On the Width of Labeled Trees

Jean-François Marckert
Université de Nancy 1

April 12, 1999

[summary by Christine Fricker]

Abstract

We consider A_n the set of all rooted labeled trees with n nodes. We denote by Z_i the number of nodes at distance i from the root and by $W_n = \max_{0 \leq i \leq n} Z_i$ the width of the tree. The aim of the talk is to present results on convergence of moments of W_n (correctly renormalized) to those of the maximum of the normalized Brownian excursion and to give a tight bound for the rate of convergence. For the proof, the connections between especially breadth first search random walk on trees, random walk with Poissonian increment, parking function and empiric process of mathematical statistics are described.

The results presented in this talk were obtained jointly with P. Chassaing.

1. Introduction

A rooted labeled tree with n nodes is a connected graph with n vertices and $n-1$ edges where a vertex, the root, is specified. Tight bounds of the width W_n of rooted labeled trees with n nodes are given, answering an open question of Odlyzko and Wilf [3] that $E(W_n)$ is between $C_1\sqrt{n}$ and $C_2\sqrt{n \log n}$. More precisely, we first present the result of Takacs [6] that $W_n/\sqrt{n}$ converges in distribution to the maximum m of the Brownian excursion, with the well-known theta distribution given by

$$\Pr(m \leq x) = \sum_{k \in \mathbb{Z}} (1 - 4k^2x^2)e^{-2k^2x^2}.$$

However weak convergence does not answer completely the question of Odlyzko and Wilf. To fill this gap, we prove that

Theorem 1. *For all $p \geq 1$, $\left|E(W_n^p/n^{p/2}) - E(m^p)\right| \leq C_p n^{-1/4}\sqrt{\log n}$ where m is the maximum of the normalized Brownian excursion.*

The moments of m are well-known and given by

$$E(m^p) = p(p-1)\Gamma(p/2)\zeta(p)2^{-p/2}.$$

For this we prove that there exists a sequence of normalized Brownian excursions of maximum m_n such that, for all $p \geq 1$,

$$E(\left|W_n/\sqrt{n} - m_n\right|^p) \leq C'_p n^{-p/4}\sqrt{\log n}$$

using that if q is defined by $1/p + 1/q = 1$ and if X and Y are two real random variables in L_p, then by Holder's inequality,

$$|E(X^p) - E(Y^p)| \leq p\left\|X - Y\right\|_p \left\|X + Y\right\|_p^{p/q}.$$

2. Relation Between Rooted Labeled Trees and Parking Functions

In hashing with linear probing or parking, we consider n cars c_i $(1 \leq i \leq n)$ arriving in this order at random in $n+1$ places $\{0, 1, \dots, n\}$, where car c_i is parking on its place h_i if h_i is still empty, otherwise car c_i is trying places $h_{i+1} \bmod n+1, \dots$. We consider parking functions, i.e. sequences $(h_i)_{1 \leq i \leq n}$ such that place n is empty. A parking function is alternatively characterized by the sequence $(A_k = \{i, h_i = k\})_{0 \leq k \leq n}$ of sets of cars that arrive on place k, with $x_k = \operatorname{card} A_k$. If y_k is the number of cars that tried once to park on place k,

$$y_k = y_{k-1} - 1 + x_k, \qquad y_0 = x_0.$$

The fact that place n is the empty place is given by

$$y_k \geq 1 \quad (0 \leq k \leq n-1), \qquad y_n = 0$$

or equivalently

$$(1) \qquad \sum_{i=0}^{k} x_i - k \geq 1 \quad (0 \leq k \leq n-1), \qquad \sum_{i=0}^{n} x_i - n = 0.$$

