2005 International Conference on Analysis of Algorithms DMTCS proc. **AD**, 2005, 167–172

The profile of unlabeled trees

Bernhard Gittenberger[1]†

[1] *Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, A-1040 Wien, Austria.* gittenberger@dmg.tuwien.ac.at

We consider the number of nodes in the levels of unlabeled rooted random trees and show that the joint distribution of several level sizes (where the level number is scaled by $\sqrt{n}$) weakly converges to the distribution of the local time of a Brownian excursion evaluated at the times corresponding to the level numbers. This extends existing results for simply generated trees and forests to the case of unlabeled rooted trees.

Keywords: unlabeled trees, profile, Brownian excursion, local time

1 Introduction

We consider the profile of unlabeled rooted random trees, sometimes called Pólya trees, because the enumeration theory of Pólya (1937) allows an analytical treatment of this class of trees by means of generating functions. The profile is defined as the stochastic process built of the level sizes of a random tree, where a level is a set of nodes at a fixed distance from the root. The distance from the root, scaled by $\sqrt{n}$ serves as time index of this process.

The first investigations of the profile of random trees seem to go back to Stepanov (1969) who derived explicit formulas for the distribution of the size of one level. Further papers deal mainly with simply generated trees as defined by Meir and Moon (1978). Kolchin (see Kolchin (1977, 1986)) related the level size distributions to distributions occurring in particle allocation schemes. Later Takács (1991) derived another expression for the level sizes by means of generating functions. Aldous (1991) conjectured two functional limit theorems for the profile in two different ranges which were proved in Drmota and Gittenberger (1997); Gittenberger (1998). Drmota (1997) studied restrictions of the profile to nodes of fixed degree. An extension to random forests of simply generated trees is given by Gittenberger (2002).

Later other tree classes have been considered as well. The profile of random binary search trees has been first studied by Chauvin et al. (2001) and later by Drmota (2004) and Drmota and Hwang (2005a). Random recursive trees have been investigated recently by Drmota and Hwang (2005b) and van der Hofstad et al. (2002).

The plan of the proof is as follows. We will first set up the generating functions of the trees where nodes in certain levels are marked. This function is given as solution of a recurrence relation which has to be analyzed in detail. Then we are able to show that the finite dimensional distributions (fdd's) of the profile, i.e., the distributions of the sizes of several levels considered simultaneously, converge to the fdd's of Brownian excursion local time.

We want to remark that for a functional limit theorem, as it is available for the other tree classes mentioned above, the convergence of the fdd's is not sufficient. Therefore it would be necessary to establish tightness (see Billingsley (1968) of the process. This requires to estimate certain moments of the differences of the sizes of two given levels and could be done by the same approach. But the generating functions appearing here are very involved and hence we are not able to solve this problem at the moment.

2 Preliminaries and Results

First we collect some results for unlabeled unrooted trees. Let $\mathcal{Y}_n$ denote the set of unlabeled rooted trees of size n and y_n be the cardinality of this set. Pólya (1937) already discussed the generating function

$$y(z) = \sum_{n\geq 1} y_n z^n$$

†This work has been supported by the Austrian Science Foundation FWF, grant P16053-N05.

and showed that the radius of convergence ρ satisfies $0 < \rho < 1$ and that $z = \rho$ is the only singularity on the circle of convergence $|z| = \rho$. He also showed that $y(z)$ satisfies the functional equation

$$y(z) = z \exp\left(\sum_{i\geq 1} \frac{y(z^i)}{i}\right).$$

Later Otter (1948) showed that $y(\rho) = 1$ as well as the asymptotic expansion

$$y(z) = 1 - b(\rho - z)^{1/2} + c(\rho - z) + d(\rho - z)^{3/2} + \cdots \tag{1}$$

which he used to deduce that

$$y_n \sim \frac{b\sqrt{\rho}}{2\sqrt{\pi}} n^{-3/2} \rho^{-n}.$$

Furthermore he calculated the first constants appearing in this expansion: $\rho \approx 0.3383219$, $b \approx 2.6811266$, and $c = b^2/3 \approx 2.3961466$.

