*Proceedings of the 27th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms
Kraków, Poland, 4-8 July 2016*

Cost functionals for large random trees

Jean-François Delmas[1], Jean-Stéphane Dhersin[2], Marion Sciauveau[1†]

[1] *CERMICS, École des Ponts, UPE, Champs-sur-Marne, France*
[2] *Université Paris 13, LAGA, 93430 Villetaneuse, France*

Additive tree functionals allow to represent the cost of many divide-and-conquer algorithms. We give an invariance principle for such tree functionals for the Catalan model (random tree uniformly distributed among the full binary ordered tree with given number of internal nodes). This relies on the natural embedding of binary trees into the Brownian excursion and then on elementary L^2 computations. We recover results first given by Fill and Kapur (2004) and then by Fill and Janson (2009).

Keywords: random binary tree, cost functional, toll function, Brownian excursion, continuum random tree

1 Introduction

1.1 Additive functionals and toll functions

Additive functionals on binary trees allow to represent the cost of algorithms such as "divide and conquer", see Kapur's PhD thesis [24] and Fill and Kapur [16]. For T a rooted full binary ordered tree, we set $|T|$ its cardinal, $\emptyset$ its root, $L(T)$ and $R(T)$ the left-sub-tree and right-sub-tree of the root of T. A functional F on binary trees is called an additive functional if it satisfies the following recurrence relation:

$$F(T) = F(L(T)) + F(R(T)) + b_{|T|}, \tag{1}$$

for all trees T such that $|T| \geq 1$ and with $F(\emptyset) = 0$. The given sequence $(b_n, n \geq 1)$ is called the toll function. Notice that:

$$F(T) = \sum_{v \in T} b_{|T_v|}, \tag{2}$$

where T_v is the sub-tree above v whose root is v.

We give some examples of commonly used toll functions or index functions related to additive functional. For $v, w \in T$, we say that w is an ancestor of v and write $w \leq v$ if $v \in T_w$. For $u, v \in T$, we denote by $u \wedge v$, the most recent common ancestor of u and v: $u \wedge v$ is the only element of T such that: $w \leq u$ and $w \leq v$ implies $w \leq u \wedge v$. We shall denote by d the graph distance in T.

- The total size of the tree T, $|T|$, corresponds to the additive functional with toll function $b_n = 1$.

[†]This work is partially supported by DIM RDMath IdF

- The total path length of T is defined by $P(T) = \sum_{v\in T} d(\emptyset, v)$. We have that $P(T) + |T|$ is the additive functional with toll function $b_n = n$ as

$$\sum_{w\in T} |T_w| = \sum_{w\in T}\sum_{v\in T} \mathbf{1}_{\{w\leq v\}} = \sum_{v\in T} \big(1 + d(\emptyset, v)\big) = |T| + P(T).$$

- The shape functional of a tree T is the additive functional with toll function $b_n = \log(n)$. (This functional will not be covered by the main results of this paper.)

- The Wiener index of the tree T is defined by $W(T) = \sum_{u,v\in T} d(u, v)$. Notice that $d(u, v) = d(\emptyset, u) + d(\emptyset, v) - 2d(\emptyset, u \wedge v)$. This implies that $W(T) = 2|T| \sum_{w\in T} |T_w| - 2\sum_{w\in T} |T_w|^2$ as

$$\sum_{w\in T} |T_w|^2 = \sum_{w\in T}\sum_{u,v\in T} \mathbf{1}_{\{w\leq u\wedge v\}} = \sum_{u,v\in T} \big(1 + d(\emptyset, u \wedge v)\big) = |T|^2 + \sum_{u,v\in T} d(\emptyset, u \wedge v).$$

  According to (2), the functional $\sum_{w\in T} |T_w|^2$ is an additive functional with toll function $b_n = n^2$. And thus the Wiener index of a full binary tree is a combination of two additive functionals.

