The contour of size n general planar trees

Jean-François Marckert

Université de Versailles

45 Avenue des Etats Unis

78035 Versailles Cedex

marckert@math.uvsq.fr

The contour of a tree t is the polygonal function constructed from the heights of the leaves of t. We prove that the contour of size n general planar trees, suitably normalized, converges to the Brownian excursion.

Introduction

We denote by G the set of general planar trees, i.e. ordered planar trees with unrestricted node degrees, and by G_n the subset of G which contains the trees with n nodes. The cardinal of G_n is given by $C_{n-1} = \binom{2(n-1)}{n-1}/n$, the $n-1^{th}$ Catalan number.

A tree t from G_n has a random number $N(t)$ of leaves which can be numbered from the most left one to the most right one $l(1), \cdots, l(N(t))$. The height $H_t(l(i))$ of the i^{th} leaf of t is defined to be the distance from $l(i)$ to the root. The sequence $\big(H_t(l(i))\big)_{i=1,\cdots,N(t)}$ is called the contour process of t. By convention, the height of a non-existing leaf is zero.

The aim of this paper is to prove the following theorem:

Theorem 1 *The contour process converges to the Brownian excursion (for the Skohorod topology).*

$$\Big(\frac{H(l(\lfloor \frac{nx}{2} \rfloor))}{\sqrt{2n}}\Big)_{x\in[0,1]} \overset{weakly}{\longrightarrow} \big(e(x)\big)_{x\in[0,1]}.$$

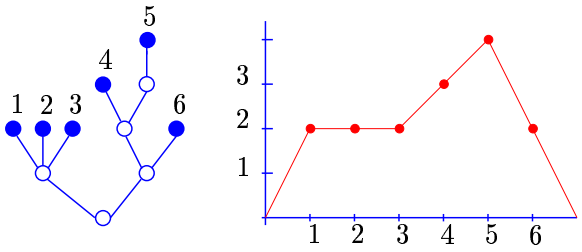

Figure 1 : A tree and its associated contour sequence.

Set $\hat{H}_n$, the function whose graph joints linearly the set of points $\left(\frac{k}{n}, \frac{H(l(k/2))}{\sqrt{2n}}\right)_{k\in[\![0,n]\!]}$. $(\hat{H}_n(x))_{x\in[0,1]}$ converges weakly to the Brownian excursion in the set of continuous function on [0,1] endowed with the uniform topology.

Bernoulli excursions and simple random walks

We denote by $RW(2n)$ the set of simple random walks with $2n$ steps. An element $\omega = (S_0(\omega), S_1(\omega), \cdots, S_{2n}(\omega))$ of RW_{2n} satisfies:

$$S_0(\omega) = 0 \text{ and for } 0 \leq k \leq 2n-1, S_{k+1}(\omega) = S_k(\omega) \pm 1,$$

the 2^{2n} path being equiprobable. The set $Es(2n)$ of Bernoulli excursions with $2n$ steps is the subset of $RW(2n)$ whose elements satisfy:

$$S_{2n}(\omega) = 0, S_k(\omega) > 0 \text{ for any } k \in \{1, \cdots, 2n-1\}, \tag{1}$$

each of the C_{n-1} elements of $Es(2n)$ being equiprobable. A Bernoulli excursion is a simple random walk conditioned to satisfy conditions (1).

The contour seen on the depth first walk

The depth first search (see Aldous [1] p.260), gives a rich one to one correspondence between $Es(2n)$ and G_n:

$$\begin{array}{rcl} \omega : G_n & \longrightarrow & Es(2n) \\ t & \longmapsto & \omega(t). \end{array} \tag{2}$$

To construct $\omega(t)$, the *depth first walk*, consider a fly, walking with unit speed around t, clockwise,

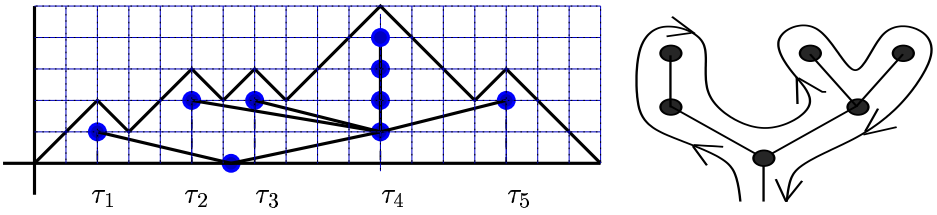

Figure 2 *A tree and its associated Bernoulli excursion.*

from the root to the root. The height (+1) of the fly in function of the time gives the depth first walk. It follows for example, that the height $h(t)$ of the tree is the maximum of the associated Bernoulli excursion minus 1:

$$h(t) = \max\{\omega(t) - 1\}.$$

Hence, the law of the height of the trees in G_n, suitably normalized, converges to the law of the maximum of the Brownian excursion (when n goes to $+\infty$).