Let $L_n(t)$ denote the number of nodes at distance t from the root of a unlabeled rooted tree of size n. If t is not an integer, then define $L_n(t)$ by linear interpolation:

$$L_n(t) = (\lfloor t\rfloor + 1 - t)L_n(\lfloor t\rfloor) + (t - \lfloor t\rfloor)L_n(\lfloor t\rfloor + 1), \quad t \geq 0.$$

We will show the following theorem.

Theorem 1 *Let*

$$l_n(t) = \frac{1}{\sqrt{n}} L_n\left(t\sqrt{n}\right)$$

and $l(t)$ denote the local time of a standard scaled Brownian excursion. Then we have for any d and any choice of fixed numbers $t_1, \ldots, t_d$ the following limit theorem:

$$(l_n(t_1), \ldots, l_n(t_d)) \xrightarrow{w} \frac{1}{b\sqrt{\rho}} \left(\left(\frac{b}{2\sqrt{2}} \cdot t_1\right), \ldots, \left(\frac{b}{2\sqrt{2}} \cdot t_d\right)\right),$$

as $n \to \infty$, where b is the constant of Equation (1).

3 Combinatorial Setup

In order to compute the distribution of the number of nodes in some given levels in a tree of size n we have to calculate the number $Y(k_1 m_1 k_2 m_2 \cdots k_d m_d n)$ of trees of size n with m_i nodes in level k_i, $i = 1, \ldots, d$ and normalize by y_n.

Therefore we introduce the functions $y(k|z, u, s)$, where s may depend on z and u, be given by the recurrence relation

$$\begin{aligned} y(0|z, u, s) &= us \\ y(k+1|z, u, s) &= z \exp\left(\sum_{i\geq 1} \frac{y(k|z^i, u^i, s)}{i}\right), \quad k \geq 0. \end{aligned} \tag{2}$$

Then we have

$$\begin{aligned} y(k_1, \ldots, k_d|z, u_1, \ldots, u_d) &= \sum_{m_1,\ldots,m_d,n\geq 0} y(k_1 m_1 k_2 m_2 \cdots k_d m_d n) u_1^{m_1} \cdots u_d^{m_d} z^n \\ &= y(k_1|\,(z, u_1, y(k_2 - k_1|\,(z, u_2, \ldots y(k_d - k_{d-1}|\,(z, u_d, y(z)) \ldots\,) \end{aligned}$$

Note that for $d = 1$ we have $y(k|z, u) = y(k|z, u, y(z))$. The characteristic function of the joint distribution of $\frac{1}{\sqrt{n}} L_n(k_1), \ldots, \frac{1}{\sqrt{n}} L_n(k_d)$ is then given by

$$\phi_{k_1 \cdots k_d n}(t_1, \ldots, t_d) = \frac{1}{y_n} [z^n] y(k_1, \ldots, k_d| \left(z, e^{it_1/\sqrt{n}}, \ldots e^{it_d/\sqrt{n}}\right).$$

This coefficient will be calculated by singularity analysis (see Flajolet and Odlyzko (1990)). Hence the following lemma will be the crucial step of the proof. Note that we will use $y(k|z, u, s)$ as generating functions for certain tree families consisting of unlabeled trees with some nodes marked (in particular, all nodes in a certain level). Thus we will make this restriction in the sequel.

