Cost functionals for large random trees

Marion Sciauveau
Joint work with J-F. Delmas and J-S. Dhersin

CERMICS (ENPC) and LAGA (Paris 13)

Séminaire des doctorants - CERMICS - 26 septembre 2016

Some notations for binary trees

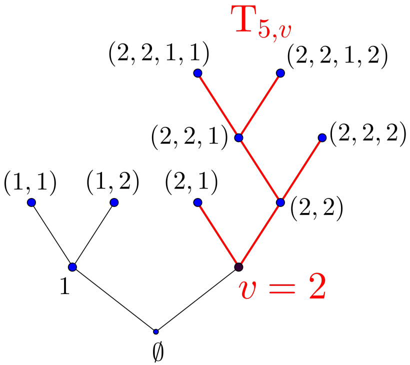


Figure: Binary trees with 5 internal nodes

- T_n rooted full binary ordered tree with n internal nodes
- $|\mathrm{T}_n|$: the cardinal of T_n
- $L(\mathrm{T}_n)$: the left-sub-tree of T_n
- $R(\mathrm{T}_n)$: the right-sub-tree of T_n
- the sub-tree $\mathrm{T}_{n,v}$ of T_n with root v

Random binary trees

A random binary tree is a binary tree selected at random from some probability distribution on binary trees. We often used two models of probability:

1. **Catalan model**: random tree uniformly distributed among the full binary ordered trees with given number of internals nodes. In others words, the probability that a particular tree occurs is $\frac{1}{C_n}$ where C_n is the n^{th} Catalan number.

2. **Random permutation model**: binary search trees are recursive binary trees with keys associated with the internal nodes. The keys are given by a given random permutation of the numbers $\{1, 2, \cdots, n\}$. In the RPM model, each permutation is equally likely.

Comparison between the two models

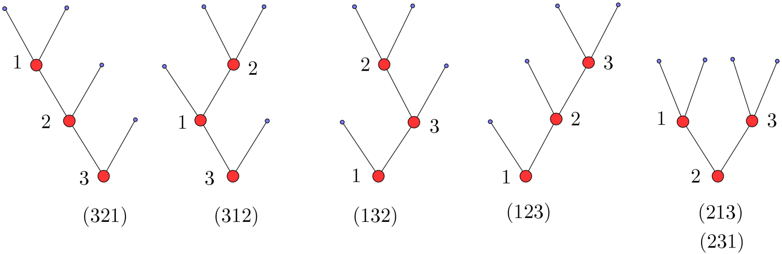


Figure: $C_3 = 5$ rooted full binary ordered trees with 3 internal nodes

Definition of cost functionals

Additive functional

A functional F on binary trees is called an additive functional if it satisfies the following recurrence relation:

$$F(\mathrm{T}) = F(L(\mathrm{T})) + F(R(\mathrm{T})) + b_{|\mathrm{T}|}$$

for all trees T such that $|\mathrm{T}| \geq 1$ and with $F(\emptyset) = 0$

Remark:

1. $(b_n, n \geq 1)$ is called the toll function
2. $$F(\mathrm{T}_n) = \sum_{v \in \mathrm{T}_n} b_{|\mathrm{T}_{n,v}|}$$

Some examples of additive functionals

Index	Total size	Total path length	Wiener index
Expression	$\lvert \mathrm{T}_n \rvert$	$P(\mathrm{T}_n) = \sum_v d(\emptyset, v)$	$W(\mathrm{T}_n) = \sum_{u,v} d(u,v)$
Toll function	$b_n = 1$	$b_n = n$	$b_n = n$ and $b_n = n^2$
Additive functional	$\sum_v 1$	$P(\mathrm{T}_n) = \sum_v \lvert \mathrm{T}_{n,v} \rvert - \lvert \mathrm{T}_n \rvert$	$W(\mathrm{T}_n) = 2\lvert \mathrm{T}_n \rvert \sum_w \lvert \mathrm{T}_{n,w} \rvert - 2\sum_w \lvert \mathrm{T}_{n,w} \rvert^2$
Scaling factor		$\lvert \mathrm{T}_n \rvert^{-\frac{1}{2}}$	$\lvert \mathrm{T}_n \rvert^{-\frac{3}{2}}$
Asymptotics (a.s.)		$2\int_0^1 e(s)ds$	$4\int_0^1 e(s)ds - 4\int_{[0,1]^2} ds\,dt\, m_e(s,t)$

Motivation (1)

Goal: study the asymptotics of cost functional with toll function of type $b_n = n^\beta$ for $\beta > 0$.

