \sum_{\ell, r \geq 0} f_{\ell, r}(u - \alpha_{n, k})^\ell (v - \beta_{n, h})^r,$$

where $f_{\ell, r} := f_{u^\ell v^r}(\alpha_{n, k}, \beta_{n, h}) / (\ell! r!)$, and then integrating term by term to obtain the formal expansion

$$C_{k, h}(n) \sim \sum_{\ell, r \geq 0} f_{\ell, r} \Xi_\ell(n, k) \Xi_r(n, h), \tag{2.5}$$

where

$$\begin{aligned} \Xi_\ell(n, k) &:= [u^\ell](u - \alpha_{n, k})^\ell n^u \\ &= \frac{(\log n)^k}{k!} \sum_{0 \leq j \leq \ell} \binom{\ell}{j} (-\alpha_{n, k})^{\ell-j} \frac{k \cdots (k - j + 1)}{(\log n)^j} \quad (\ell \geq 0). \end{aligned}$$

The first few values of Ξ_r are as follows:

$$\begin{aligned} \Xi_0(n, k) &= 1, & \Xi_1(n, k) &= 0, & \Xi_2(n, k) &= -\frac{k}{(\log n)^2}, \\ \Xi_3(n, k) &= \frac{2k}{(\log n)^3}, & \Xi_4(n, k) &= \frac{3k(k - 2)}{(\log n)^4}. \end{aligned}$$

Since $\Xi_r(n, k)$ equals $(\log n)^{-r}$ times a polynomial in k of degree $\lfloor \frac{1}{2}r \rfloor$, the double sum on the right-hand side of (2.5) can be regrouped and gives an asymptotic expansion when $k = O(\log n)$. The error analysis is similar to those in Hwang (1995), (1997), and we find that (2.5) holds uniformly for $1 \leq k, h \leq 2 \log n - \omega_n (\log n)^{1/2}$, where ω_n is any sequence tending to infinity. The error term $[u^k v^h] C_{k, h}(n) O(n^{-1})$ appearing in (2.4) is handled similarly, and is asymptotically negligible.

From the explicit forms of the Ξ_ℓ , we obtain the expansion

$$\begin{aligned} C_{k, h}(n) &= \frac{(\log n)^{k+h}}{k! h!} \left\{ f_{0,0} - \frac{1}{\log n} (f_{2,0} \alpha_{n, k} + f_{0,2} \beta_{n, h}) \right. \\ &\quad + \frac{1}{(\log n)^2} (3(f_{4,0} \alpha_{n, k}^2 + f_{0,4} \beta_{n, h}^2) + f_{2,2} \alpha_{n, k} \beta_{n, h}) \\ &\quad \left. + 2(f_{3,0} \alpha_{n, k} + f_{0,3} \beta_{n, h}) + O((\log n)^{-3}) \right\}, \tag{2.6} \end{aligned}$$

which is sufficient for our purposes.

2.2.1. *Special cases.* Assume that $0 \leq \alpha, \beta < 2$. If $\alpha, \beta \notin \{0, 1\}$ then

$$f_{0,0} = f(\alpha_{n, k}, \beta_{n, h}) \rightarrow f(\alpha, \beta) \neq 0,$$

and we obtain

$$\text{cov}(Y_{n, k}, Y_{n, h}) \sim f(\alpha, \beta) \frac{(\log n)^{k+h}}{k! h!}$$

uniformly for $1 \leq k, h \leq 2 \log n - \omega_n(\log n)^{1/2}$. This proves Theorem 1.4 in the case in which $\alpha, \beta \notin \{0, 1\}$. It also implies that

$$\text{var}(X_{n,k}) \sim f(\alpha, \alpha) \frac{(\log n)^{2k}}{k!^2} \quad (1 \leq k \leq 2 \log n - \omega_n \sqrt{\log n}).$$

If $\alpha = \beta = 1$ then, by (2.6) and the approximations

$$f_{0,0} \sim c_2 \frac{s_{n,k} t_{n,h}}{(\log n)^2}, \quad f_{0,2} \sim -c_1 \frac{s_{n,k}}{\log n}, \quad f_{2,0} \sim -c_1 \frac{t_{n,h}}{\log n}, \quad f_{2,2} \sim c_0,$$

where $k = \log n + s_{n,k}, h = \log n + t_{n,h}$, and the c_j are defined as in (1.3), we obtain

$$\text{cov}(Y_{n,k}, Y_{n,h}) \sim \frac{p(s_{n,k}, t_{n,h}) (\log n)^{k+h}}{(\log n)^2 k! h!},$$

where p is as given in (1.2). This also implies that $\text{var}(X_{n,k}) \sim p(s_{n,k}, s_{n,k})(\log n)^{2k-2}/(k!)^2$.

If $\alpha = 0$ and $\beta \in (0, 1)$ then, as above, we have

$$\text{cov}(Y_{n,k}, Y_{n,h}) \sim \begin{cases} -\frac{(\log n)^{k+h-1}}{(k-1)! h! \Gamma(\beta+1)} (\psi(\beta+1) - 1 + \gamma) & \text{if } \beta \neq 1, \\ \frac{(\log n)^{k+h-2}}{(k-1)! h! \Gamma(\beta+1)} ((1 - \frac{1}{6}\pi^2)s_{n,h} + 2 - \gamma - \zeta(3) - \frac{1}{4}\pi^2 + \frac{1}{6}\pi^2\gamma) & \text{if } \beta = 1, \end{cases}$$

so that $\rho(Y_{n,k}, Y_{n,h}) \rightarrow 0$ in both cases.

The case in which $\beta = 1$ and $\alpha \neq 1$ is treated similarly.