Lemma 1 *Set* $w = us - y(z)$ *and* $z = \rho\left(1 + \frac{x}{n}\right)$. *Assume that* $|w| = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ *and* $z - \rho \to 0$ *in such a way that* $\arg(z-\rho) \neq 0$ *and*

$$1 - \frac{C_2}{\sqrt{n}} \leq \left|1 - \sqrt{1 - \frac{z}{\rho}}\right| \leq 1 + \frac{C}{\sqrt{n}}.$$

Furthermore, assume that s *is either* $y(z)$ *or* $y(k_1, \dots, k_d|z, u_1, \dots, u_d)$ *for some* d. *Then* $y(k|z,u,s)$ *admits the local representation*

$$y(k|z,u,s) = y(z) + \frac{y(z)^k w}{1 - \frac{1-y(z)^k}{2(1-y(z))}w + \mathcal{O}\left(\left|\frac{1-y(z)^{2k}}{1-y(z)^2}\right| |w|^2\right)} \tag{3}$$

uniformly for $k = \mathcal{O}(\sqrt{n})$.

The proof is presented in the next section.

4 The Local Behaviour of $y(k|\cdot)$

Note that for all choices of the initial condition s in (2) which are relevant in this context, the sequence $y(k|z,u,s)$ tends to $y(z)$, as k tends to infinity. (Of course we have to make precise the allowed range of z and u.) Hence we will first derive a-priori estimates for the differences

$$w_k(z,u) = y(k|z,u,s) - y(z)$$

and then proceed by bootstrapping.

First observe that if $s = y(z)$ (the other cases behave similarly) we have

$$\begin{aligned} |w_k(z,u)| &= \left|\sum_n \sum_{m<n} Y_{kmn}(u^m - 1)z^n\right| \\ &\leq 2\sum_n \sum_{m<n} Y_{kmn}|zu|^n - 2\sum_n Y_{k0n}|z|^n \\ &= 2y(\geq k|\,|zu|) \end{aligned} \tag{4}$$

where $y(\geq k|z)$ denote the generating function enumerating trees with height at least k.

Lemma 2 *Consider the space* Ω *of all functions* $u(z)$, *which are analytic in* $B := \{z \in \mathbf{C} : |z| \leq \rho^2 + \varepsilon\}$ *and satisfy* $u(0) = 0$, *equipped with the norm*

$$\|u\| = \max_{z \in B}\left|\frac{u(z)}{z}\right|.$$

Then the operator T *defined by*

$$Tu = z\exp\left(\sum_{i\geq 1}\frac{u(z^i)}{i}\right)$$

is a contraction for functions in the set $\{u \in \Omega : \|u\| < 3\}$.

Remark *Note that since* $y(z)/z$ *has only positive coefficients, we have* $\|y(z)\| \leq y(\rho)/\rho < 3$ *and hence it will be sufficient that* T *is a contraction for functions with norm bounded by* $(1+\varepsilon)\|y\|$.

Proof: We have

$$\begin{aligned} \left|\frac{Tu - Tv}{z}\right| &= \left|\exp\left(\sum_{i\geq 1}\frac{u(z^i)}{i}\right) - \exp\left(\sum_{i\geq 1}\frac{v(z^i)}{i}\right)\right| \\ &\leq \left|\sum_{i\geq 1}\frac{u(z^i) - v(z^i)}{i}\right| \exp\left(\max\left(\left|\sum_{i\geq 1}\frac{u(z^i)}{i}\right|, \sum_{i\geq 1}\left|\frac{v(z^i)}{i}\right|\right)\right) \\ &\leq \|u - v\|\sum_{i\geq 1}\left|\frac{z^i}{i}\right| \exp\left(\sum_{i\geq 1}\left|\frac{z^i}{i}\right| \max\left(\|u\|, \|v\|\right)\right) \\ &\leq L\|u - v\| \end{aligned}$$

where L can be chosen less than 1, since $\max(\|u\|, \|v\|) < 3$ and

$$\sum_{i\geq 1} \left|\frac{z^i}{i}\right| \leq \log \frac{1}{1-(\rho^2+\varepsilon)} < 0.13$$

for sufficiently small ε. □

Note that

$$y(\geq k|z) = y(z) - y(< k|z) = T^{k-1}y(z) - T^{k-1}z.$$

Hence $\|y(\geq k|\cdot)\| \leq L^{k-1}\|y(z) - z\|$ and thus we get the bound

$$|w_k(z^i, u^i)| \leq CL^k |zu|^i \tag{5}$$

where C can be chosen less than $1/2$. This gives rise to the next a priori estimate.