- The Sackin index (or external path length) of the tree T, used to study the balance of the tree, is similar to the total path length of T when one considers only the leaves: $S(T) = \sum_{v\in\mathcal{L}(T)} d(\emptyset, v)$, where the set of leaves is $\mathcal{L}(T) = \{v \in T; |T_v| = 1\}$. Using that for a full binary tree we have $|T| = 2|\mathcal{L}(T)| - 1$, we deduce that $2S(T) = \sum_{w\in T} |T_w| - 1$. The Colless index of the tree T is defined as $C(T) = \sum_{v\in T} \big||\mathcal{L}(L_v)| - |\mathcal{L}(R_v)|\big|$. Since T is a full binary tree, we get $2|\mathcal{L}(L_v)| - 2|\mathcal{L}(R_v)| = |L_v| - |R_v|$ and $|L_v| + |R_v| = |T_v| - 1$. We obtain that $2C(T) = \sum_{w\in T} |T_w| - |T| - 2\sum_{v\in T} \min(|L_v|, |R_v|)$. The cophenetic index of the tree T, which is used in [27] to study the balance of the tree, is defined by $\mathrm{Co}(T) = \sum_{u,v\in\mathcal{L}(T),\, u\neq v} d(\emptyset, u\wedge v)$. Using again that T is a full binary tree, we get $4\mathrm{Co} = 4\sum_w |\mathcal{L}(T_w)|(|\mathcal{L}(T_w)| - 1) - 4|\mathcal{L}(T)|(|\mathcal{L}(T)| - 1) = \sum_{w\in T} |T_w|^2 - |T|^2 - |T| + 1$.

1.2 Asymptotics for additive functionals in the Catalan model

We consider the Catalan model: let T_n be a random tree uniformly distributed among the set of full binary ordered trees with n internal nodes (and thus $n+1$ leaves), which has cardinal $C_n = (2n)!/[(n!^2)(n+1)]$. In particular, we have:

$$|T_n| = 2n + 1.$$

Recall that T_n is a (full binary) Galton-Watson tree (also known as simply generated tree) conditioned on having n internal nodes. It is well known, see Takàcs [34], Aldous [3, 4] and Janson [21], that $|T_n|^{-3/2}P(T_n)$ converges in distribution, as n goes to infinity, towards $2\int_0^1 B_s\, ds$, where $B = (B_s, 0 \leq s \leq 1)$ is the normalized Brownian excursion. This result can be seen as a consequence of the convergence in distribution of T_n (in fact the contour process) properly scaled towards the Brownian continuum tree whose contour process is B, see [3] and Duquesne [9], or Duquesne and Le Gall [10] in the setting of Brownian excursion. For a combinatorial approach, which can be extended to other families of trees, see also Fill and Kapur [15, 17] or Fill, Flajolet and Kapur [13].

In [16], the authors considered the toll functions $b_n = n^\beta$ with $\beta > 0$ and they proved that with a suitable scaling the corresponding additive functional $F_\beta(T_n)$ converge in distribution to a limit, say Y_β. The distribution of Y_β is characterized by its moments. (In [16], the authors consider also the toll function $b_n = \log(n)$.) See also Janson and Chassaing [23] for asymptotics of the Wiener index, which is a consequence of the joint convergence in distribution of $(F_1(T_n), F_2(T_n))$ with a suitable scaling and Blum, François and Janson [6] for the convergence of the Sackin and Colless indexes. It is announced in Fill and Janson [14] that for $\beta > 1/2$, Y_β can be represented as a functional of the normalized Brownian excursion. More precisely, for $\beta > 1/2$, Y_β is distributed as $\phi_\beta(B)$, where for any non-negative continuous function h defined on $[0,1]$:

$$\phi_\beta(h) = \beta \int_0^1 [t^{\beta-1} + (1-t)^{\beta-1}] h(t)\, dt - \frac{1}{2}\beta(\beta-1) \int_{[0,1]^2} |t-s|^{\beta-2} [h(t) + h(s) - 2m_h(s,t)]\, dsdt,$$