Difference with the contour of binary tree

the contour of general trees and the contour of binary trees studied by Gutjahr & Pflug [3]. The first one is that, in the context of general trees, the contour does not characterize the tree.The second one, which is in the core of the present paper is relation (2), that is DFW is a simple conditioned random walk; this is not at all the case for the depth first walk associated to binary trees.

For a simple walk ω of $RW(2n)$, we define the sequence $(\tau_k(\omega))_{k\geq 0}$ of indices on which ω reaches a local maximum; more formally:

$$\begin{aligned}\tau_1 &= \inf\{j | j \geq 1, S_j > S_{j+1}\} \\ \tau_k &= \inf\{j | j > \tau_{k-1}, S_{j-1} < S_j, S_j > S_{j+1}\}.\end{aligned}$$

We define the r.v. K by:

$$K = \max\{k, \tau_k \leq 2n\}.$$

It is well known, and easy to check that, if we consider the depth first Bernoulli excursion ω associated to a tree t, the sequence $(S_{\tau_k})_{k\geq 1}$ satisfies:

$$\begin{aligned}S_{\tau_k}(\omega(t)) &= H_t(l(k)) + 1, \\ K(\omega(t)) &= N(t).\end{aligned}$$

Study of r.v. on $ES(2n)$ via corresponding r.v. on $RW(2n)$

The key point of this paper is the following remark:
if there exists a moderate deviations principle (or large deviations principle) for a functional of the simple random walk, there exists a bound for the analogous principle on the excursion. Indeed, assume that A_n is a subset of $RW(2n)$ such that,

$$\mathbb{P}(\omega \in A_n) = o(n^{-\beta}) \quad \text{then} \quad \mathbb{P}(\omega \in A_n | \omega \in Es(n)) = o(n^{-\beta+3/2}). \tag{3}$$

This is just a consequence of

$$\mathbb{P}(Es(2n) | RW(2n)) = 2^{-2n} C_{n-1} \sim \frac{n^{-3/2}}{4\sqrt{\pi}}.$$

This remarks which may appear to be very weak, is fundamental in this paper. Indeed, almost all r.v. variables are much more difficult to handle on $ES(2n)$ than on $RW(2n)$. But, concentration like inequalities on $RW(2n)$ remains "true" on $ES(2n)$.

Moderate deviations of the sequence $(\tau_k)_{k\geq 1}$

Set $\Omega_1(2n)$ the subset of $Es(2n)$ defined by:

$$\omega \in \Omega_1(2n) \Leftrightarrow \Big(\omega \in Es(2n) \text{ s.t. } \forall k, 1 \leq k \leq K(\omega), \big|\tau_k(\omega) - 4k\big| < n^{3/5}\Big).$$

Lemma 2 *If n is large enough, we have:*

$$\mathbb{P}\big(\Omega_1(2n)\big) \geq 1 - \exp(-n/6)$$

Proof : We use here a principle of "moderate deviations" on the set of values $\tau_1, \tau_2, \cdots$. The law of τ_k is simple on $RW(2n)$: $\tau_{k+1} - \tau_k$ is the sum of two independent geometric distributed r.v. with parameter $1/2$ (we denote this law by $\mathcal{G}(1/2)$). So, one has on $RW(2n)$:

$$\mathbb{E}(\tau_{k+1} - \tau_k) = 4, \mathrm{Var}(\tau_{k+1} - \tau_k) = 4.$$

τ_1 is the sum of two $\mathcal{G}(1/2)$ distributed r.v. minus 1. If $(g_i)_{i\geq 1}$ denotes a sequence of $\mathcal{G}(1/2)$ independent r.v., one has:

$$\text{on } RW(2n), \quad \tau_k \stackrel{law}{=} \Big(\sum_{j=1}^{2k} g_j\Big) - 1.$$