A labeled tree with vertices $\{0, 1, \dots, n\}$ rooted at 0 is also characterized by the sequence of disjoint sets $(A_k)_{0 \leq k \leq n}$ whose union is $\{1, \dots, n\}$ and the $x_k = \operatorname{card} A_k$ satisfying (1). Indeed, A_k $(k \geq 1)$ (respectively A_0) is defined as the set of new neighbors of the smallest element in A_{k-1} (respectively 0) and (1) is the condition for the tree to be connected and to have root 0. The number of A_k $(k \geq 1)$ with cardinality x_k $(k \geq 1)$ is proportional to the product of Poisson probabilities $e^{-1}/x_k!$. In other words the corresponding unlabeled tree is a Galton-Watson tree with Poisson(1) progeny, constrained to have $n+1$ nodes. Thus, the sequence $y = (y_k)_{0 \leq k \leq n}$ is the discrete excursion with length n of a random walk with increments $x_k - 1$. It is well-known that $(y_{\lfloor nt \rfloor}/\sqrt{n})_{0 \leq t \leq 1}$ converges in distribution to $(e(t))_{0 \leq t \leq 1}$ where e is a normalized Brownian excursion and $\max_k y_k/\sqrt{n}$ converges in distribution to $m = \max_{0 \leq t \leq 1} e(t)$, which is theta-distributed.

The random walk (y_k) (introduced in [1] and [5]) gives the profile of the tree (Z_k) as a subsequence of (y_k)

$$Z_{k+1} = y_{l(k)}$$

where $l(k) = \sum_{i=1}^{k} Z_i$ and the width W_n as

$$W_n = \max_k Z_k = \max_k y_{l(k)}.$$

We prove in the following proposition that $W_n = \max_k y_{l(k)}$ has the same behavior as $\max_k y_k$.

Proposition 1. *For each $p \geq 1$,*

$$\|W_n - \max_k y_k\|_p = O(n^{1/4}(\log n)^{3/4}).$$

This result is based on the slow variation of the sequence $y = (y_k)_{0 \leq k \leq n}$. Indeed, $\Omega_c(n)$ defined as the set of sequences $y = (y_k)_{0 \leq k \leq n}$ such that, for all k and m such that $k + m \leq n$,

$$|y_{m+k} - y_m| \leq c\sqrt{k \log n}$$

satisfies the following lemma.

Lemma 1. *For all $\alpha > 0$, there exists $c > 0$ such that, for all n,*

$$1 - \Pr(\Omega_c(n)) = o(n^{-\alpha}).$$

This can be proved using (see Petrov [4]) that, if (Y_k) is a random walk with increments X_k satisfying $E(X_k) = 0$ and for some $\alpha > 0$, $E(\exp(\alpha|X_k|) < \infty$, then there exists $T, C_1, C_2 > 0$ such that

$$\Pr(|Y_k| \geq x) \leq \begin{cases} 2e^{-\frac{x^2}{4C_1}} & \text{if } 0 \leq x \leq C_1 T, \\ 2e^{-C_2 x} & \text{if } x \geq C_1 T. \end{cases}$$

Then it remains to prove that $E((\max_k y_k/\sqrt{n})^p) \to E(m^p)$ and to estimate the rate of convergence. This is the object of the next section.

3. Parking Functions and Empiric Processes

Consider the sequence $(U_i)_{1\leq i\leq n}$ of n i.i.d. random variables uniformly distributed on $[0,1]$. Let $F_n(t)$ be the empiric distribution for $(U_i)_{1\leq i\leq n}$ i.e.

$$F_n(t) = \frac{\text{card}\{i \in \{1,\ldots,n\}, U_i \leq t\}}{n}, \qquad (0 \leq t \leq 1).$$

Process $(F_n(t))$ converges to $(F(t)) = (t)$, the distribution function of the uniform distribution. Especially, the empiric process $(\alpha_n(t)) = (\sqrt{n}(F_n(t) - F(t)))_{0\leq t\leq 1}$ converges in distribution to the Brownian bridge $(b(t))_{0\leq t\leq 1}$.

There is a precise connection between parking functions and empiric processes. Indeed, consider the sequence $(U_i)_{1\leq i\leq n}$ of i.i.d. random variables uniformly distributed on $[0,1]$ and realize parking in the following way: If $U_i \in \left[\frac{k-1}{n+1}, \frac{k}{n+1}\right[$, then car c_i tries to park first on place $h_i = k$. The last empty place V is given in terms of the empiric process.

