Parking functions, empirical processes, and the width of rooted labeled trees

Philippe Chassaing
Institut Elie Cartan
Vandoeuvre, France
chassain@iecn.u-nancy.fr

Jean-François Marckert
Université de Versailles St-Quentin en Yvelines
Versailles, France
marckert@math.uvsq.fr

Submitted: August 31, 1999; Accepted: February 8, 2001.
MR Subject Classifications: 05C05, 60J65, 60J80, 62G30

Abstract

This paper provides tight bounds for the moments of the width of rooted labeled trees with n nodes, answering an open question of Odlyzko and Wilf (1987). To this aim, we use one of the many one-to-one correspondences between trees and parking functions, and also a precise coupling between parking functions and the empirical processes of mathematical statistics. Our result turns out to be a consequence of the strong convergence of empirical processes to the Brownian bridge (Komlós, Major and Tusnády, 1975).

Key words. Rooted labeled trees, moment, width, Brownian excursion, empirical processes, hashing with linear probing, parking.

1 Introduction

An order $n+1$ *labeled tree* is a connected graph with set of vertices $\{0, 1, 2, 3, ..., n\}$, and with n edges. If we specify one vertex to be the root, we have a *rooted* labeled tree. According to Cayley (1889) the number of such trees is $(n+1)^n$.

For τ chosen at random in the set of order $n+1$ rooted labeled trees, let $G_k^{(n)}(\tau)$ denote the number of nodes at distance k from the root of τ, and let $H_n(\tau)$ denote the maximum distance of a node from the root, the *height* of τ ; $(G_k^{(n)})_{k\geq 0}$ is the *profile* of the tree. The *width* $W_n(\tau)$ is defined by

$$W_n = \max_{0\leq k\leq H_n} G_k^{(n)}.$$

Odlyzko and Wilf (1987) used a Perron-Frobenius-like theory to derive asymptotics for the cumulative function of W_n. They also proved that

$$C_1\sqrt{n} \le E(W_n) \le C_2\sqrt{n\log n},$$

and left the first term in the asymptotic of $E(W_n)$ as an open question.

Let $\ell(t)$ denote the local time of the normalized Brownian excursion $e(.)$ at level t, i.e.

$$\ell(t) = \lim_{\varepsilon\to 0_+} \frac{1}{\varepsilon}\int_0^1 I_{[t,t+\varepsilon]}\left(e(u)\right)\, du.$$

Aldous [1] conjectured that $t \longmapsto G^{(n)}_{\lfloor t\sqrt{n}\rfloor}/\sqrt{n}$ would converge weakly, as a stochastic process, to $t \longmapsto \ell(t)/2$. Aldous's conjecture was settled by Drmota and Gittenberger [9]. As noted by these last authors, their result entails the weak convergence of $W_n/\sqrt{n}$ to the maximum m of the Brownian excursion, as $\ell(t)$ is itself a Brownian excursion changed of time [5]. Previously, the weak convergence of $W_n/\sqrt{n}$ to m was proven directly by Takács (1993).

However weak convergence does not answer completely the question of Odlyzko & Wilf, as it does not yield convergence of the first moment, and even less the speed of this convergence. The aim of our paper is to fill this gap. Our proof uses the breadth first search (BFS) random walk [3, 27], following Takács [28], who used the BFS random walk to prove convergence of moments of the width for binary trees, or general unlabeled trees, by a clever use of the ballot theorem. For rooted labeled trees, we need an additional ingredient: a close connection between rooted labeled trees and *empirical processes* of mathematical statistics [26], which, we believe, has interest in itself. For instance, this connection gives an alternative $O(n)$ algorithm, for the generation of a random rooted labeled tree, to the $O(n)$ algorithm using Prüfer-Knuth's correspondence (see [16, 20]). It also allows to analyze the size of parking blocks during the phase transition [7]. Note that Aldous, or Drmota and Gittenberger's results are actually about *general* simple trees. Rooted labeled trees are a special case of simple trees, but an important one [16, 20].

Recall [5, 8, 15] that the maximum m of the Brownian excursion satisfies

$$\begin{aligned}\Pr(m \le x) &= \sum_{-\infty<k<+\infty} (1-4k^2x^2)e^{-2k^2x^2},\\ E(m) &= \sqrt{\frac{\pi}{2}},\end{aligned}$$

and, for $r > 1$,

$$E(m^r) = 2^{-r/2} r(r-1)\Gamma\left(\frac{r}{2}\right)\zeta(r).$$

We shall say that m is *theta-distributed* by reference to Jacobi's Theta function. Incidentally, it is also well known that theta-distributed random variables occur as a limit for the height of trees: see Rényi and Szekeres (1967) for rooted labeled trees, Flajolet and Odlyzko (1982) for general simple trees.

Let us state the main result of this paper:

Theorem 1.1 *For $p \geq 1$,*

$$E\left(n^{-p/2}W_n^p\right) - E(m^p) = O_p\left(n^{-1/4}\sqrt{\log n}\right).$$

As a special case,

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

One of the motivations of Odlyzko and Wilf, when they study the width of labeled trees, is to give a tight estimate for the average bandwidth of this class of tree.

2 The breadth first search random walk

From now on, we assume, without consequences for $W_n(\tau)$'s distribution, that τ is drawn at random in the subset Ω_n of labeled trees *rooted at* 0. The BFS of the rooted labeled tree starts with the root, 0, and is implemented by maintaining a *queue* Q, that is initially (0). Then, at each of the n following stages of the BFS, the vertex x at the head of the queue is removed from the queue, and all "new" neighbors of x are added at the end of the queue, in increasing order. At step 0, the search produces the set A_0 of neighbors

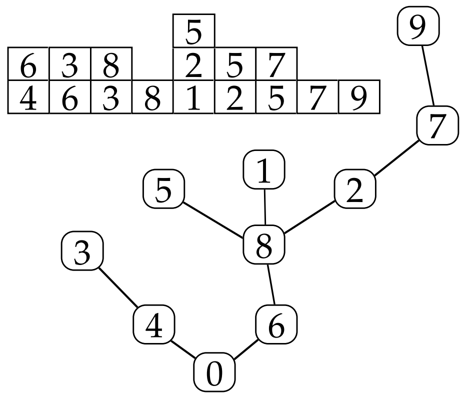

Figure 1: Successive states of the queue.