The law convergence of $(\tau_k - 4k)/\sqrt{4k}$ to the normal law provides the following deviations:

Lemma 3 *On $RW(2n)$, there exists a constant c_1 such that for n large enough:*

$$\mathbb{P}\big(\exists k \ s.t. \ 1 \leq k \leq K(\omega), \big|\tau_k - 4k\big| \geq n^{3/5}\big) \leq e^{-c_1 n^{1/5}}.$$

Proof : K is random, but surely $K \leq n$. Hence,

$$\mathbb{P}\big(\exists k \text{ s.t. } 1 \leq k \leq K(\omega), \big|\tau_k - 4k\big| \geq n^{3/5}\big) \leq \sum_{k=1}^{n} \mathbb{P}\big(\big|\tau_k - 4k\big| \geq n^{3/5}\big). \tag{4}$$

According to Petrov [5] (p.52 and Lemma 5 p.55), since geometric random variable has exponential moments, there exists a constant $\gamma > 0$ such that, for any $1 \leq k \leq n$,

$$\mathbb{P}(|\tau_k - 4k| \geq n^{3/5}) \leq 2e^{-\frac{n^{6/5}}{\gamma n}}. \tag{5}$$

Equations (4) and (5) yields Lemma 3. □

Lemma 3 and equation (3) can be apply to end the proof of Lemma 2:

$$\begin{aligned}\mathbb{P}\big(\complement\Omega_1(2n)\big) &= \mathbb{P}(\exists k \text{ s.t. } \big|\tau_k - 4k\big| \geq n^{3/5} | Es(2n)\big) \\ &\leq e^{-n^{1/6}}\end{aligned}$$

if n large enough. □

Moderate variation of excursions

Set $\Omega_2(2n)$ the subset of $Es(2n)$ defined by:

$$\omega \in \Omega_2(2n) \Leftrightarrow \Big(\omega \in Es(2n) \text{ s.t. } \forall m,k \in \{0,\cdots,2n\}, m+k \ \leq 2n, |S_{m+k}(\omega) - S_m(\omega)| \leq 4\sqrt{k\log n}\Big).$$

Lemma 4 *If n is large enough, one has:*

$$\mathbb{P}(\Omega_2(2n)) \geq 1 - n^{-4}.$$

First, we prove an analogous Lemma on $RW(2n)$. Following Odlyzko [4], we consider the subset $\Omega_3(2n)$ of $RW(2n)$ defined by:

$$\omega \in \Omega_3(2n) \Leftrightarrow \Big(\omega \in RW(2n) \text{ s.t. } \forall m,k \in \{0,\cdots,2n\}, m+k \ \leq 2n, |S_{m+k}(\omega) - S_m(\omega)| \leq 4\sqrt{k\log n}\Big).$$

Odlyzko[4] shows:

Lemma 5

$$\mathbb{P}(\Omega_3(n)) = O(n^{-6}).$$

Proof : We have

$$\begin{aligned}\mathbb{P}(\exists k, \exists m, |S_{m+k} - S_m| \geq 4\sqrt{k\log n}) &\leq \sum_{m=0}^{2n-1}\sum_{k=1}^{2n-m} \mathbb{P}(|S_{m+k} - S_m| \geq 4\sqrt{k\log n}) \\ &= O(n^2)\exp(-\frac{4^2\log n}{2}) \\ &= O(n^{-6}), \qquad (6)\end{aligned}$$

where the second inequality follows from Chernov bounds (Bollobás [2] p.12.): for any $\alpha \geq 0$ and $k \in \mathbb{N}$,

$$\mathbb{P}(|S_k| > \alpha) \leq 2\exp(-\frac{\alpha^2}{2k}). \ \square \tag{7}$$

The end of the proof of Lemma 5 follows once again (3):

$$\begin{aligned} \mathbb{P}(\complement\Omega_2(2n)) &= \mathbb{P}\big(\complement\Omega_3(2n)\big|Es(2n)\big) \\ &= O(n^{-6}n^{3/2}) = O(n^{-9/2}). \ \square \end{aligned}$$