with

$$m_h(s,t) = \inf_{u \in [s\wedge t, s\vee t]} h(u). \tag{3}$$

Furthermore, for $\beta = 1$, this reduces to $\phi_1(h) = \int_0^1 h$ and for $\beta > 1$ we also have:

$$\phi_\beta(h) = \beta(\beta-1) \int_{[0,1]^2} |t-s|^{\beta-2} m_h(s,t)\, dsdt. \tag{4}$$

We use the natural embedding of T_n into the Brownian excursion, see [4], so that the convergence in distribution of the additive functional is then an a.s. convergence (which holds simultaneously for all $\beta > 1/2$) and also give the fluctuations for this a.s. convergence. From this convergence, we also provide another representation of $\phi_\beta(h)$ which is a natural by-product of the a.s. convergence.

Remark 1.1 *The method presented in this paper based on the embedding of T_n into a Brownian excursion can not be extended directly to other models of trees such as binary search trees, recursive trees or simply generated trees.*

Concerning binary search trees (or random permutation model or Yule trees), see [31] and [32] for the convergence of the external path length (which corresponds in our setting to the Sackin index), [28] for toll function $b_n = n^\beta$, [29] for the Wiener index (and [21] for simply generated trees), [6] (and [18] for other trees) for the Sacking and Colless indexes.

Concerning recursive trees, see [26], [8] for the convergence of the total path length and [29] for the Wiener index. In the setting of recursive tree, then (1) is a stochastic fixed point equation, which can be analyzed using the approach of [33].

Remark 1.2 *One can replace the toll function $b_{|T|}$ in (1) by a function of the tree, say $\boldsymbol{b}(T)$. For example, if one consider $\boldsymbol{b}(T) = \mathbf{1}_{\{|T|=1\}}$, then the corresponding additive functional $F(T) = |\mathcal{L}(T)|$ gives the number of leaves. The case of "local" toll function $\boldsymbol{b}$ (with finite support or fast decreasing rate) has been considered in the study of fringe trees, see [2], [7] for binary search trees, and [22] for simply generated trees and [19] for binary search trees and recursive trees.*

See [20] for the study of the phase transition on asymptotics of additive functionals with toll functions $b_n = n^\beta$ on binary search trees between the "local" regime (corresponding to $\beta \le 1/2$) and the "global" regime ($\beta > 1/2$). The same phase transition is observed for the Catalan model, see [16]. Our main result, see Theorem 3.1, concerns specifically the "global" regime.

2 Binary trees in the Brownian excursion

We begin by recalling the definition of a real tree, see [12], and some elementary properties of the Brownian continuum random tree, see [25]. A real tree is a metric space $(\mathcal{T}, d)$ which satisfies the following two properties for every $x, y \in \mathcal{T}$:

(i) There exists a unique isometric map $f_{x,y}$ from $[0, d(x, y)]$ into $\mathcal{T}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$.

(ii) If ϕ is a continuous injective map from $[0, 1]$ into $\mathcal{T}$ such that $\phi(0) = x$ and $\phi(1) = y$, then we have $\phi([0, 1]) = f_{x,y}([0, d(x, y)])$.

Equivalently, a metric space $(\mathcal{T}, d)$ is a real tree if and only if $\mathcal{T}$ is connected and d satisfies the four point condition:

$$d(s, t) + d(x, y) \le \max(d(s, x) + d(t, y), d(s, y) + d(t, x)) \quad \text{for all} \quad s, t, x, y \in \mathcal{T}.$$

A rooted real tree is a real tree $(\mathcal{T}, d)$ with a distinguished vertex $\emptyset$ called the root. In the following paragraphs, we will only consider compact rooted real trees.