A change of variable $u \mapsto wv$ is useful for the remaining case of when $\alpha = \beta = 0$; then, a similar analysis gives

$$\text{cov}(Y_{n,k}, Y_{n,h}) \sim \frac{(\log n)^{k+h-1}}{(k-1)! (h-1)! (k+h-1)}. \tag{2.7}$$

Alternatively, we can use the exact expression (see van der Hofstad *et al.* (2002))

$$E(Y_{n,k} Y_{n,h}) = \sum_{0 \leq j \leq k} \binom{2j+h-k}{j+h-k} [w^{j+h+1}] \binom{n-1+w}{n-1},$$

obtained from expanding the right-hand side of (1.5), and then proceed as we did above (the two terms with indices $j = k - 1, k$ suffice to obtain (2.7)).

Proof of Corollary 1.3. When $\alpha, \beta \in (0, 2), \alpha \neq 1$, we have, by (1.4),

$$\lim_{\beta \rightarrow 1} \frac{f(\alpha, \beta)}{\sqrt{f(\alpha, \alpha) f(\beta, \beta)}} = \text{sgn}(1 - \beta) \frac{\psi(\alpha + 1) - \alpha + \gamma}{\sqrt{c_2 f(\alpha, \alpha) \Gamma(\alpha + 1)}};$$

thus, the sign-change follows. The case in which $\alpha = 1$ can be checked similarly.

The proofs of the other corollaries to Theorem 1.1 are straightforward and are therefore omitted.

3. Profile and width

Profiles of trees are closely related to many other shape parameters. In this section, we briefly discuss the connection between profile and width, starting from the derivation of an asymptotic estimate for the expected width, i.e. from the proof of (1.7). Then we consider the level polynomial $M_n(z) := \sum_k Y_{n,k} z^k$, which will be seen to be a convenient tool for proving (1.6) and for relating the limit properties of the profile to those of the total path length (and other weighted path lengths).

3.1. The expected width

Since the width is defined by $W_n = \max_k Y_{n,k}$, we have, by the estimate (2.2),

$$E(W_n) \geq \max_k E(Y_{n,k}) = \frac{n}{\sqrt{2\pi \log n}} (1 + O((\log n)^{-1/2})).$$

However, it is less clear how to obtain a tight upper bound. The arguments introduced in Chauvin *et al.* (2001) can be used to prove the almost-sure convergence result (1.6) (see below for a sketch of the proof), but do not lead to an effective upper bound for the expected width. Instead, we introduce a new argument, reducing the determination of an upper bound to estimating the mean and the variance of some differences between level sizes, and show that the above lower bound is indeed tight.

We start with a probabilistic lemma.

Lemma 3.1. *Let $Z(t)$ be a stochastic process on the space of continuous functions on $[0, 1]$. Assume that there exist constants $\lambda \geq 0$ and $\theta > 1$ such that*

$$P(|Z(t_1) - Z(t_2)| \geq \delta) = O(|t_1 - t_2|^\theta \delta^{-\lambda}) \tag{3.1}$$

uniformly for all $t_1, t_2 \in [0, 1]$. Then we have

$$P\left(\max_{|s-t| \leq \varepsilon} |Z(s) - Z(t)| \geq \delta\right) = O(\varepsilon^{\theta-1} \delta^{-\lambda}). \tag{3.2}$$

Proof. We modify the proof of Theorem 12.3 of Billingsley (1968). First, our assumption (3.1) is exactly Billingsley (1968, Equation (12.50)) with $F(t) = t$. It follows that, for $\varepsilon > 0$ (with $1/\varepsilon$ an integer; cf. Billingsley (1968, Equation (12.57))),

$$\sum_{j < 1/\varepsilon} P\left(\sup_{j\varepsilon \leq s \leq (j+1)\varepsilon} |Z(s) - Z(j\varepsilon)| \geq \delta\right) = O(\varepsilon^{\theta-1} \delta^{-\lambda}).$$

Similarly, we obtain

$$\sum_{j < 1/\varepsilon} P\left(\sup_{j\varepsilon \leq s \leq (j+1)\varepsilon} |Z(s) - Z((j+1)\varepsilon)| \geq \delta\right) = O(\varepsilon^{\theta-1} \delta^{-\lambda}).$$

Now, suppose that there exist $s, t \in [0, 1]$ with $|s - t| \leq \varepsilon$ and $|Z(s) - Z(t)| \geq \delta$. Then there exists a $j < 1/\varepsilon$ with $\max(|s - j\varepsilon|, |t - j\varepsilon|) < \varepsilon$ and $\max(|Z(s) - Z(j\varepsilon)|, |Z(t) - Z(j\varepsilon)|) \geq \frac{1}{2}\delta$. Consequently,

$$\begin{aligned} P\left(\max_{|s-t| \leq \varepsilon} |Z(s) - Z(t)| \geq \delta\right) &\leq \sum_{j < 1/\varepsilon} P\left(\sup_{j\varepsilon \leq s \leq (j+1)\varepsilon} |Z(s) - Z(j\varepsilon)| \geq \frac{1}{2}\delta\right) \\ &\quad + \sum_{j < 1/\varepsilon} P\left(\sup_{j\varepsilon \leq s \leq (j+1)\varepsilon} |Z(s) - Z((j+1)\varepsilon)| \geq \frac{1}{2}\delta\right) \\ &= O(\varepsilon^{\theta-1} \delta^{-\lambda}). \end{aligned}$$

This proves (3.2) for all ε such that $1/\varepsilon$ is an integer. However, the general case also follows from the O -estimate.

Lemma 3.2. *Let $\Delta := h - k$ and $\bar{Y}_{n,k} := Y_{n,k} - E(Y_{n,k})$. Then, uniformly for $k, h = \log n + o(\log n)$,*

$$E((\bar{Y}_{n,k} - \bar{Y}_{n,h})^2) = O(n^2 \Delta^2 (\log n)^{-3}). \tag{3.3}$$

Proof. In some ranges, we may apply the results of the previous section for the covariance of $Y_{n,k}$ and $Y_{n,h}$, but they do not lead to a uniform estimate, in terms of $|k - h|^2$, in the whole range when $\alpha = \beta = 1$. Here, we instead give a self-contained proof of (3.3).