End of the proof of Theorem 1

For each k, set $\tilde{S}_k = S_{\tau_k}$. We have

$$\mathbb{P}(\exists k \text{ s.t. } \left|\tilde{S}_k - S_{4k}\right| \geq n^{2/5}) = O(n^{-4}).$$

$$\begin{aligned} \mathbb{P}(\exists k \text{ s.t. } \left|\tilde{S}_k - S_{4k}\right| \geq n^{2/5}) \leq\ & \mathbb{P}(\complement\Omega_1(2n)) + \mathbb{P}(\complement\Omega_2(2n)) + \\ & \mathbb{P}(\exists k \text{ s.t. } \left|\tilde{S}_k - S_{4k}\right| \geq n^{2/5}, \Omega_1(2n), \Omega_2(2n)) \end{aligned}$$

The last term in the right hand side is 0 if n is sufficiently large, since in $\Omega_1(2n) \cap \Omega_2(2n)$, we have:

$$|S_{\tau_k} - S_{4k}| \leq 4\sqrt{|\tau_k - 4k| \log n} \leq 4\sqrt{n^{3/5} \log n}.$$

Hence, with probability $1 - O(n^{-4})$, we have:

$$\sup_{t\in[0,1]} \left\{ \left| \frac{\tilde{S}_{[nt/2]} - S_{[2nt]}}{\sqrt{n}} \right| \right\} = O(n^{-1/10}).$$

Since

$$\Big(\frac{S_{2nt}}{\sqrt{2n}}\Big)_{t\in[0,1]} \overset{weakly}{\longrightarrow} \big(e(t)\big)_{t\in[0,1]},$$

we have:

$$\Big(\frac{\tilde{S}_{nt/2}}{\sqrt{2n}}\Big)_{t\in[0,1]} \overset{weakly}{\longrightarrow} \big(e(t)\big)_{t\in[0,1]}$$

Conclusion

One refers to Aldous [1] for an overview on the depth first walk associated to Galton Watson trees. In the setting of Galton Watson trees with progeny $(p_i)_{i\geq 0}$ conditioned by their total population (say n here), he shows that the DFS excursion $X_1, \cdots X_{2n}$ (Aldous calls this walk the search depth) associated to a tree, satisfies:

$$\Big(\frac{X_{[2nt]}}{\sqrt{n}}, 0 \leq t \leq 1\Big) \stackrel{weakly}{\longrightarrow} \Big(2\sigma^{-1}e(t), 0 \leq t \leq 1\Big), \tag{8}$$

where σ^2 is the variance of the critical Galton Watson process from the same exponentially family as $(p_i)_{i\geq 0}$ (see [1] p.26). The set of general planar trees with n nodes (endowed with the uniform law) can be view as a Galton Watson process with shifted geometric progeny $(p_i = 2^{-i+1})_{i\geq 0}$ (for which $\sigma^2 = 2$); the particularity of general planar tree, is that the depth first walk is exactly a simple random walk conditioned to be an excursion. In any other cases, the discrete process $(X_k)_{k=0,\cdots,2n}$ is not a conditioned random walk. The number of leaves $N(t)$ of a Galton Watson process with progeny $(p_i)_{i\geq 0}$ conditioned by the total population n, satisfies:

$$\frac{N(t)}{n} \stackrel{proba}{\longrightarrow} p_0.$$

If the "np_0" corresponding values τ_k are "well distributed" in $[0, 2n]$ one can prove a similar theorem as Theorem 1. So, one have the following

Conjecture : *The contour process $H_{l(t)}$ of a critical Galton Watson process with progeny distribution $(p_i)_{i=0,\cdots}$ and variance σ^2 satisfies:*

$$\Big(\frac{H_{l([p_0\, n\, t])}}{\sqrt{n}}\Big)_{t\in[0,1]} \stackrel{weakly}{\longrightarrow} \big(\frac{2}{\sigma}e(t)\big)_{t\in[0,1]}.$$

References

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

[2] Bollobás, B. *Random Graphs* , Academic Press, p.12 (1985).

[3] Gutjahr, W.; Pflug, G.C. (1992), *The asymptotic contour process of a binary tree is a Brownian excursion.* Stoch. Proc. Appl. 41, No.1, 69-89.

[4] Odlyzko, A.M. (1995), *Search for the maximum of a random walk*, R.S.A. Vol.6 p.275-295 .

[5] Petrov, V.V. (1975), *Sums of Independant Random Variables*, Springer .