One can use continuous functions to encode compact rooted real trees as follows. Let h be a non-negative continuous function defined on $[0, 1]$ such that $h(0) = h(1) = 0$. For every $x, y \in [0, 1]$, we set $d_h(x, y) = h(x) + h(y) - 2m_h(x, y)$, where m_h is defined in (3). It is easy to check that d_h is symmetric and satisfies the triangle inequality. The relation $\sim_h$ defined on $[0, 1]^2$ by $x \sim_h y \Leftrightarrow d_h(x, y) = 0$ is an equivalence relation. Let $\mathcal{T}_h = [0, 1]/\sim_h$ be the corresponding quotient space. The function d_h on $[0, 1]^2$ induces a function on $\mathcal{T}_h^2$, which we still denoted by d_h, and which is a distance on $\mathcal{T}_h$. It is not difficult to check that $(\mathcal{T}_h, d_h)$ is then a compact real tree. We denote by $\mathbf{p}$ the canonical projection from $[0, 1]$ into $\mathcal{T}_h$. Thus, the metric space $(\mathcal{T}_h, d_h)$ is a compact real tree which can be viewed as a rooted real tree by setting $\emptyset = \mathbf{p}(0)$.

Let $B = (B_t, 0 \le t \le 1)$ be a normalized Brownian excursion. Informally, B is just a linear standard Brownian path started from the origin and conditioned to stay positive on $(0, 1)$ and to come back to 0 at time 1. For $\alpha > 0$, let $e = \sqrt{2/\alpha}\, B$ and let $\mathcal{T}_e$ denote the Brownian tree. The continuum random tree introduced by Aldous corresponds to $\alpha = 1/2$ and the Brownian tree associated to the normalized Brownian excursion corresponds to $\alpha = 2$. We shall keep the parameter α so that the two previous cases are easy to read on the results.

Let $(U_n, n \in \mathbb{N}^*)$ be a sequence of independent random variables uniform on $[0, 1]$, independent of e. We denote by $\mathcal{T}_n$ the random tree spanned by the $n + 1$ points $\mathbf{p}(U_1), \ldots, \mathbf{p}(U_{n+1})$ that is the smallest connected subset of $\mathcal{T}_e$ that contains $\mathbf{p}(U_1), \ldots, \mathbf{p}(U_{n+1})$ and the root. The tree $\mathcal{T}_n$ has exactly $2n + 1$ nodes. There is a natural order on $\mathcal{T}_n$ given by the order of its external nodes $\mathbf{p}(U_1), \ldots, \mathbf{p}(U_{n+1})$. Let T_n be the corresponding trees when one forget about the branch lengths. It is well known that T_n is uniform among the full binary ordered trees with n internal nodes. See Figure (1) for an example with $n = 4$.

Let $(h_1, \ldots, h_{2n+1})$ be the branch lengths of the tree $\mathcal{T}_n$ given in the lexicographical order. We recall, see [4], [30] (Theorem 7.9) or [11], that the density of $(h_1, \ldots, h_{2n+1})$ is given by:

$$f_n(h_1, \ldots, h_{2n+1}) = 2 \frac{(2n)!}{n!} \frac{\alpha^{n+1}}{L_n} \, \mathrm{e}^{-\alpha L_n^2} \, \mathbf{1}_{\{h_1 > 0, \ldots, h_{2n+1} > 0\}},$$

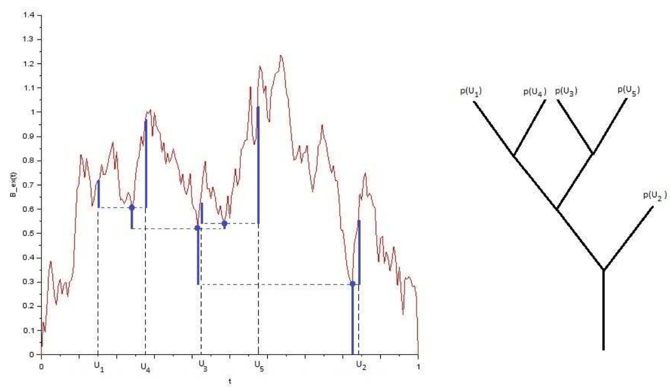

Fig. 1: The Brownian excursion, $\mathcal{T}_n$ (for $n = 4$) and T_n.