Assume, without loss of generality, that $h \geq k$. By (2.4), we have

$$E((\bar{Y}_{n,k} - \bar{Y}_{n,h})^2) = ([u^k v^k] - 2[u^k v^h] + [u^h v^h]) f(u, v) n^{u+v} (1 + O(n^{-1})) + O(\delta_{k,h} n (\log n)^{-1/2}).$$

It suffices to find upper bounds for the dominant term

$$J := ([u^k v^k] - 2[u^k v^h] + [u^h v^h]) f(u, v) n^{u+v} = \frac{1}{(2\pi)^2} \iint_{\mathcal{D}} e^{-ikx -iky} (1 - 2e^{-i\Delta y} + e^{-i\Delta(x+y)}) f(e^{ix}, e^{iy}) n^{e^{ix} + e^{iy}} dx dy,$$

where $\mathcal{D} := [-\pi, \pi]^2$. Now,

$$\begin{aligned} 1 - 2e^{-i\Delta y} + e^{-i\Delta(x+y)} &= (1 - e^{-i\Delta y})^2 + e^{-i\Delta y} (e^{-i\Delta x} - 1 + i\Delta x) \\ &\quad - e^{-i\Delta y} (e^{-i\Delta y} - 1 + i\Delta y) + e^{-i\Delta y} (i\Delta y - i\Delta x) \\ &=: Q_1 + Q_2 - Q_3 + Q_4. \end{aligned}$$

Let

$$J_m := \frac{1}{(2\pi)^2} \iint_{\mathcal{D}} Q_m e^{-ikx -iky} f(e^{ix}, e^{iy}) n^{e^{ix} + e^{iy}} dx dy \quad (m = 1, \dots, 4).$$

By the elementary inequalities $|e^{iw} - 1| \leq |w|$ for real w and $1 - \cos w \geq c_5 w^2$ for $|w| \leq \pi$, where $c_5 := 2/\pi^2$, we have

$$|J_1| \leq \frac{n^2 \Delta^2}{(2\pi)^2} \iint_{\mathcal{D}} y^2 |f(e^{ix}, e^{iy})| n^{-c_5(x^2+y^2)} dx dy.$$

This, together with the uniform bound

$$|f(e^{ix}, e^{iy})| = O(|xy|)$$

for $x, y \in \mathcal{D}$, yields

$$\begin{aligned} |J_1| &= \left(n^2 \Delta^2 \iint_{\mathcal{D}} |x||y|^3 n^{-c_5(x^2+y^2)} dx dy \right) \\ &= O(n^2 \Delta^2 (\log n)^{-3}). \end{aligned}$$

Similarly, by the inequality $|e^{iw} - 1 - iw| \leq \frac{1}{2}|w|^2$ for real w , we have

$$|J_m| = O(n^2 \Delta^2 (\log n)^{-3}) \quad (m = 2, 3).$$

For the integral J_4 , we use the expansion

$$f(e^{ix}, e^{iy}) = c_2 i^2 xy + O(|xy||x + y|)$$

and obtain $J_4 = J_5 + J_6$, where

$$\begin{aligned} J_5 &:= \frac{c_2 \Delta}{(2\pi)^2} \iint_{\mathcal{D}} i^3 (y - x)xy \exp\{-ikx -iky -i\Delta y\} n^{e^{ix}+e^{iy}} \, dx \, dy, \\ J_6 &= O\left(n^2 |\Delta| \iint_{\mathcal{D}} |xy|(|x| + |y|)^2 n^{-c_5(x^2+y^2)} \, dx \, dy\right) \\ &= O(n^2 |\Delta| (\log n)^{-3}). \end{aligned}$$

For J_5 , we use the expansion

$$e^{-i\Delta y} = 1 + O(|\Delta||y|)$$

and the relation

$$\iint_{\mathcal{D}} (y - x)xy e^{-ik(x+y)} n^{e^{ix}+e^{iy}} \, dx \, dy = 0$$

(which follows by symmetry), so that

$$\begin{aligned} J_5 &= O\left(n^2 \Delta^2 \iint_{\mathcal{D}} |x||y|^2(|x| + |y|)n^{-c_5(x^2+y^2)} \, dx \, dy\right) \\ &= O(n^2 \Delta^2 (\log n)^{-3}) \end{aligned}$$

uniformly for $k, h = \log n + o(\log n)$. This completes the proof of (3.3).

Lemma 3.3. *Uniformly for $k, h = \log n + o(\log n)$,*

$$|E(Y_{n,k} - Y_{n,h})| = O(n|\Delta|(\log n)^{-1}) \tag{3.4}$$

and, uniformly for $k = \log n + O(1)$ and $h = \log n + o((\log n)^{2/3})$,

$$|E(Y_{n,k}) - E(Y_{n,h})| \sim \frac{n}{\sqrt{2\pi \log n}} \left(1 - \exp\left\{-\frac{(k - h)^2}{2 \log n}\right\}\right). \tag{3.5}$$

Proof. Assume that $|k - \log n| \leq |h - \log n|$. By Cauchy’s integral formula,

$$E(Y_{n,k}) - E(Y_{n,h}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikx} (1 - e^{-i\Delta x}) \frac{n^{e^{ix}}}{\Gamma(1 + e^{ix})} (1 + O(n^{-1})) \, dx.$$

In the first case, when $k, h = \log n + o(\log n)$, we apply the inequality $|1 - e^{-i\Delta x}| \leq |\Delta x|$ and the same arguments as above, yielding

$$|E(Y_{n,k}) - E(Y_{n,h})| = O(|\Delta|n(\log n)^{-1})$$

uniformly in k, h . This proves (3.4).

The approximation (3.5) follows from a straightforward application of the usual saddle-point method.