of vertex 0, so that after step 0 the queue contains exactly the elements of A_0, but not 0 anymore. At step 1, the search produces the set A_1 of new neighbors of the smallest element x in A_0, so that after step 1 the queue contains $A_0 \cup A_1 - \{x\}$. Let A_k denote the set of new elements in the queue after step k, and let

$$a_k = \#A_k.$$

A labeled tree τ with vertices $\{0, 1, 2, 3, ..., n\}$, rooted at 0, is described by a sequence of disjoint sets $(A_i)_{0 \leq i \leq n}$, whose union is $\{1, 2, ..., n\}$, and whose cardinalities $a_i = \#A_i$ satisfy the following set of constraints

$$a_0 \geq 1,$$

$$\begin{array}{rcl} a_0 + a_1 - 1 & \geq & 1, \\ & .... & \\ a_0 + a_1 + ... + a_k - k & \geq & 1, \\ & .... & \\ a_0 + a_1 + ... + a_{n-1} - n + 1 & \geq & 1, \\ a_0 + a_1 + ... + a_n - n & = & 0. \end{array} \tag{2.1}$$

Constraints (2.1) are necessary and sufficient conditions for a tree to be connected, or for the queue to become empty only after step n.

We call *BFS random walk* the sequence $y^{(n)} = \left(y_k^{(n)}(\tau)\right)_{0 \leq k \leq n}$ of queue lengths: $y_k^{(n)}(\tau)$ denotes the number of vertices in the queue after step k, defined by $y_0^{(n)} = a_0$ and

$$\begin{array}{rcl} y_k^{(n)} & = & a_0 + a_1 + ... + a_k - k, \\ y_k^{(n)} - y_{k-1}^{(n)} & = & a_k - 1. \end{array}$$

The proof of Theorem 1.1 relies on the expression of the profile and of the width of the tree in term of the BFS random walk: observe that $G_1^{(n)} = y_0^{(n)}$, $G_2^{(n)} = y_{G_1^{(n)}}^{(n)}$. More generally, at step $G_1^{(n)} + G_2^{(n)} + ... + G_k^{(n)}$, we explore the last vertex at a distance k from the root, and the queue contains exactly the vertices at distance $k+1$ from the root, leading to

$$G_{k+1}^{(n)} = y_{G_1^{(n)} + G_2^{(n)} + ... + G_k^{(n)}}^{(n)}.$$

Actually, this is Kendall's embedding of a Galton-Watson process in the process of queue lengths, when studying a single-server queue [23].

Thus W_n is the maximum of a sample of $y_i^{(n)}$. Due to slow variation of the sequence $(y_k^{(n)})_{0 \leq k \leq n}$, this sample turns out to be "representative", in the sense that the maximum of the sample is close to the maximum of the whole sequence.

Proposition 2.1 *For any $p \geq 1$*

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

The proof is given in the next Section. In Section 4, we use a connection between labeled trees and empirical processes, more easily explained with the help of parking functions, to prove the next Proposition.

Proposition 2.2 *In some probability space, there exists a sequence m_n of theta-distributed random variables and a sequence of copies of $y^{(n)}$ such that, for any $p \geq 1$,*

$$\|\max_k y_k^{(n)} - m_n\sqrt{n}\|_p = O_p(\log n).$$

As a consequence, we have

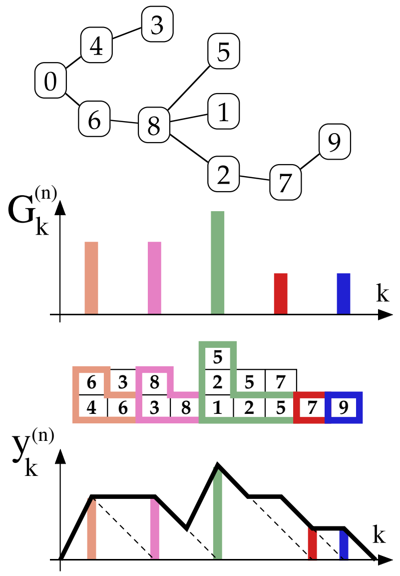

Figure 2: Embedding of the profile in the BFS random walk.

Proposition 2.3 *In some probability space, there exists a sequence m_n of theta-distributed random variables and a sequence of copies of W_n such that, for any $p \geq 1$,*

$$\left\| \frac{W_n}{\sqrt{n}} - m_n \right\|_p = O_p\left(n^{-1/4} (\log n)^{1/2} \right).$$

Then

$$\begin{aligned}\left| E\left[\left(\frac{W_n}{\sqrt{n}} \right)^p \right] - E(m^p) \right| &\leq p \max\left(\left\| \frac{W_n}{\sqrt{n}} \right\|_p, \|m\|_p \right)^{p-1} \left| \left\| \frac{W_n}{\sqrt{n}} \right\|_p - \|m_n\|_p \right| \\ &= O_p\left(n^{-1/4} (\log n)^{1/2} \right),\end{aligned}$$

leading to Theorem 1.1.

3 Proof of Proposition 2.1

The number of n-tuples $(A_i)_{0 \leq i \leq n}$ with cardinalities $(a_i)_{0 \leq i \leq n}$,

$$\frac{n!}{a_0! a_1! \ldots a_n!},$$

is proportional to the product of Poisson probabilities $e^{-1}/a_i!$, so, if a labeled tree τ, rooted at 0, is drawn at random, the corresponding sequence $(a_i(\tau))_{0 \leq i \leq n}$ has the distribution of independent Poisson random variables with mean value 1, *conditioned to satisfy constraints* (2.1) (see Spencer (1997)). In other words, the corresponding unlabeled tree is a Galton-Watson tree with Poisson(1) progeny, constrained to have $n+1$ nodes, and A_k is the progeny of the k^{th} node visited by the BFS.