Lemma 3 *Suppose that* $w_0 = \mathcal{O}(1/\sqrt{n})$, $|y| = 1 + \Theta(1/\sqrt{n})$, *and* $k = \mathcal{O}\left(|w_0|^{-1}\right)$. *Then*

$$w_i = \mathcal{O}\left(w_0 y^i\right).$$

Proof: The proof runs by induction. Assume $w_k \leq (1+C_0|w_0|)$ and note that since we have in the range for z and u which is given by the assumptions of the lemma $|y(k|z,u)| \leq y(|z|(1+\varepsilon)$, we immediately get that $ze^{y+\theta_k w_k}$ and $z\exp\left(\frac{y(z^i)+\theta_{ki}w_k(z^i,u^i)}{i}\right)$ are uniformly bounded if θ_k and θ_{ki} lie in the unit interval. Therefore the Taylor expansion

$$\begin{aligned} y(k+1|z,u) &= z\exp\left(y(z) + w_k(z,u) + \sum_{i\geq 2} \frac{y(z^i) + w_k(z^i,u^i)}{i}\right) \\ &= y(z)\exp\left(w_k(z,u) + \sum_{i\geq 2}\frac{w_k(z^i,u^i)}{i}\right) \end{aligned} \tag{6}$$

yields

$$\begin{aligned} |w_{k+1}(z,u)| = &\left| y(z)w_k(z,u) + \frac{y(z)}{2} + \mathcal{O}\left(w_k^3\right) + \mathcal{O}\left(\sum_{i\geq 2}\frac{w_k(z^i,u^i)}{i}\right)\right| \\ &(1 + C_1|w_0| + C_2|w_k|)\,|w_k| \\ &\leq \left(1 + C_1|w_0| + C_2(1+C_0|w_0|)^k|w_0|\right)(1+C_0|w_0|)^k|w_0| \end{aligned}$$

Hence $|w_k| \leq (1+C_0|w_0|)^k|w_0| = \mathcal{O}\left(w_0 y^i\right)$ and the proof is complete. □

Now we continue with bootstrapping. By expanding (6) into a Taylor series and using (5) we get

$$w_{k+1} = yw_i + \frac{y}{2}w_k^2 + \mathcal{O}\left(|w_k|^3\right) + \mathcal{O}\left(L^k\right).$$

This recursion can be solved asymptotically by setting $q_k = \frac{y^k}{w_k}$ which yields

$$q_{k+1} = q_k - \frac{y}{2}y^{k-1} + \mathcal{O}\left(|w_k||y|^k\right) + \mathcal{O}\left(\frac{L^k|y|^k}{|w_k|^2}\right)$$

Since $L^k = \mathcal{O}\left(|w_k|^3\right)$, this implies

$$q_k = \frac{1}{w_0} - \frac{1}{2}\frac{1-y^k}{1-y} + \mathcal{O}\left(|w_0|\left|\frac{1-y^{2k}}{1-y^2}\right|\right)$$

and this implies (3).

5 Finite Dimensional Distributions

For $d = 1$ we have

$$\phi_{k,n}(t) = \frac{1}{y_n}[z^n]y(k|\left(z, e^{it/\sqrt{n}}, y(z)\right)$$
$$= \frac{1}{2\pi i y_n}\int_\Gamma y(k|\left(z, e^{is/\sqrt{n}}, y(z)\right)\frac{dz}{z^{n+1}} \quad (7)$$

where the contour $\Gamma = \gamma \cup \Gamma_4$ consists of a truncated Hankel contour $\gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ around the singularity (at distance $1/\sqrt{n}$ closed by a circular arc Γ_4).