where $L_n = \sum_{k=1}^{2n+1} h_k$ denotes the total length of $\mathcal{T}_n$. Notice that the edge-lengths are independent of the shape of the tree T_n. It is then easy to deduce that the density of L_n is given by:

$$f_{L_n}(x) = 2\frac{\alpha^{n+1}}{n!} x^{2n+1}\, \mathrm{e}^{-\alpha x^2}\, \mathbf{1}_{\{x>0\}}.$$

And we have, see [1] that $L_n/\sqrt{n}$ converges a.s. towards $1/\sqrt{\alpha}$. Furthermore, elementary computations give that $(h_1, \dots, h_{2n+1})$ has the same distribution as $(L_n\Delta_1, \dots, L_n\Delta_{2n+1})$, where $\Delta_1, \dots, \Delta_{2n+1}$ represents the lengths of the $2n + 1$ intervals obtained by cutting $[0, 1]$ at $2n$ points $V_1, \dots, V_{2n}$, where $V_1, \dots, V_{2n}$ are $2n$ independent uniform random variables on $[0, 1]$ and independent of L_n. We thus deduce the following elementary Lemma.

Lemma 2.1 *The random vector $(h_1, \dots, h_{2n+1})$ has the same distribution as:*

$$\left(L_n \frac{E_1}{S_{2n+1}}, \dots, L_n \frac{E_{2n+1}}{S_{2n+1}} \right),$$

where $E_1, \dots, E_{2n+1}$ are $2n + 1$ independent exponential random variables, independent of L_n, and $S_{2n+1} = \sum_{k=1}^{2n+1} E_k$.

We end this section with a result on the Brownian excursion. We set m for m_e defined in (3). For $s \in [0, 1]$ and $r \in [0, e_s)$, the length of the excursion of e above r straddling s is given by:

$$\sigma_{r,s} = \int_0^1 \mathbf{1}_{\{m(s,t)\geq r\}} dt.$$

For $\beta \geq 0$, we set:

$$Z_\beta = \int_0^1 ds \int_0^{e_s} dr\ \sigma_{r,s}^{\beta-1}. \tag{5}$$

The next result is proved using the representation of Brownian excursion from Biane [5].

Lemma 2.2 *Let $\beta > 0$. We have:*

$$\mathbb{P}\left(Z_\beta < +\infty\right) = \begin{cases} 0 & \text{if } \beta \leq 1/2, \\ 1 & \text{if } \beta > 1/2. \end{cases}$$

The following result based on elementary computations allows to recover the formulation of our Corollary 3.2 given in [16] and [14], see (4).

Lemma 2.3 *We have $Z_1 = \int_0^1 e_s\, ds$ and for $\beta > 1$:*

$$Z_\beta = \frac{1}{2}\,\beta(\beta-1)\int_{[0,1]^2} |t-s|^{\beta-2}\, m(s,t)\, ds\, dt.$$

3 Results

Inspired by (2), we consider the following random measure A_n associated to the tree T_n defined as follows. For any non-negative function defined on $[0,1]$, we set:

$$A_n(f) = |T_n|^{-\frac{3}{2}} \sum_{v\in T_n} |T_v| f\left(\frac{|T_v|}{|T_n|}\right),$$

where we recall that T_v is the sub-tree above v with root v and $|T_n| = 2n+1$. The case $f(x) = x^{\beta-1}$ corresponds to the additive functional on T_n given by (2) with toll function $b_n = n^\beta$ up to the scaling factor $|T_n|^{-(\frac{1}{2}+\beta)}$.