As a consequence, the sequence $y^{(n)} = (y_k^{(n)})_{0\le k\le n}$ is a random walk with length n and i.i.d. increments $a_i - 1$, conditioned to satisfy (2.1). Set

$$M_n = \max_k y_k^{(n)}.$$

The aim of this section is to bound the difference between M_n and W_n. Essentially, we follow the line of proof of [28, formula 63, page 200], but we improve Takács's bounds with the help of Petrov's Theorem 3.2. Let $x \vee y$ denote the maximum of x and y, and let $\Omega_\delta(n)$ be the set of sequences $y = (y_k)_{k=0,\dots,n}$ that satisfy

$$|y_{m+k} - y_m| \le \delta \left(\log n \vee \sqrt{k \log n}\right)$$

whenever $k \ge 0$, $m \ge 0$ and $m + k \le n$. We have

Proposition 3.1 *Given any positive number α there exists a constant $\kappa(\alpha)$, not depending on n, such that*

$$\Pr\left(y^{(n)} \notin \Omega_{\kappa(\alpha)}(n)\right) = o_\alpha(n^{-\alpha}).$$

Proof. Let $(N_k)_{0\le k\le n}$ be a sequence of independent random variables with mean 1, Poisson-distributed, and let $t = (t_k)_{0\le k\le n}$ be the random walk with increments $N_k - 1$. Let $\Delta(n)$ denote the set of sample paths y that satisfy constraints (2.1). As a consequence of Spencer's key remark,

$$\begin{aligned}\Pr(y \notin \Omega_\delta(n)) &= \Pr(t \notin \Omega_\delta(n) \mid t \in \Delta(n)) \\ &\le \frac{\Pr(t \notin \Omega_\delta(n))}{\Pr(t \in \Delta(n))}.\end{aligned}$$

According to Otter's formula [23], we have

$$\Pr(t \in \Delta(n)) = \frac{1}{n}\Pr(t_n = 0),$$

so due to the standard local limit theorem [11, Ch. 4, Th. 4.2.1] we obtain

$$\Pr(t \in \Delta(n)) = \Theta(n^{-3/2}).$$

Thus we are to prove Proposition 3.1 only for the unconditioned random walk t, but this is a consequence of the next Theorem [22, p.52-55].

Theorem 3.2 *(Petrov, 1975) Let Y_k be a random walk with i.i.d. increments X_k satisfying simultaneously*

- *$E(X_k) = 0$, and*

- *for some positive constant* α, $E(e^{\alpha|X_k|}) < +\infty$,

then:

i) there exists two positive real constants g *and* T *such that*

$$E(\exp(\lambda X_1)) \leq \exp(g\lambda^2) \text{ for } |\lambda| < T,$$

ii) for $(Y_k)_{k\geq 1}$ *defined as above, we have*

$$\begin{aligned}
\Pr(|Y_k| \geq x) &\leq 2\exp\left(-\frac{x^2}{4kg}\right) \quad \text{if } 0 \leq x \leq kgT, \\
&\leq 2\exp\left(-\frac{Tx}{2}\right) \quad \text{if } x \geq kgT.
\end{aligned}$$

For $\delta \geq gT$, Theorem 3.2 yields

$$\begin{aligned}
\Pr(t \notin \Omega_\delta(n)) &\leq \Pr\left(\exists m, k \mid |t_{m+k} - t_m| \geq \delta\left(\log n \vee \sqrt{k\log n}\right)\right) \\
&\leq n\sum_{k=1}^{n} \Pr\left(|t_k| \geq \delta\left(\log n \vee \sqrt{k\log n}\right)\right) \\
&\leq 2n \sum_{k=1}^{\frac{\delta^2\log n}{T^2g^2}} n^{-\delta T/2} + 2n \sum_{k=\frac{\delta^2\log n}{T^2g^2}}^{n} n^{-\delta^2/4g} \\
&\leq \frac{2\delta^2\log n}{T^2g^2} n^{1-\delta T/2} + 2n^{2-\delta^2/4g}.
\end{aligned}$$

For δ large enough, the last term is $o_\alpha(n^{-\alpha})$. ◇

For the end of the proof of Proposition 2.1, recall that $G_i^{(n)} = y_{m(i)}$, in which $m(1) = 0$ and $m(i+1) = m(i) + G_i^{(n)}$. Consider an integer k such that $y_k = M_n$: for some i, $k \in [m_i, m_{i+1}[$, so that

$$\begin{aligned}
0 \leq M_n - W_n &\leq M_n - G_i^{(n)} \\
&\leq \delta\left(\log n \vee \sqrt{(k - m(i))\log n}\right) I_{\Omega_\delta(n)} + n\left(1 - I_{\Omega_\delta(n)}\right) \\
&\leq \delta\left(\log n \vee \sqrt{G_i^{(n)}\log n}\right) I_{\Omega_\delta(n)} + n\left(1 - I_{\Omega_\delta(n)}\right) \\
&\leq \delta\left(\log n + \sqrt{M_n\log n}\right) I_{\Omega_\delta(n)} + n\left(1 - I_{\Omega_\delta(n)}\right) \qquad (3.2) \\
&\leq \delta\left(\log n + \sqrt{\delta\sqrt{n}\log^{3/2} n}\right) I_{\Omega_\delta(n)} + n\left(1 - I_{\Omega_\delta(n)}\right).
\end{aligned}$$

Thus, owing to Proposition 3.1, for a suitable choice of δ,

$$\begin{aligned}
E\left(|W_n - M_n|^p\right) &\leq \delta^p\left(\log n + \sqrt{\delta\sqrt{n}\log^{3/2} n}\right)^p + n^p \Pr\left(y^{(n)} \notin \Omega_{\delta(p)}(n)\right) \\
&= O_p\left(n^{p/4}(\log n)^{3p/4}\right).
\end{aligned}$$

This last estimate holds true under hypothesis of finite exponential moments for the progeny. Actually, to obtain a complete proof of Proposition 2.1, we need to decrease the exponent of $\log n$ from $3p/4$ to $p/2$. In the special case of labeled trees (Poisson progeny), we shall prove at the end of the next Section, as a consequence of the DKW inequality for empirical processes, that

Lemma 3.3 *For $p \geq 1$, $E(M_n^{p/2}) = O_p\left(n^{p/4}\right)$.*

For a suitable choice of δ, relation (3.2) and Lemma 3.3 yield Proposition 2.1.