The contribution of Γ_4 turns out to be small and if $z \in \gamma$, then the local expansion (3) is valid. Furthermore we have with $z = \rho\left(1 + \frac{x}{n}\right)$ and $k = \kappa\sqrt{n}$ the following expansions:

$$\left(e^{it/\sqrt{n}} - 1\right)y(z) \sim \frac{it}{\sqrt{n}}$$
$$1 - y(z) \sim b\sqrt{-\frac{x}{n}}$$
$$y(z)^k \sim \exp\left(-b\kappa\sqrt{-x}\right)$$

and hence

$$y(k|\left(z, e^{it/\sqrt{n}}, y(z)\right) = y(z) + w_k\left(z, e^{it/\sqrt{n}}\right)$$

where

$$w_k = \frac{1}{\sqrt{n}}\frac{it\sqrt{-x}\exp\left(-\frac{b\kappa}{2}\sqrt{-x}\right)}{\sqrt{-x}\exp\left(-\frac{b\kappa}{2}\sqrt{-x}\right) - \frac{it}{b\sqrt{\rho}}\sinh\left(\frac{b\kappa}{2}\sqrt{-x}\right)}.$$

Insertion into (7) yields the characteristic function of the distribution of $l(b\kappa/2\sqrt{2}\,)/b\sqrt{\rho}$ as desired.

Now we can proceed by induction. We have for $d = 2$

$$\phi_{k,k+h,n}(t_1, t_2) = \frac{1}{y_n}[z^n]y(k|\left(z, e^{it_1/\sqrt{n}}, y(h|\left(z, e^{it_2/\sqrt{n}}, y(z)\right)\right)$$

and

$$y(k|\left(z, e^{it_1/\sqrt{n}}, y(h|\left(z, e^{it_2/\sqrt{n}}, y(z)\right)\right) = y(z) + \tilde{w}_k$$

where $\tilde{w}_k$ can be estimated similarly by application of Lemma 1. This step can be repeated easily for $d > 2$ and in this way we get the characteristic functions of the fdd's of $l((b/2\sqrt{2})\cdot t)/b\sqrt{\rho}$ as desired.

References

D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *Ann. Appl. Probab.*, 11 (4):1042–1062, 2001.

M. Drmota. On nodes of given degree in random trees. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1996)*, pages 31–44. VSP, Utrecht, 1997.

M. Drmota. Profile and height of random binary search trees. *J. Iranian Statistical Society*, 3(2):117–138, 2004.

M. Drmota and B. Gittenberger. On the profile of random trees. *Random Structures Algorithms*, 10(4): 421–451, 1997.

M. Drmota and H. K. Hwang. Bimodality and phase transitions in the profile variance of random binary search trees. *SIAM J. Discr. Math.*, 2005a. To appear.

M. Drmota and H. K. Hwang. Profile of random trees: correlation and width of random recursive trees and binary search trees. *Adv. Appl. Prob.*, 2005b. Submitted.

P. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

B. Gittenberger. Convergence of branching processes to the local time of a Bessel process. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 423–438, 1998.

B. Gittenberger. On the profile of random forests. In *Mathematics and computer science, II (Versailles, 2002)*, Trends Math., pages 279–293. Birkhäuser, Basel, 2002.

V. F. Kolchin. Branching processes, random trees and a generalized particle distribution scheme. *Mat. Zametki*, 21(5):691–705, 1977.

V. F. Kolchin. *Random mappings*. Translation Series in Mathematics and Engineering. Optimization Software Inc. Publications Division, New York, 1986.

A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

R. Otter. The number of trees. *Ann. Math.*, 49(2):583–599, 1948.

G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68:145–254, 1937.

V. E. Stepanov. The distribution of the number of vertices in the layers of a random tree. *Teor. Verojatnost. i Primenen.*, 14:64–77, 1969.

L. Takács. Conditional limit theorems for branching processes. *J. Appl. Math. Stochastic Anal.*, 4(4): 263–292, 1991.

R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. On the covariance of the level sizes in random recursive trees. *Random Structures & Algorithms*, 20(4):519–539, 2002.