3.2. An upper bound for the expected width

Let $k_0 = \lfloor \log n \rfloor$, $\Lambda := \lfloor (\log n)^{1/4+\xi_n} \rfloor$, and $L := \lfloor (\log n)^{1/4} \rfloor$, where $\xi_n \in (0, \frac{1}{2})$ will be specified below. We use the following upper bound:

$$\begin{aligned} W_n &\leq \max_{0 \leq |j| \leq L} Y_{n,k_0+j\Lambda} + \max_{|k-h| \leq \Lambda} |\bar{Y}_{n,k} - \bar{Y}_{n,h}| \\ &\quad + \max_{|k-h| \leq \Lambda} |E(Y_{n,k} - Y_{n,h})| + \sum_{|k-k_0| \geq (\log n)^{1/2+\xi_n}} Y_{n,k} \\ &=: W_n^{(1)} + W_n^{(2)} + W_n^{(3)} + W_n^{(4)}. \end{aligned}$$

We show that, when taking expectations, the term Y_{n,k_0} in $W_n^{(1)}$ is dominant and all other terms are asymptotically of smaller order than $E(Y_{n,k_0})$.

We start with $W_n^{(4)}$. By (2.2),

$$\begin{aligned} E(W_n^{(4)}) &= O\left(\sum_{|k-k_0| \geq (\log n)^{1/2+\xi_n}} \frac{(\log n)^k}{k!}\right) \\ &= O(n \exp\{-\frac{1}{2}(\log n)^{2\xi_n}\} (\log n)^{-\xi_n}); \end{aligned}$$

see Hwang (1997). If we choose

$$\xi_n := \frac{\log \log \log n}{\log \log n}$$

then

$$E(W_n^{(4)}) = o(n(\log n)^{-1}).$$

For $W_n^{(3)}$ we have, by (3.4) with $k, h = \log n + O(L\Lambda)$ and by (2.2) with k and h outside this range,

$$\begin{aligned} \max_{|h-k| \leq \Lambda} |E(Y_{n,h} - Y_{n,k})| &= O(n\Lambda(\log n)^{-1}) \\ &= O(n(\log n)^{-3/4+\xi_n}). \end{aligned}$$

We then apply Lemmas 3.1 and 3.2 to prove that

$$E(W_n^{(2)}) = O\left(\frac{n}{\sqrt{\log n}} \left(\frac{\Lambda}{(\log n)^{3/2-\xi_n}}\right)^{1/2}\right). \tag{3.6}$$

To do so, we first define $Y_n(t)$, $-1 \leq t \leq 1$, by

$$Y_n(t) = \bar{Y}_{n,k_0+t(\log n)^{1/2+\xi_n}} \frac{(\log n)^{1-\xi_n}}{n}$$

when $t(\log n)^{1/2+\xi_n}$ is an integer, and by linear interpolation otherwise. By Lemma 3.2, we have

$$E((Y_n(s) - Y_n(t))^2) = O((s - t)^2)$$

uniformly for $s, t \in [-1, 1]$. By Chebyshev’s inequality,

$$P(|Y_n(s) - Y_n(t)| \geq w) = O((s - t)^2 w^{-2}).$$

Let $\eta_n := \Lambda(\log n)^{-1/2-\xi_n}$. It follows, by Lemma 3.1, that

$$P\left(\max_{|s-t|\leq\eta_n} |Y_n(s) - Y_n(t)| \geq w\right) = O(\eta_n w^{-2})$$

and, consequently, that

$$\begin{aligned} E\left(\max_{|s-t|\leq\eta_n} |Y_n(s) - Y_n(t)|\right) &= \int_0^\infty P\left(\max_{|s-t|\leq\eta_n} |Y_n(s) - Y_n(t)| \geq w\right) dw \\ &= O(\eta_n^{1/2}). \end{aligned}$$

This and the definition of $Y_n(t)$ imply (3.6), which can be written as

$$E(W_n^{(2)}) = O(n(\log n)^{-9/8+2\xi_n}).$$

Thus, it remains to find an upper bound for $W_n^{(1)}$. By the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} E(W_n^{(1)}) &\leq E \sum_{|j|\leq L} Y_{n,k_0+j\Lambda} \mathbf{1}_{\{Y_{n,k_0+j\Lambda}=W_n^{(1)}\}} \\ &\leq \sum_{|j|\leq L} E(Y_{n,k_0+j\Lambda}^2)^{1/2} P(Y_{n,k_0+j\Lambda} = W_n^{(1)})^{1/2} \\ &\leq \frac{n}{\sqrt{2\pi \log n}} + O(n(\log n)^{-1}) \\ &\quad + O\left(\frac{n}{\sqrt{\log n}} \sum_{1\leq|j|\leq L} P(Y_{n,k_0+j\Lambda} = W_n^{(1)})^{1/2}\right). \end{aligned}$$

Here we have used the estimates

$$E(Y_{n,k_0}^2)^{1/2} = \frac{n}{\sqrt{2\pi \log n}} + O(n(\log n)^{-1})$$

and $E(Y_{n,k}^2)^{1/2} = O(n/(\log n)^{1/2})$; see Drmota and Hwang (2005).

Now set $D_j := Y_{n,k_0} - Y_{n,k_0+j\Lambda}$ for $1 \leq |j| \leq L$. Then we have

$$\begin{aligned} P(Y_{n,k_0+j\Lambda} = W_n^{(1)}) &\leq P(D_j \leq 0) \\ &\leq P(|D_j - E(D_j)| \geq E(D_j)) \\ &\leq \frac{\text{var}(D_j)}{E(D_j)^2}, \end{aligned}$$

by Chebyshev’s inequality.

By Lemma 3.2, we have

$$\text{var}(D_j) = O\left(\frac{n^2}{(\log n)^3} j^2 \Lambda^2\right).$$

This and (3.5) imply that

$$P(D_j \leq 0)^{1/2} = O\left(\frac{|j|\Lambda}{\log n} \left(1 - \exp\left\{-\frac{j^2 \Lambda^2}{2 \log n}\right\}\right)^{-1}\right)$$

for $1 \leq |j| \leq L$, and it follows that

$$\begin{aligned} \sum_{1 \leq |j| \leq L} P(D_j \leq 0)^{1/2} &= O\left(\frac{\Lambda}{\log n} \int_1^L x \left(1 - \exp\left\{-\frac{x^2 \Lambda^2}{2 \log n}\right\}\right)^{-1} dx\right) \\ &= O(\Lambda^{-1} (\log n)^{2\xi_n}). \end{aligned}$$

Thus,

$$E(W_n^{(1)}) \leq \frac{n}{\sqrt{2\pi \log n}} + O(n(\log n)^{-3/4+\xi_n}).$$

Collecting these estimates gives

$$E(W_n) \leq \frac{n}{\sqrt{2\pi \log n}} (1 + O((\log n)^{-1/4} \log \log n)),$$

which proves (1.7).