4 Proof of Proposition 2.2

4.1 Rooted labeled trees and parking functions

As $y^{(n)}$ is distributed like a random walk with i.i.d. increments conditioned on first return to 0 being at time n (cf. (2.1)), it rescales to Brownian excursion:

$$\left(\frac{y^{(n)}_{\lfloor nt\rfloor}}{\sqrt{n}}\right)_{0\leq t\leq 1} \overset{weakly}{\longrightarrow} (e(t))_{0\leq t\leq 1},$$

and thus

$$\frac{\max_k y_k^{(n)}}{\sqrt{n}} \overset{weakly}{\longrightarrow} m = \max_{0\leq t\leq 1} e(t).$$

In this section we prove the more demanding convergence of moments, through a coupling labeled trees-empirical processes more easily explained through parking functions.

A first correspondence between parking functions and acyclic functions was discovered by Schützenberger (1968). The description of the equivalent connection between labeled trees rooted at 0 and parking functions, through the BFS random walk, is more convenient for our purpose. In hashing with linear probing, or parking [13, 17], we consider the case with n cars and $n+1$ places $\{0, 1, 2, ..., n\}$, car c_k parking on place p_k if p_k is still empty, that is, if a car with a smaller index did not park on place p_k before. Otherwise car c_k tries places $(p_k + 1) \bmod n+1$, $(p_k + 2) \bmod n+1$, ..., until it finds an empty place. We consider *parking functions* (resp. *confined* sequences) in the terminology of [14] (resp. of [13, 17]), that is sequences $\omega = (p_k)_{1\leq k\leq n}$ such that the last empty place is place n. Such a parking function ω is alternatively characterized by the sequence $\left(\tilde{A}_i(\omega)\right)_{0\leq i\leq n}$, where

$$\tilde{A}_i(\omega) = \{k \mid p_k = i\}$$

is the set of cars whose first try is place i.

Let $\tilde{a}_i(\omega)$ denote $\#\tilde{A}_i(\omega)$, and let $\tilde{y}_k^{(n)}(\omega)$ denote the number of cars that tried, successfully or not, to park on place k. For $k \neq 0$, we have

$$\begin{aligned}\tilde{y}_k^{(n)} &= \tilde{y}_{k-1}^{(n)} - 1 + \tilde{a}_k \\ &= \tilde{a}_0 + \tilde{a}_1 + ... + \tilde{a}_k - k,\end{aligned}$$

since either place $k-1$ is occupied by car c_i and, among the $\tilde{y}_{k-1}^{(n)}$ cars that visited place $k-1$, only c_i won't visit place k, or place $k-1$ is empty: only $k-1=n$, $k=0$, belongs to this last case, and clearly

$$\tilde{y}_0^{(n)} = \tilde{a}_0.$$

So a sequence $(\tilde{A}_i)_{0\le i\le n}$ is associated with a confined parking scheme if and only if $(\tilde{a}_i)_{0\le i\le n}$ satisfies the constraints (2.1), since a place k is empty only if $\tilde{y}_k^{(n)}(\omega)=0$.

Finally, observing that each of the $(n+1)^{n-1}$ sequences $(\tilde{A}_i)_{0\le i\le n}$ that satisfies (2.1) defines simultaneously a unique parking function (confined sequence) ω for n cars on $n+1$ places and a unique order $n+1$ labeled tree $\tau(\omega)$ rooted at 0, we obtain

Proposition 4.1 *There exists a one-to-one correspondence $\omega \to \tau(\omega)$ between parking functions and trees, such that for any k and ω*

$$y_k^{(n)}\left(\tau(\omega)\right) = \tilde{y}_k^{(n)}(\omega).$$

As a consequence, note that if $D(n+1,n)$ denotes the *total displacement* of cars, we have

$$\begin{aligned} D(n+1,n) &= -n + \sum_{k=0}^{n} y_k^{(n)} \\ &= -n + (n+1)^{3/2} \int_0^1 \frac{y^{(n)}_{\lfloor (n+1)t \rfloor}}{\sqrt{n+1}} dt. \end{aligned}$$

Thus

$$n^{-3/2} D(n+1,n) \quad \overset{weakly}{\longrightarrow} \quad \int_0^1 e(t)dt,$$

and we recover here *partly* the convergence of moments of the total displacement towards the moments of the Airy law, already obtained by Flajolet et al. [13]: the Airy law is known as the law of the area below the Brownian excursion. At Subsection 4.5 we shall complete this alternative proof with the help of the connection parking functions – empirical processes.

4.2 Empirical processes

Consider a sequence of independent random variables $(U_i)_{i\ge 1}$, each of them uniform on $[0,1]$. Let $F_n(t)$ denote the *empirical distribution function* for $(U_i)_{1\le i\le n}$, defined for $t \in [0,1]$ by

$$F_n(t) = \frac{\#\{i \,|\, 1 \le i \le n \text{ and } U_i \le t\}}{n}.$$

We recall a few facts about the convergence of the empirical distribution function towards the distribution function $F(t)=t$ of the uniform law [26]. The speed of convergence of many interesting statistics is revealed by the *empirical process*

$$\alpha_r(t) = \sqrt{r}(F_r(t) - F(t)),$$

that satisfies

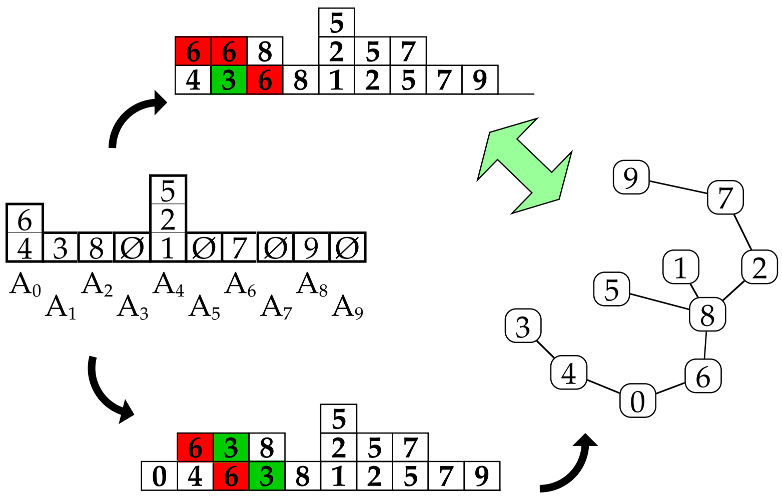

Figure 3: Correspondence trees $\leftrightarrow$ parking.