Proposition 2. *There is a unique $T(n)$ in $\{0,1,\ldots,n\}$ such that*

$$\alpha_n(\frac{T(n)}{n+1}) = \min_{1\leq j\leq n} \alpha_n(\frac{j}{n+1}).$$

Moreover, $T(n) = V$.

It is easy to deduce that

$$\left|\frac{\max_k y_k}{\sqrt{n}} - (\sup_{0\leq t\leq 1} \alpha_n(t) - \inf_{0\leq t\leq 1} \alpha_n(t))\right| \leq \frac{1+2\epsilon_n}{\sqrt{n}}$$

where $\epsilon_n = \sqrt{n}\sup_{0\leq t\leq 1} |\alpha_n(\frac{\lfloor (n+1)t\rfloor}{n+1}) - \alpha_n(t)|$ satisfies the following proposition.

Proposition 3. *There exists A, C and K such that for all x and n,*

$$\Pr(\epsilon_n \geq C\log n + x) \leq An^{1-KC}e^{-Kx}. \tag{2}$$

Then $\alpha_n(t)$ is replaced by a Brownian bridge $b_n(t)$ using the following result of Komlos, Major and Tusnady [2].

Theorem 2. *There exists a sequence of Brownian bridges $(b_n)_{n\geq 1}$ and A, M, $\mu > 0$ such that for all n and x*

$$\alpha_n(t) = b_n(t) + \frac{c_n(t)}{\sqrt{n}}$$

where $C_n = \sup_{0\leq t\leq 1} |c_n(t)|$ verifies for all x

$$\Pr(C_n \geq A\log n + x) \leq Me^{-\mu x}. \tag{3}$$

Then

$$\left|\frac{\max_k y_k}{\sqrt{n}} - (\sup_{0\le t\le 1} b_n(t) - \inf_{0\le t\le 1} b_n(t))\right| \le \frac{1+2(\epsilon_n + C_n)}{\sqrt{n}}$$

where C_n is introduced in Theorem 2. Using the fact that, if T is the almost surely unique point such that $b(T) = \min_{0\le t\le 1} b(t)$, then $e = (e(t))_{0\le t\le 1}$, defined by $e(t) = b((T+t) \bmod 1) - b(T)$, is a normalized Brownian excursion independent of T, one has

$$\left|\frac{\max_k y_k}{\sqrt{n}} - \sup_{0\le t\le 1} e_n(t)\right| \le \frac{1+2(\epsilon_n + C_n)}{\sqrt{n}}. \tag{4}$$

Relations (2) and (3) give that $\|\epsilon_n + C_n\|_p$ is bounded by $K_p \log n$ and thus (4) gives the following result.

Theorem 3. *For each $p \ge 1$,*

$$\left\|m_n - \frac{\max_k y_k}{\sqrt{n}}\right\|_p = O(\frac{\log n}{\sqrt{n}}).$$

It is then easy to deduce Theorem 1.

Bibliography

[1] Aldous (David). – Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability*, vol. 25, n° 2, 1997, pp. 812–854.

[2] Komlós (J.), Major (P.), and Tusnády (G.). – An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 34, n° 1, 1976, pp. 33–58.

[3] Odlyzko (Andrew M.) and Wilf (Herbert S.). – Bandwidths and profiles of trees. *Journal of Combinatorial Theory. Series B*, vol. 42, n° 3, 1987, pp. 348–370.

[4] Petrov (V. V.). – *Sums of independent random variables.* – Springer-Verlag, New York-Heidelberg, 1975, x+346p. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[5] Spencer (Joel). – Enumerating graphs and Brownian motion. *Communications on Pure and Applied Mathematics*, vol. 50, n° 3, 1997, pp. 291–294.

[6] Takács (Lajos). – Limit distributions for queues and random rooted trees. *Journal of Applied Mathematics and Stochastic Analysis*, vol. 6, n° 3, 1993, pp. 189–216.