3.2.1. *A possible refinement of the error term in (1.7).* If we had the estimates

$$E((\bar{Y}_{n,k} - \bar{Y}_{n,h})^{2m}) = O(n^{2m} \Delta^{2m} (\log n)^{-3m}) \quad (m = 2, 3, \dots),$$

for $k, h \sim \log n$, then the error term $O((\log n)^{-1/4} \log \log n)$ in the approximation to the expected width would be improved to $O((\log n)^{-1/2+\varepsilon})$ for some $\varepsilon > 0$. To prove these moment estimates, we could apply induction and the approach used in Fuchs *et al.* (2005), but the details would be very messy.

3.3. Asymptotics of the level polynomials

The proof of the almost-sure convergence (1.6) follows from the martingale arguments introduced in Chauvin *et al.* (2001). Thus, we only sketch a few steps of the proof here.

We first observe that the normalized random function $\mathcal{M}_n(z) := M_n(z)/E(M_n(z))$ (where, recall, $M_n(z) := \sum_k Y_{n,k} z^k$) is a martingale. Roughly, this reflects the fact that, by construction, the new label has the same probability of being attached to any of the existing nodes. Also, by (2.2),

$$E(M_n(z)) = \binom{n-1+z}{n-1}.$$

By the martingale convergence theorem (see Hall and Heyde (1980)), $\mathcal{M}_n(\alpha)$ converges almost surely to a limit $M(\alpha)$ for any finite, positive α . Then, by the recursive definition (2.1) of $Y_{n,k}$, and similarly to the contraction method (see Fuchs *et al.* (2005)), we deduce that

$$M(\alpha) \stackrel{D}{=} \alpha U^\alpha M(\alpha) + (1-U)^\alpha M(\alpha)^*,$$

where $M(\alpha)^* \stackrel{D}{=} M(\alpha)$, U is a uniform[0, 1] random variable, and $M(\alpha)$, $M(\alpha)^*$, and U are independent. This implies that $M(\alpha) \stackrel{D}{=} Y(\alpha)$ for every $\alpha > 0$.

Interestingly, this limit relation also extends to complex values of α .

Lemma 3.4. *For any compact set in $\{z \in \mathbb{C} : |z - 1| < 1\}$, the martingale $\mathcal{M}_n(z)$ converges almost surely, uniformly, and in L_2 , to its limit $M(z)$ (which is also an analytic function).*

The key step of the proof is to use an explicit expression for $E(M_n(z_1)M_n(z_2))$ (see (1.5)), and to use Kolmogorov’s criterion in combination with vector martingale theorems. Finally,

we recover $Y_{n,k}$ almost surely (and uniformly for $1 - \varepsilon \leq k/\log n \leq 1 + \varepsilon$, for some $\varepsilon > 0$) via Cauchy’s integral formula:

$$\begin{aligned} Y_{n,k} &= \frac{1}{2\pi i} \oint_{|z|=\alpha_{n,k}} M_n(z) z^{-k-1} dz \\ &\sim \frac{1}{2\pi i} \oint_{|z|=\alpha_{n,k}} M(z) E(M_n(z)) z^{-k-1} dz \\ &\sim M(\alpha_{n,k}) \frac{1}{2\pi i} \oint_{|z|=\alpha_{n,k}} E(M_n(z)) z^{-k-1} dz \\ &\sim M(\alpha) E(Y_{n,k}). \end{aligned}$$

We omit all technical details. Note that the radius $\alpha_{n,k} := k/\log n$ in the contour integration is a natural choice because it is the saddle point of the integrand $E(M_n(z)) z^{-k-1}$. Since $M(z)$ is almost surely an analytic function and $M(1) = 1$, it follows that

$$W_n = \max_k Y_{n,k} \sim \max_k E(Y_{n,k}) \sim \frac{n}{\sqrt{2\pi \log n}}$$

almost surely. This completes the proof of (1.6).

3.4. Total path length

Corollary 3.1. *Let T_n denote the total path length in a random recursive tree of n nodes. Then $\mathcal{M}'_n(1)$ is a martingale, and*

$$\mathcal{M}'_n(1) = \frac{T_n - E(T_n)}{n} \xrightarrow{D} Y'(1)$$

almost surely and in L_2 .

Proof. Since $T_n = \sum_k k Y_{n,k}$, we have $\mathcal{M}'_n(1) = (T_n - E(T_n))/n$ by the definition of $M_n(z)$. From Lemma 3.4, it follows that

$$\begin{aligned} \mathcal{M}'_n(1) &= \frac{1}{2\pi i} \int_{|z-1|=\delta < 1} z^{-2} \mathcal{M}_n(z) dz \\ &\rightarrow \frac{1}{2\pi i} \int_{|z-1|=\delta < 1} z^{-2} M(z) dz \\ &= M'(1) = Y'(1) \end{aligned}$$

almost surely.

This result is already known; see Mahmoud (1991) and Dobrow and Fill (1999). However, our approach also gives

$$\mathcal{M}_n^{(m)}(1) \rightarrow M^{(m)}(1) \quad (m \geq 1)$$

almost surely and in L_2 . In particular, when $m = 2$, we have

$$\frac{1}{n} \sum_k k(k-1)(Y_{n,k} - \mu_{n,k}) - \frac{2}{n} E(T_n)(T_n - E(T_n)) \rightarrow M''(1).$$

Note that $\mathcal{M}_n^{(m)}(1)$ is also a martingale for $m \geq 1$.

4. Profile of random binary search trees

In this section, we give the corresponding results for the profiles of random BSTs. The proofs are similar to those for random recursive trees and are thus omitted. Recall that $X_{n,k}$ and $I_{n,k}$ denote the numbers of external nodes and internal nodes, respectively, at level k in a random BST of n elements.