Theorem 4.2 *(Donsker, 1952)*

$$(\alpha_r(t))_{t\in[O,1]} \stackrel{weakly}{\longrightarrow} (b(t))_{t\in[O,1]},$$

$b(t)$ being the Brownian bridge.

Thus the first error term is of order $O(1/\sqrt{r})$. The second error term is given by the following Theorem of "strong convergence":

Theorem 4.3 *(Komlós, Major & Tusnády, 1975) Given U_1, U_2, .. uniform on $[0,1]$ and independent, there exists a sequence $(b_n)_{n\geq 1}$ of Brownian bridges such that for all n and x,*

$$\Pr\left(\sup_{0\leq t\leq 1}|\alpha_n(t)-b_n(t)| \geq \frac{A\log n + x}{\sqrt{n}}\right) \leq Me^{-\mu x},$$

where A, M and μ are positive absolute constants.

Equivalently, we can write

$$F_n(t) = F(t) + \frac{b_n(t)}{\sqrt{n}} + \frac{r_n(t)}{n},$$

in which $r_n(t)$ denotes $\sqrt{n}\,(\alpha_n(t)-b_n(t))$, and satisfies

$$\Pr\left(\sup_{0\leq t\leq 1}|r_n(t)| \geq A\log n + x\right) \leq Me^{-\mu x}.$$

KMT's Theorem is the last ingredient we need to estimate $\|W_n\|_p$.

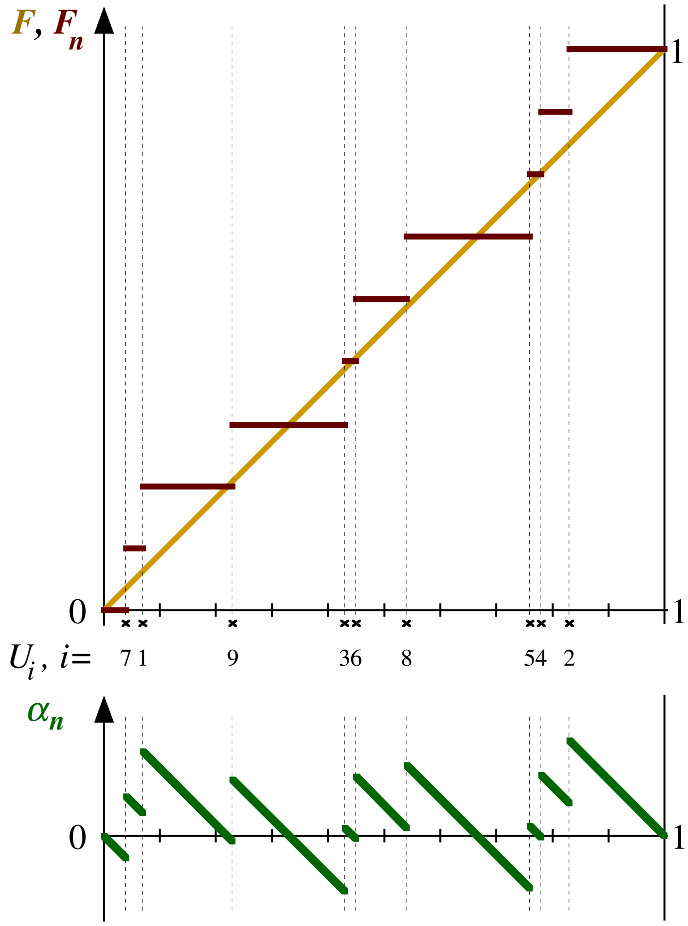

Figure 4: Empirical distribution F_n, empirical process α_n.

4.3 Parking functions and empirical processes

Let $(U_i)_{1\leq i\leq n}$ denote a sequence of i.i.d. random variables, each of them uniform on $[0,1]$, and let the first try of car c_i be at place

$$p_i = \lceil (n+1)U_i \rceil\,,$$

assuming that place $n+1$ is also place 0. Let D_i denote the set of cars whose first try is place i, set $d_i = \#D_i$, and let $z_k^{(n)}$ denote the number of cars that tried, successfully or not, to park on place k. Let $V(\omega)$ denote the last empty place.

Compared with Subsection 4.1, we have some changes: the "parking" functions, or hashing sequences, $\omega = (k \to p_k, \quad 0 \leq k \leq n)$, are not confined anymore, and there are now $(n+1)^n$ such functions ω, clearly equiprobable ; $V(\omega)$ is not n anymore: V is now random uniform on $\{0, 1, 2, ..., n\}$.

Let α_n be the empirical process of $(U_1, U_2, ..., U_n)$. We have

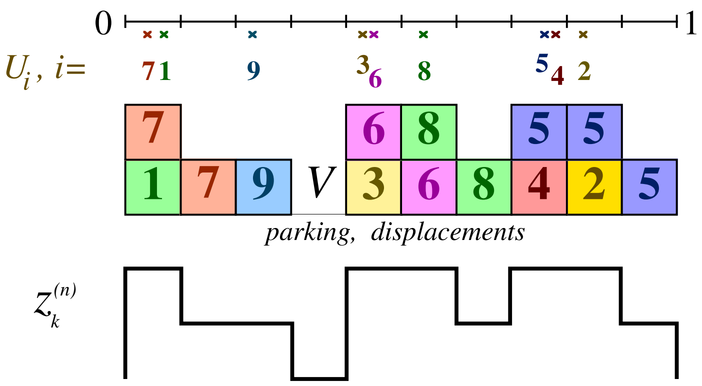

Figure 5: BFS random walk associated to $(U_1, U_2, \ldots, U_n)$

Proposition 4.4 *The relation*

$$\alpha_n\left(\frac{T(n)}{n+1}\right) = \min_{0\le k\le n} \alpha_n\left(\frac{k}{n+1}\right),$$

defines a unique number $T(n)$ *between* 0 *and* n. *Furthermore,*

$$T(n) = V.$$

As a consequence, $T(n)$ is uniformly distributed on $\{0, 1, 2, \ldots, n\}$. Also, the empty place V does not depend on the chronology (the D_i's), but only on the sequence $(d_i)_{0\le i\le n}$, since we have

$$\alpha_n\left(\frac{k}{n+1}\right) = \frac{1}{\sqrt{n}}\left(d_1 + d_2 + \ldots + d_k - k\,\frac{n}{n+1}\right).$$