We define the following random measure associated to the excursion e:

$$\Phi(f) = \sqrt{2\alpha}\int_0^1 ds \int_0^{e_s} dr\; f(\sigma_{r,s}).$$

We now state our main result on the invariance principle.

Theorem 3.1 *Almost surely, for all $f \in \mathcal{C}^0((0,1])$ such that $\lim_{x\downarrow 0} x^{\frac{1}{2}-\varepsilon} f(x) = 0$ for some $0 < \varepsilon < \frac{1}{2}$, we have:*

$$\lim_{n\to+\infty} A_n(f) = \Phi(f).$$

Proof: We only present the main ideas of the proof, as the detailed proofs will be given in a forthcoming paper. Let f be a smooth enough function defined on $[0,1]$. We first notice that $A_n(f)$ is well approximated by:

$$\begin{aligned} A_{n,1}(f) &= |T_n|^{-\frac{3}{2}} \sum_{v\in T_n} (|T_v|+1) f\left(\frac{|T_v|}{|T_n|}\right) \\ &= 2|T_n|^{-\frac{3}{2}} \sum_{u\in\mathcal{L}(T_n)} \sum_{v\in T_n,\, v\leq u} f\left(\frac{|T_v|}{|T_n|}\right), \end{aligned}$$

where we recall that $\mathcal{L}(T)$ denotes the leaves of the tree T and $|T| = 2|\mathcal{L}(T)| - 1$ in a full binary tree. The precise distribution of the heights $(h_1,\ldots,h_{2n+1})$ given in Lemma 2.1 and the fact that $L_n/\sqrt{n}$

converges a.s. towards $1/\sqrt{\alpha}$ gives that h_v is close to $L_n/(2n+1)$ that is of $1/(2\sqrt{\alpha n})$. In the spirit of the law of the large number, using L^2 computations, we obtain that $A_{n,1}(f)$ is well approximated by:

$$\begin{aligned} A_{n,2}(f) &= 4\sqrt{\alpha n}\,|T_n|^{-\frac{3}{2}} \sum_{u\in\mathcal{L}(T_n)} \sum_{v\in T_n,\, v\leq u} f\left(\frac{|T_v|}{|T_n|}\right) h_v \\ &= 4\sqrt{\alpha n}\,|T_n|^{-\frac{3}{2}} \sum_{k=1}^{n+1} \int_0^{e_{U_k}} dr\, f\left(\frac{2X_{r,k}+1}{|T_n|}\right), \end{aligned}$$

where $X_{r,k}+1$ denotes the number of integers $i\in\{1,\ldots,n+1\}$ such that the random variable U_i belongs to the same excursion interval of e above level r as U_k, that is $m(e_{U_i}, e_{U_k}) > r$. Conditionally on e, the random variable $X_{r,k}$ is binomial with parameter (n, σ_{r,U_k}). In particular, for large n, $2X_{r,k}+1$ is close to $2n\sigma_{r,U_k}$ and thus $(2X_{r,k}+1)/|T_n|$ is close to σ_{r,U_k}. Using $|T_n| = 2n+1$, the smoothness of f and L^2 computations, we get that $A_{n,2}(f)$ is well approximated by:

$$A_{n,3}(f) = \sqrt{2\alpha}\,\frac{1}{n+1}\sum_{k=1}^{n+1}\int_0^{e_{U_k}} dr\, f\left(\sigma_{r,U_k}\right).$$

Then use the law of large number (conditionally on e) to get that a.s.

$$\lim_{n\to+\infty} A_{n,3}(f) = \sqrt{2\alpha}\int_0^1 ds\int_0^{e_s} dr\, f\left(\sigma_{r,s}\right).$$