4.1. External nodes

It has been known, since the work of Lynch (1965), that

$$\sum_k E(X_{n,k})u^k = \binom{n + 2u - 1}{n} \quad (n \geq 0);$$

see also Françon (1977) or Mahmoud (1992).

Lemma 4.1. For $n \geq 0$,

$$\begin{aligned} \sum_{k,h} E(X_{n,k}X_{n,h})u^k v^h &= \frac{2uv}{2u + 2v - 2uv - 1} \binom{n + 2u + 2v - 2}{n} \\ &+ \frac{2u + 2v - 4uv - 1}{2u + 2v - 2uv - 1} \binom{n + 2uv - 1}{n}. \end{aligned}$$

This simplifies Lemma 4 in Chauvin *et al.* (2001).

From this lemma we deduce, by singularity analysis (see Flajolet and Odlyzko (1990)), that

$$E(X_{n,k}X_{n,h}) = 2^{k+h} [u^k v^h] \phi(u, v) n^{u+v-2} (1 + O(n^{-1})) + O\left(\delta_{k,h} \frac{(2 \log n)^k}{k! n}\right)$$

uniformly for $\alpha, \beta \in [2 - 2^{1/2} + \varepsilon, 2 + 2^{1/2} - \varepsilon]$, for any $\varepsilon > 0$, where

$$\phi(u, v) := \frac{uv}{(2u + 2v - uv - 2)\Gamma(u + v - 1)} - \frac{1}{\Gamma(u)\Gamma(v)}. \tag{4.1}$$

Theorem 4.1. For $\alpha, \beta \in (2 - 2^{1/2}, 2 + 2^{1/2})$, the correlation coefficient $\rho(X_{n,k}, X_{n,h})$ is asymptotic to

$$\begin{aligned} &\frac{\phi(\alpha, \beta)}{\sqrt{\phi(\alpha, \alpha)\phi(\beta, \beta)}} && \text{if } \alpha, \beta \notin \{1, 2\}, \\ &\frac{\phi'_v(\alpha, \beta)t_{n,h} - \frac{1}{2}\phi''_{v^2}(\alpha, \beta)}{\sqrt{\phi(\alpha, \alpha)p(\beta, \beta; t_{n,h}, t_{n,h})}} && \text{if } \alpha \notin \{1, 2\}, \beta \in \{1, 2\}, \\ &\frac{p(\alpha, \beta; s_{n,k}, t_{n,h})}{\sqrt{p(\alpha, \alpha; s_{n,k}, s_{n,k})p(\beta, \beta; t_{n,h}, t_{n,h})}} && \text{if } \alpha, \beta \in \{1, 2\}, \end{aligned}$$

where

$$p(j, \ell; s, t) := \phi''_{uv}(j, \ell)st - \frac{1}{2}(j\phi'''_{u^2v}(j, \ell)t + \ell\phi'''_{uv^2}(j, \ell)s) + \frac{1}{4}j\ell\phi^{(4)}_{u^2v^2}(j, \ell).$$

Unlike the profile of recursive trees, the limiting correlation coefficients of $\rho(X_{n,k}, X_{n,h})$ undergo sharp sign-changes at $\beta = 1$ and $\beta = 2$; see Figures 5 and 6.

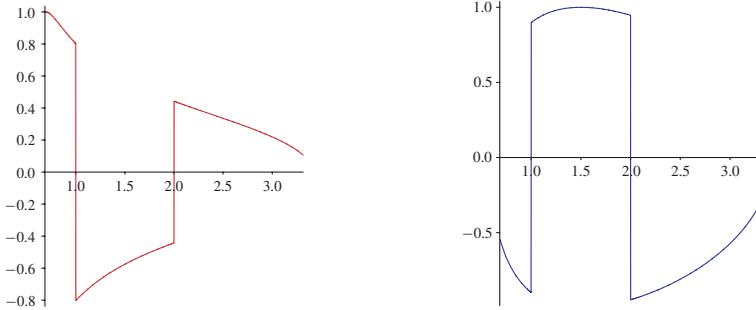


FIGURE 5: The limiting correlation coefficients, plotted against β . There are two sharp sign-changes in $\phi(\alpha, \beta)/[\phi(\alpha, \alpha)\phi(\beta, \beta)]^{1/2}$. Here $\alpha = 0.7$ (left) and $\alpha = 1.5$ (right).

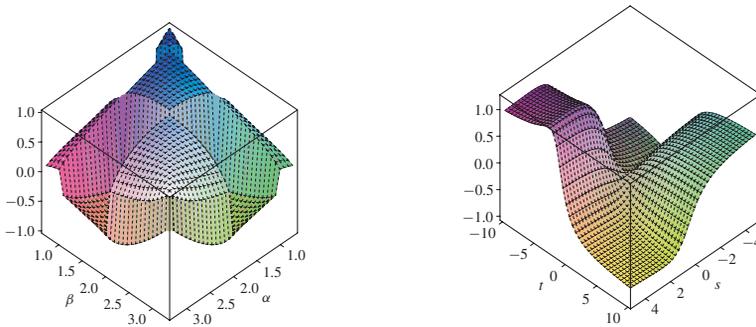


FIGURE 6: Three-dimensional renderings of the limiting correlation coefficients, with $\alpha, \beta \in (2 - 2^{1/2}, 2 + 2^{1/2}) \setminus \{1, 2\}$ (left) and $\alpha = 1, \beta = 2$ (right).

4.1.1. *Width.* The same arguments as above lead to

$$E\left(\max_k X_{n,k}\right) = \frac{n}{\sqrt{4\pi \log n}} \left(1 + O((\log n)^{-1/4} \log \log n)\right).$$

This result is new. The corresponding almost-sure convergence was established in Chauvin *et al.* (2001).