Proof : Set $\theta(n,i) = \sqrt{n}\,\alpha_n(i/n+1)$. For $0 \le i < j \le n+1$, $\theta(n,i) = \theta(n,j)$ can occur only if $(i-j)\frac{n}{n+1}$ is an integer, i.e. if $(i,j) = (0, n+1)$, as the fractional parts of $\theta(n,j) - \theta(n,i)$ and $(i-j)\frac{n}{n+1}$ are the same: the number of cars whose first try belongs to $\{i+1, i+2, \ldots, j\}$ is given by

$$d_{i+1} + d_{i+2} + \ldots + d_j = \theta(n,j) - \theta(n,i) - (i-j)\frac{n}{n+1}.$$

Thus $i \longmapsto \alpha_n(i/n+1)$ reaches its minimum only once in $\{0, 1, 2, \ldots, n\}$, and $T(n)$ is well defined. For $k = 1, 2, \ldots, n+1$, we have

$$\begin{aligned}\theta(n, V+k) - \theta(n, V) &= d_{V+1} + d_{V+2} + \ldots + d_{V+k} - k\frac{n}{n+1} \\ &= z^{(n)}_{V+k} + k - 1 - k\frac{n}{n+1} \\ &= z^{(n)}_{V+k} + \frac{k}{n+1} - 1, \qquad (4.3)\end{aligned}$$

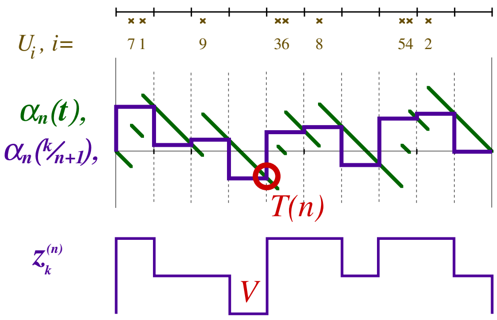

Figure 6: The empirical process and the empty place.

the second equality, as already seen in Subsection 4.1, due to the fact that $z^{(n)}_{V+1} = d_{V+1}$, but for $k \neq V$, $z^{(n)}_{k+1} = z^{(n)}_k - 1 + d_{k+1}$. Finally, for $k = 1, 2, ..., n$, $z^{(n)}_{V+k} \geq 1$ so the last term is positive, that is, $k \to \theta(n, k)$ reaches its minimum at point V. ♢

Proposition 4.4 yields a surprisingly precise coupling between the sequence $z^{(n)}$ and the empirical process α_n associated with $(U_i)_{1\leq i\leq n}$: for $0 \leq k \leq n$, set

$$w^{(n)}_k = \frac{n-k}{n+1} + \sqrt{n}\left[\alpha_n\left(\left\{\frac{k+1+T(n)}{n+1}\right\}\right) - \alpha_n\left(\frac{T(n)}{n+1}\right)\right],$$

and let $w^{(n)} = \left(w^{(n)}_k\right)_{0\leq k\leq n}$. As a byproduct of (4.3), we obtain

Corollary 4.5

$$\left(z^{(n)}_{V+1+k}\right)_{0\leq k\leq n} = w^{(n)}.$$

Now, if we define, for $\omega = (p_k)_{1\leq i\leq n}$,

$$T\omega = (1 + p_k)_{1\leq i\leq n},$$

we observe that the sequence $\hat{y}^{(n)} = \left(z^{(n)}_{V(\omega)+k+1}(\omega)\right)_{0\leq k\leq n}$ is invariant under T, while

$$V(T\omega) = 1 + V(\omega).$$

It follows that V is uniform and independent of $\hat{y}^{(n)}$, so that, on one hand, the conditional distribution of $\hat{y}^{(n)}$ given that $V = n$ is the same as the unconditional distribution of $\hat{y}^{(n)}$. On the other hand, the conditional distribution of $\hat{y}^{(n)}$ given that $V = n$ is the distribution of the sequence $z^{(n)}$ under the hypothesis of equiprobability of *confined* sequences, that is, the distribution as the sequence $\tilde{y}^{(n)}$ of Subsection 4.1. Finally,

Proposition 4.6 *The BFS random walk $y^{(n)}$ satisfies*

$$y^{(n)} \overset{(law)}{=} w^{(n)}.$$

This connection between BFS random walks and empirical processes is close in spirit to a coding of parking functions given page 14 of [14], and the correspondence trees-parking schemes of Subsection 4 is close to the one explained ibidem page 17. This explicit coupling also reminds of similarities between the Cayley tree function, or the Borel distribution [4, Section 2.2] in one hand, and expressions omnipresent in [26, Chap. 9] about empirical processes, in the other hand (see Exercice 1, p. 345 or formulas of Birnbaum & Pyke p. 386). After a short digression, we explain in the last subsection how Proposition 4.6, together with KMT Theorem, yields Proposition 2.2.

4.4 Generation of a random labeled tree

An easy extension of Proposition 4.6 says that $(d_{V+1}, d_{V+2}, \ \dots \ , d_{V+n}, d_V)$ satisfies constraints (2.1), and that one can generate a random labeled tree τ rooted at 0, with the help of $(U_1, ..., U_n)$, computing first $T(n)$ $(= V)$ and setting

$$A_i(\tau) = D_{V+i+1}(\omega) = \{1 \leq k \leq n \ \mid \ \lceil (n+1)U_k \rceil = V + i + 1\}.$$

This algorithm does not compare unfavorably to the algorithm based on the Prufer-Knuth's correspondence between labeled trees and n-tuples of $\{0, 1, ..., n\}^n$ (see [16, p.389-391], or [20, Chap. 2]), as it takes $O(n)$ to compute the subsets $A_i(\tau)$ and $O(n)$ to draw τ, given the subsets $A_i(\tau)$.