We deduce that for all (smooth enough) functions, we have a.s. $\lim_{n\to+\infty} A_n(f) = \Phi(f)$. Since the considered family of smooth functions is convergence determining, this implies that a.s. $(A_n, n\in\mathbb{N}^*)$ converges towards Φ (for the weak convergence or vague convergence of finite measure on $[0,1]$). This in turns gives that a.s. for all continuous functions, $\lim_{n\to+\infty} A_n(f) = \Phi(f)$. More work is required to extend this result to the class of functions considered in the Theorem. □

According to Lemma 2.2, the random variable Z_β defined by (5) is a.s. finite (resp. infinite) if $\beta > 1/2$ (resp. $0 < \beta \leq 1/2$). Considering the function $f(x) = x^{\beta-1}$ for $\beta > 0$, we easily deduce from Theorem 3.1 the following convergence. For $n\in\mathbb{N}^*$, we set:

$$Z_\beta^{(n)} = \frac{1}{\sqrt{2\alpha}}\,|T_n|^{-(\beta+\frac{1}{2})}\sum_{v\in T_n}|T_v|^\beta.$$

Corollary 3.2 *We have almost surely, for all $\beta > 0$,*

$$\lim_{n\to+\infty} Z_\beta^{(n)} = Z_\beta.$$

Remark 3.3 *Corollary 3.2 gives directly that $(|T_n|^{-3/2}\sum_{v\in T_n}|T_v|, |T_n|^{-5/2}\sum_{v\in T_n}|T_v|^2)$ is asymptotically distributed as $\sqrt{2\alpha}\,(Z_1, Z_2)$. Since, according to [6] or [18], the quantity $\sum_{v\in T_n}\min(|L_v|,|R_v|)$ is of smaller magnitude than $|T_n|^{3/2}$, we can directly recover the joint asymptotic distribution of the total length path, the Wiener, Sackin, Colless and cophenetic indexes defined in Section 1.1 for the Catalan model. More precisely, we have the following a.s. convergence as n goes to infinity:*

$$\left(\frac{P(T_n)}{|T_n|^{3/2}}, \frac{W(T_n)}{|T_n|^{5/2}}, \frac{S(T_n)}{|T_n|^{3/2}}, \frac{C(T_n)}{|T_n|^{3/2}}, \frac{\mathrm{Co}(T_n)}{|T_n|^{5/2}}\right) \to \sqrt{2\alpha}\left(Z_1, 2(Z_1-Z_2), \frac{Z_1}{2}, \frac{Z_1}{2}, \frac{Z_2}{4}\right).$$

The next proposition gives the fluctuations corresponding to the invariance principles of Corollary 3.2 when $\beta \geq 1$. Notice the speed of convergence in the invariance principle is of order $|T_n|^{-1/4}$ and the limiting variance is (up to a multiplicative constant) given by $Z_{\beta'}$ with $\beta' = 2\beta$.

Proposition 3.4 *For all $\beta \geq 1$, we have the following convergence in distribution as n goes to infinity:*

$$\left(|T_n|^{1/4}(Z_\beta^{(n)} - Z_\beta),\, Z_\beta^{(n)}\right) \to \left((2\alpha)^{-1/4}\sqrt{Z_{2\beta}}\; G, Z_\beta\right),$$

where G is a centered reduced Gaussian random variable independent of the excursion e.

The contribution to the fluctuations is given by the error of approximation of $A_{n,1}(f)$ by $A_{n,2}(f)$ with $f(x) = x^{\beta-1}$, see notations from the proof of Theorem 3.1. This corresponds to the fluctuations coming from the approximation of the branch lengths $(h_v, v \in T_n)$ by their mean, which relies on the explicit representation on their joint distribution given in Lemma 2.1. In particular, there is no other contribution to the fluctuations from the approximation of the continuum tree $\mathcal{T}_e$ by the sub-tree $\mathcal{T}_n$.

References

[1] R. Abraham and J.-F. Delmas. Record process on the continuum random tree. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(1):225–251, 2013.