4.2. Internal nodes

For internal nodes, we have

$$E(I_{n,k}) = [u^k] \frac{1 - \binom{n+2u-1}{n}}{1 - 2u} = 2^k [u^k] \frac{1 - \binom{n+u-1}{n}}{1 - u};$$

see Brown and Shubert (1984) or Mahmoud (1992).

Lemma 4.2. For $n \geq 0$,

$$\sum_{k,h} E(I_{n,k}I_{n,h})u^k v^h = \frac{1}{(1-2u)(1-2v)} \left(1 - \binom{n+2u-1}{n} - \binom{n+2v-1}{n} \right) + \frac{2uv}{(1-2u)(1-2v)(2u+2v-2uv-1)} \binom{n+2u+2v-2}{n} - \frac{1}{2u+2v-2uv-1} \binom{n+2uv-1}{n}.$$

From this lemma it follows, again by singularity analysis, that

$$E(X_{n,k}X_{n,h}) = 2^{k+h} [u^k v^h] \varphi(u, v) n^{u+v-2} (1 + O(n^{-1})) + O\left(\frac{(2 \log n)^k}{k!n} + \frac{(2 \log n)^h}{h!n}\right)$$

uniformly for $\alpha, \beta \in [2 - 2^{1/2} + \varepsilon, 2 + 2^{1/2} - \varepsilon]$, for any $\varepsilon > 0$, where

$$\varphi(u, v) := \frac{\phi(u, v)}{(1-u)(1-v)}$$

(ϕ being as defined in (4.1)).

Theorem 4.2. For $\alpha, \beta \in (2 - 2^{1/2}, 2 + 2^{1/2})$, the correlation coefficient $\rho(X_{n,k}, X_{n,h})$ is asymptotic to

$$\begin{aligned} & \frac{\varphi(\alpha, \beta)}{\sqrt{\varphi(\alpha, \alpha)\varphi(\beta, \beta)}} && \text{if } \alpha, \beta \notin \{2\}, \\ & \frac{\varphi'_v(\alpha, 2)t_{n,h} - \frac{1}{2}\varphi''_v(\alpha, 2)}{\sqrt{\varphi(\alpha, \alpha)q(t_{n,h}, t_{n,h})}} && \text{if } \alpha \neq 2, \beta = 2, \\ & \frac{q(s_{n,k}, t_{n,h})}{\sqrt{q(s_{n,k}, s_{n,k})q(t_{n,h}, t_{n,h})}} && \text{if } \alpha = \beta = 2, \end{aligned}$$

where

$$q(s, t) := \varphi'''_{uv}(2, 2)st - (\varphi'''_{u^2v}(2, 2)s + \varphi'''_{uv^2}(2, 2)t) + \varphi^{(4)}_{u^2v^2}(2, 2).$$

Figure 7 depicts the single sign-change of the limiting correlation coefficient $\varphi(\alpha, \beta)/(\varphi(\alpha, \alpha)\varphi(\beta, \beta))^{1/2}$; cf. Figures 5 and 6.

Note that $\varphi(1, 1) = c_2 = 2 - \frac{1}{6}\pi^2$. Thus, $\rho(I_{n,k}, I_{n,h}) \rightarrow 1$ when (i) $k, h \sim \alpha \log n$ with $\alpha \neq 2$, and (ii) $k, h \sim 2 \log n$ and $|k - 2 \log n|, |h - 2 \log n| \rightarrow \infty$.

4.3. An intuitive interpretation of the sign-change

For internal nodes, the behavior and the corresponding intuitive interpretation of the limiting correlation coefficients are similar to those of $Y_{n,k}$ (of recursive trees). The double sign-change of the limit of $\rho(X_{n,k}, X_{n,h})$ is explained roughly as follows. First observe that

$$E(I_{n,k}) \sim \begin{cases} 2^k - \frac{E(X_{n,k})}{1 - \alpha_{n,k}} & \text{if } 1 \leq k \leq \log n - (\log n)^{2/3-\varepsilon}, \\ 2^k \Phi\left(\frac{\log n - k}{\sqrt{\log n}}\right) & \text{if } |k - \log n| \leq (\log n)^{2/3-\varepsilon}, \\ \frac{E(X_{n,k})}{\alpha_{n,k} - 1} & \text{if } k \geq \log n + (\log n)^{2/3-\varepsilon}, \end{cases}$$

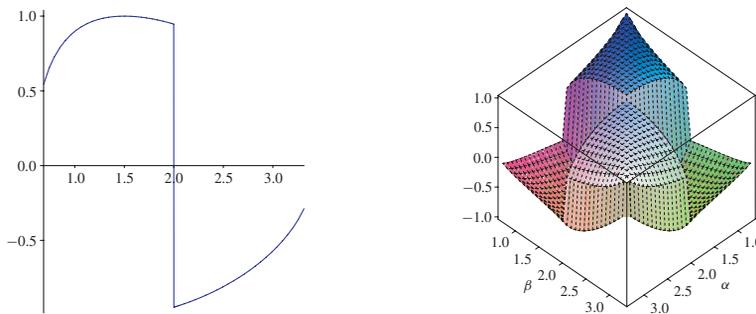


FIGURE 7: The asymptotic correlation coefficient plotted against β with $\alpha = 1.5$ (left), and rendered in three dimensions for $\alpha, \beta \in (2 - 2^{1/2}, 2 + 2^{1/2})$ (right).

for any $\varepsilon > 0$, where $\Phi(x)$ is the standard normal distribution function; see Fuchs *et al.* (2005). This says roughly that the levels up to $(1 - \varepsilon) \log n$ are full of internal nodes (since, in this range, $E(X_{n,k}) = o(2^k)$), with less room for external nodes; outside this range, the number of internal nodes at each level is asymptotically of the same order as that of external nodes. Thus, if $X_{n,k}$ with, say, $\alpha \in (1, 2)$ has more nodes, then there are also more internal nodes at this and neighboring levels, which implies that there are fewer nodes available both at levels less than or equal to $(1 - \varepsilon) \log n$ and at levels greater than or equal to $(2 + \varepsilon) \log n$. This is similar to the interpretation for recursive trees given in the introduction.