In the next Subsection we assume that the random labeled trees are generated using this algorithm. As a consequence, in Proposition 4.6, it is an equality between random variables that holds, and not merely an equality between probability distributions.

4.5 Proof of Proposition 2.2

We recall that

Theorem 4.7 *(Vervaat, 1979) Let $b = (b(t))_{0 \leq t \leq 1}$ be a Brownian bridge, and let T be the almost surely unique point such that $b(T) = \min_{0 \leq t \leq 1} \ b(t)$. Then T is uniform and $e = (e(t))_{0 \leq t \leq 1}$, defined by $e(t) = b(\{T + t\}) - b(T)$, is a normalized Brownian excursion, independent of T.*

Theorem 4.3 asserts the existence, on the same probability space as $(U_k)_{k \geq 1}$, of a sequence of Brownian bridges $(b_n)_{n \geq 1}$, that approximate closely the sequence $(\alpha_n)_{n \geq 1}$. According to Theorem 4.7, together with $(b_n)_{n \geq 1}$ comes a sequence of Brownian excursions $(e_n)_{n \geq 1}$, whose maxima,

$$\max_{0 \leq t \leq 1} e_n(t) = \max_{0 \leq t \leq 1} b_n(t) - \min_{0 \leq t \leq 1} b_n(t),$$

are precisely the random variables m_n of Proposition 2.3. Set

$$\begin{aligned} Q_n &= \sqrt{n} \sup_{0 \leq t \leq 1} \left| \alpha_n \left(\frac{\lfloor (n+1)t \rfloor}{n+1} \right) - \alpha_n(t) \right|, \\ R_n &= \sqrt{n} \sup_{0 \leq t \leq 1} \left| \alpha_n(t) - b_n(t) \right|. \end{aligned}$$

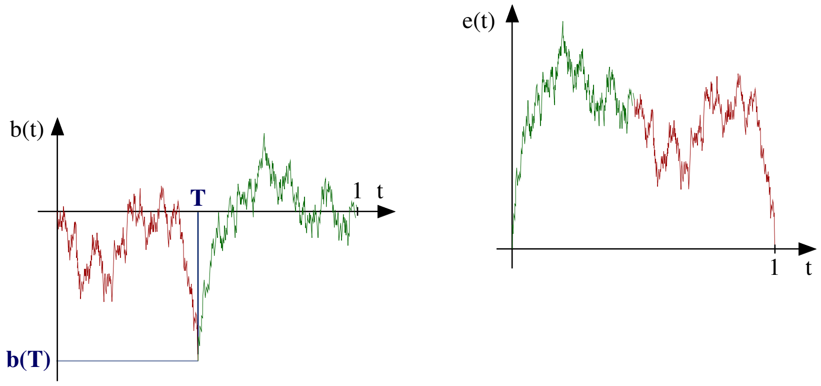

Figure 7: Vervaat's decomposition.

Due to the construction of $y_k^{(n)}$ in Subsection 4.4, we have

$$\begin{aligned}\left|\frac{\max_k y_k^{(n)}}{\sqrt{n}} - m_n\right| &\le \left|\frac{\max_k y_k^{(n)}}{\sqrt{n}} - (\sup_{0\le t\le 1}\alpha_n(t) - \inf_{0\le t\le 1}\alpha_n(t))\right| + \frac{2R_n}{\sqrt{n}}\\ &\le \frac{1+2Q_n+2R_n}{\sqrt{n}}.\end{aligned}$$

The second inequality is the point where we use Proposition 4.6. By Theorem 4.3, R_n belongs to any L_p, and

$$\|R_n\|_p = O_p(\log n).$$

Proposition 2.2 follows at once from the preceding relation and from its analog for Q_n, a consequence of the next Proposition.

Proposition 4.8 *For any positive constant K,*

$$\Pr(Q_n \ge u + \log n) \;=\; O_K\left(n^{1-K}e^{-Ku}\right).$$

Proof. Set

$$Z_n = \max_{0\le k\le n} d_k.$$

We have

$$\begin{aligned}\left|\alpha_n\left(\frac{\lfloor (n+1)t\rfloor}{n+1}\right) - \alpha_n(t)\right| &\le \frac{\sqrt{n}}{1+n} + \frac{Z_n}{\sqrt{n}}\\ &\le \frac{2\,Z_n}{\sqrt{n}},\end{aligned}$$

and

$$\begin{aligned}\Pr(2\,Z_n \ge u+\log n) &\le n\Pr(2\,d_1 \ge u+\log n)\\ &\le nE[\exp(2Kd_1)]\exp(-K\log n)e^{-Ku}\\ &\le n\left(1+\frac{e^{2K}-1}{n+1}\right)^n\exp(-K\log n)e^{-Ku},\end{aligned}$$

the first inequality due to the fact that the d_i's have the same distribution, and the third inequality because this distribution is binomial with parameters $\left(n, \frac{1}{n+1}\right)$. $\diamondsuit$

Similarly, we have

$$\begin{aligned} \left| n^{-3/2} D(n+1,n) - \int_0^1 e_n(t)dt \right| &\leq \frac{2+Q_n+R_n}{\sqrt{n}} + \left| \alpha_n \left(\frac{T(n)}{n+1} \right) - \min_t b_n(t) \right| \\ &\leq 2\ \frac{1+Q_n+R_n}{\sqrt{n}} \end{aligned}$$

leading to an error bound $\frac{\log n}{\sqrt{n}}$ for the convergence of the k^{th} moment of the total displacement to the k^{th} moment of the Airy law. Flajolet et al. have a better bound $\left(O_k\left(\frac{1}{\sqrt{n}}\right)\ ...\right)$, but the bound we obtain would hold for any smooth functional of the parking function.