[2] D. Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991.

[3] 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.

[4] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[5] P. Biane. Relations entre pont et excursion du mouvement brownien réel. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(1):1–7, 1986.

[6] M. Blum, O. François, and S. Janson. The mean, variance and limiting distriution of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Probab.*, 16(4):2195–2214, 2006.

[7] L. Devroye. Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.*, 32(1):152–171, 2002.

[8] R. P. Doborw and J. A. Fill. Total path length for random recursive trees. *Combin. Probab. Comput.*, 8(4):317–333, 1999.

[9] T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.

[10] T. Duquesne and J.-F. Le Gall. *Random trees, Lévy processes and spatial branching processes*, volume 281 of *Astérisque*. 2002.

[11] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Th. and Rel. Fields*, 131:553–603, 2005.

[12] S. Evans. *Probability and real trees*, volume 1920 of *Lecture Notes in Math.* Springer, 2008.

[13] J. A. Fill, P. Flajolet, and N. Kapur. Singularity analysis, Hadamard products, and tree recurrences. *J. Comput. Appl. Math.*, 174(2):271–313, 2005.

[14] J. A. Fill and S. Janson. Precise logarithmic asymptotics for the right tails of some limit random variables for random trees. *Ann. Comb.*, 12(4):403–416, 2009.

[15] J. A. Fill and N. Kapur. A repertoire for additive functionals of uniformly distributed m-ary search trees (extended abstract). In *2005 International Conference on Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AD, pages 105–114 (electronic).

[16] J. A. Fill and N. Kapur. Limiting distributions for additive functionals on Catalan trees. *Theoret. Comput. Sci.*, 326(1-3):69–102, 2004.

[17] J. A. Fill and N. Kapur. Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Structures Algorithms*, 26(4):359–391, 2005.

[18] D. J. Ford. Probabilities on cladograms: introduction to the alpha model. *ArXiv preprint*, 2005.

[19] C. Holmgren and S. Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Elect. J. Probab.*, 20:1–51, 2005.

[20] H.-K. Hwang and R. Neininger. Phase change of limit laws in the ”Quicksort” recurrence under varying toll functions. *SIAM J. Comput.*, 31(6):1687–1722, 2002.

[21] S. Janson. The Wiener index of simply generated random trees. *Random Struct. Algo.*, 22(4):337–358, 2003.

[22] S. Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Struct. Algo.*, To appear.

[23] S. Janson and P. Chassaing. The center of mass of the ISE and the Wiener index of trees. *Electron. Comm. Probab.*, pages 178–187, 2004.

[24] N. Kapur. *Additive functionals on random search trees*. ProQuest LLC, Ann Arbor, MI, 2003. Thesis (Ph.D.)–The Johns Hopkins University.

[25] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.

[26] H. M. Mahmoud. Limiting distribution for path lengths in recursive trees. *Probab. Engin. Inform. Sci.*, 5(1):53–59, 1991.

[27] A. Mir, F. Rossellò, and L. Rotger. A new balance index for phylogenetic trees. *Mathematical Biosciences*, 241(1):125 – 136, 2013.

[28] R. Neininger. On binary search tree recursions with monomials as toll functions. *J. Comput. Appl. Math.*, 142(1):185–196, 2002.

[29] R. Neininger. The Wiener index of random trees. *Combin., Probab. Comput.*, 11(6):587–597, 2002.

[30] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Math.* Springer, 2002.

[31] M. Régnier. A limiting distribution for ”Quicksort”. *RAIRO Theoret. Inform. and Appl.*, 23:335–343, 1989.

[32] U. Rösler. A limit theorem for ”Quicksort”. *RAIRO Theoret. Inform. and Appl.*, 25:85–100, 1991.

[33] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.

[34] L. Takàcs. On the total heights of random rooted binary trees. *J. Combin. Theory Ser. B*, 61(2):155–166, 1994.