5. Conclusions

In this paper, we have discovered a sharp sign-change phenomenon in the correlation coefficients of two level sizes in random recursive trees and random BSTs. Such sign-changes are consistent with the bimodality of the variance in the middle range ($k \sim \log n$ for recursive trees and $k \sim 2 \log n$ for BSTs).

We conclude the paper with a brief comparison of the different approaches we have taken to the study of the variance (and covariance) of profiles. In Drmota and Hwang (2005), the authors introduced two such approaches to $\text{var}(X_{n,k})$ and $\text{var}(Y_{n,k})$, one based on explicit integral representations in terms of Bessel functions and the other on explicit expressions in terms of Stirling numbers of the first kind. However, extending these approaches to $\text{var}(I_{n,k})$ is not easy. In Fuchs *et al.* (2005), the authors used an approach based on recurrence and asymptotic transfer, which applies well to all three profiles discussed in this paper, but to obtain more terms in the asymptotic expansions using this approach requires more effort. The approach we have presented in this paper is not only more general (being applicable to covariance and to more profiles) but is also useful in deriving asymptotic expansions, if necessary. Note that, by the L_2 -convergence of the normalized profiles (established by, say, the contraction method), the leading estimates for the variance or covariance can also be derived from the fixed-point equation of the limit laws. However, this approach fails in the ranges where the limit laws are degenerate.

The major open question here is: what is the limit distribution (if it exists) of the width? Is it the same as the limit law of the total path length (in both classes of random tree considered in this paper)?

Acknowledgements

We thank Luc Devroye for pointing out an error in an earlier version of the paper. Part of the work of the second author was carried out during a visit to the Institut für Stochastik und Mathematische Informatik, J. W. Goethe-Universität (Frankfurt am Main). He thanks the Institute for their hospitality and support; he is also indebted to Tsung-Hsi Tsai for helpful discussions. Furthermore, his work was partially supported by a research award of the Alexander von Humboldt Foundation and by the National Science Council of Taiwan under the grant NSC-92-2118-M-001-019.

References

- ARYA, S., GOLIN, M. J. AND MEHLHORN, K. (1999). On the expected depth of random circuits. *Combinatorics Prob. Comput.* **8**, 209–228.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- BOOTH, A. D. AND COLIN, J. T. (1960). On the efficiency of a new method of dictionary construction. *Inf. Control* **3**, 327–334.
- BROWN, G. G. AND SHUBERT, B. O. (1984). On random binary trees. *Math. Operat. Res.* **9**, 43–65.
- CHAUVIN, B., DRMOTA, M. AND JABBOUR-HATTAB, J. (2001). The profile of binary search trees. *Ann. Appl. Prob.* **11**, 1042–1062.
- DOBROW, R. P. AND FILL, J. A. (1999). Total path length for random recursive trees. *Combinatorics Prob. Comput.* **8**, 317–333.
- DRMOTA, M. AND HWANG, H.-K. (2005). Bimodality and phase transitions in the profile variance of random binary search trees. To appear in *SIAM J. Discrete Math.*
- FLAJOLET, P. AND ODLYZKO, A. M. (1990). Singularity analysis of generating functions. *SIAM J. Discrete Math.* **3**, 216–240.
- FRANÇON, J. (1977). On the analysis of algorithms for trees. *Theoret. Comput. Sci.* **4**, 155–169.
- FUCHS, M., HWANG, H.-K. AND NEININGER, R. (2005). Profiles of random trees: limit theorems for random recursive trees and binary search trees. Submitted. Available at <http://algo.stat.sinica.edu.tw/>.
- GASTWIRTH, J. L. (1977). A probability model of a pyramid scheme. *Amer. Statistician* **31**, 79–82.
- GROSSMAN, R. AND LARSON, R. G. (1989). Hopf-algebraic structure of families of trees. *J. Algebra* **126**, 184–210.
- HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York.
- HIBBARD, T. N. (1962). Some combinatorial properties of certain trees with applications to searching and sorting. *J. Assoc. Comput. Mach.* **9**, 13–28.
- HWANG, H.-K. (1995). Asymptotic expansions for the Stirling numbers of the first kind. *J. Combinatorial Theory A* **71**, 343–351.
- HWANG, H.-K. (1997). Asymptotic estimates for elementary probability distributions. *Studies Appl. Math.* **99**, 393–417.
- KNUTH, D. E. (1997). *The Art of Computer Programming*, Vol. 3, *Sorting and Searching*, 2nd edn. Addison-Wesley, Reading, MA.
- LYNCH, W. C. (1965). More combinatorial properties of certain trees. *Comput. J.* **7**, 299–302.
- MAHMOUD, H. M. (1991). Limiting distributions for path lengths in recursive trees. *Prob. Eng. Inf. Sci.* **5**, 53–59.
- MAHMOUD, H. M. (1992). *Evolution of Random Search Trees*. John Wiley, New York.
- MEIR, A. AND MOON, J. W. (1974). Cutting down recursive trees. *Math. Biosci.* **21**, 173–181.
- MITCHELL, S. L., COCKAYNE, E. J. AND HEDETNIEMI, S. T. (1979). Linear algorithms on recursive representations of trees. *J. Comput. System Sci.* **18**, 76–85.
- MOON, J. W. (1974). The distance between nodes in recursive trees. In *Combinatorics* (London Math. Soc. Lecture Notes Ser. **13**), eds T. P. McDonough and V. C. Marron, Cambridge University Press, pp. 125–132.
- NA, H. S. AND RAPOPORT, A. (1970). Distribution of nodes of a tree by degree. *Math. Biosci.* **6**, 313–329.
- TAPIA, M. A. AND MYERS, B. R. (1967). Generation of concave node-weighted trees. *IEEE Trans. Circuits Systems* **14**, 229–230.
- VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. AND VAN MIEGHEM, P. (2002). On the covariance of the level sizes in random recursive trees. *Random Structures Algorithms* **20**, 519–539.
- WINDLEY, P. F. (1960). Trees, forests and rearranging. *Comput. J.* **3**, 84–88.