Proof of Lemma 3.3. Proposition 4.6 entails

$$M_n \leq \sqrt{n}\,(1 + 2\sup_t |\alpha_n(t)|).$$

The DKW inequality [19]:

$$\Pr\left(\sup_t |\alpha_n(t)| \geq x\right) \leq 2\ \exp(-2x^2),$$

entails the desired inequality

$$\begin{aligned} E(M_n^\alpha) &= n^{\alpha/2}\alpha \int_0^{+\infty} x^{\alpha-1} \Pr\left(\frac{M_n}{\sqrt{n}} \geq x\right) dx \\ &\leq n^{\alpha/2}\alpha \left[\int_0^1 x^{\alpha-1}\,dx + 2\int_0^{+\infty} (x+1)^{\alpha-1}\ \exp(-x^2/2)\,dx\right]. \end{aligned}$$

5 Concluding remarks

Convergence of moments of the width extends easily to binary trees : the BFS random walk for a binary tree is a ruin sequence, and in the correspondence between ruin sequences and *general* trees, the maximum of the ruin sequence is within $O(1)$ of the height of the corresponding general tree. Thus we can use

Theorem 5.1 *(Flajolet-Odlyzko, 1982) The r^{th} moment of the height of a general tree with n nodes is asymptotic to $2^{-r/2}n^{r/2}E(m^r)$,*

instead of Proposition 2.2, to obtain convergence of moments of the width of binary trees. However, compared with Theorem 1.1, we lose the speed of convergence.

Asymptotics for the moments of the width of binary trees, or of general trees, can also be obtained through closed form formulas for the distribution function of the maximum of the breadth-first search random walk, using a weaker form of Proposition 2.1 [28, p. 197-201]. In a work in progress, Cyril Banderier and Philippe Flajolet study carefully asymptotics of the maximum of the the BFS random walk for general simple trees with finite degree. Together with Proposition 2.1, it gives asymptotics for moments of the width of general simple trees with finite degree. In a recent paper [10], Drmota and Gittenberger derived asymptotics of all moments (without rate) of width of *general* simple trees.

In [7], the results of Subsections 4.3 and 4.5 are generalized to study the "emergence of a giant block" of consecutive cars for a parking function. An interesting phenomenon of coalescence of blocks appears, reminiscent of the coalescence of connected components for the random graph process, during its phase transition [3].

References

[1] D. Aldous, (1991) *The continuum random tree II: an overview.* Stochastic analysis, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167, 23-70.

[2] D. Aldous, (1993) *The continuum random tree III.* Ann. of Probab. 21, No.1, 248-289.

[3] D. J. Aldous, (1997) *Brownian excursions, critical random graphs and the multiplicative coalescent.* Ann. Probab. 25, No.2, 812-854.

[4] D. J. Aldous, (1999) *Deterministic and stochastic models for coalescence (aggregation, coagulation): a review of the mean-field theory for probabilists.* Bernoulli 5, 3-48.

[5] P. Biane, M. Yor, (1987) *Valeurs principales associées aux temps locaux browniens.* Bull. Sci. Maths 111, 23-101.

[6] A. Cayley, (1889) *A theorem on trees.* Quarterly Journal of Pure and Applied Math. 23, 376-378.

[7] Ph. Chassaing, G. Louchard, (2000) *Phase transition for parking blocks, Brownian excursion and coalescence.* Available at: http://www.iecn.u-nancy.fr/∼chassain.

[8] K.L. Chung, (1976) *Excursions in Brownian motion.* Ark. för Math., 14, p.155-177.

[9] M. Drmota & B. Gittenberger, (1997) *On the profile of random trees.* Random Structures Algorithms 10, No. 4, 421–451.

[10] M. Drmota & B. Gittenberger, (2001) *The width of Galton-Watson trees.* Available at: http://www.geometrie.tuwien.ac.at/drmota/

[11] I.A. Ibragimov, Yu.V. Linnik, (1971) *Independent and stationary sequences of random variables.* Groningen, The Netherlands: Wolters-Noordhoff Publishing Company .

[12] P. Flajolet, A. Odlyzko, (1982) *The average height of binary trees and other simple trees.* J. Comp. and Sys. Sci., 25, No.2, pages ??.

[13] P. Flajolet, P. Poblete, A. Viola, (1998) *On the analysis of linear probing hashing.* Algoritmica 22, No. 4, 490-515.

[14] D. Foata, J. Riordan, (1974) *Mappings of acyclic and parking functions.* Aequationes math. 10, 10-22.

[15] D.P. Kennedy, (1976) *The distribution of the maximum Brownian excursion.* J. Appl. Probab. 13, 371-376.

[16] D. E. Knuth, (1997) *The art of computer programming. Vol. 1: fundamental algorithms.* 2nd ed., Addison-Wesley.

[17] D. E. Knuth, (1998) *Linear probing and graphs.* Algoritmica 22, No. 4, 561-568.

[18] J. Komlós, P. Major, G. Tusnády, (1975) *An approximation of partial sums of independent* RV*'s and the sample* DF*. I.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32, 111-131.

[19] P. Massart, (1990) *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality.* Ann. Probab. 18, No.3, 1269-1283.

[20] J.W. Moon, (1970) *Counting labelled trees.* Canadian Mathematical Monographs, No. 1, Canadian Mathematical Congress, Montreal.

[21] A.M. Odlyzko, H.S. Wilf, (1987) *Bandwidths and profiles of trees.* J. Comb. Theory, Ser. B 42, 348-370.

[22] V.V. Petrov, (1975) *Sums of independant random variables.* Springer .

[23] J. Pitman, (1998) *Enumerations of trees and forests related to branching processes and random walks.* In Microsurveys in Discrete Probability, ed. by D. Aldous and J. Propp. DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 41, Amer. Math. Soc., Providence, RI.

[24] A. Rényi & G. Szekeres, (1967) *On the height of trees.* J. Aust. Math. Soc. 7, 497-507.

[25] M. P. Schützenberger, (1968) *On an enumeration problem.* J. Combinatorial Theory 4, 219–221.

[26] G. R. Shorack, J. A. Wellner, (1986) *Empirical processes with applications to statistics.* Wiley.

[27] J. Spencer, (1997) *Enumerating graphs and Brownian motion.* Commun. Pure Appl. Math. 50, No. 3, 291-294.

[28] L. Takács, (1993) *Limit distributions for queues and random rooted trees.* J. Appl. Math. Stoch. Ana., 6, No.3, 189-216.

[29] W. Vervaat, (1979) *A relation between Brownian bridge and Brownian excursion.* Ann. Probab. 7, No. 1, 143-149.