Question: For $\beta > 0$,

$$\underbrace{|\mathrm{T}_n|^{?}}_{\text{scaling factor}} \underbrace{\sum_{v \in \mathrm{T}_n} |\mathrm{T}_{n,v}|^\beta}_{\text{additive functional}} \xrightarrow[n\to\infty]{} ?$$

- For $\beta > 0$, **Fill and Kapur** (2003) showed that this functional converges in distribution, after a suitable scaling, to a limit Z_β. The limit is characterized by its moments.
- **Fill and Janson** (2007) announced that for $\beta > \frac{1}{2}$, Z_β can be represented as a functional of the normalized Brownian excursion e.

Motivation (2)

- We study

$$\underbrace{|\mathrm{T}_n|^{-\frac{3}{2}}}_{\text{scaling factor}} \underbrace{\sum_{v \in \mathrm{T}_n} |\mathrm{T}_{n,v}| f\left(\frac{|\mathrm{T}_{n,v}|}{|\mathrm{T}_n|}\right)}_{\text{normalized additive functional}}$$

  for f satisfying smooth conditions

- **Aim**: derive an invariance principle for such tree functionals

- **Model**: Catalan model

Brownian tree associated to the normalized Brownian excursion

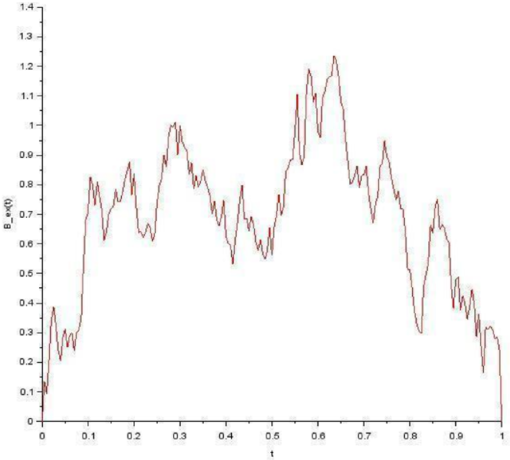
Let e be a normalized Brownian excursion on $[0,1]$ i.e. a standard Brownian motion on $[0,1]$ conditioned on being nonnegative on $[0,1]$ and on taking the value 0 at 1.

For $s,t \in [0,1]$, $s < t$, we define

$$d_e(s,t) = e(s) + e(t) - 2 \inf_{s<u<t} e(u)$$

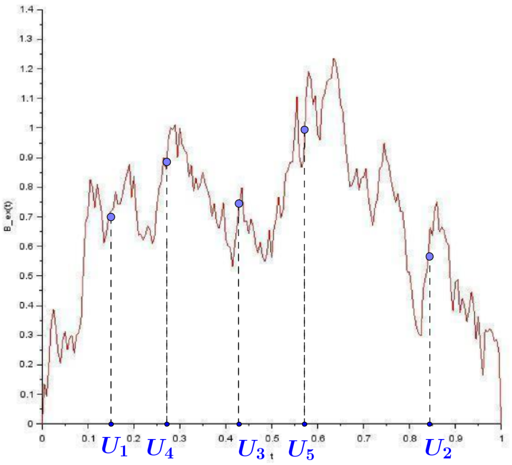
The Browian tree is defined as $\mathcal{T}_e = [0,1]/\sim_e$ where $s \sim_e t \Leftrightarrow d_e(s,t) = 0$ and we let d_e be the induced distance on the quotient. We denote by $\mathbf{p}$ the canonical projection from $[0,1]$ to $\mathcal{T}_e$.

Natural embedding of binary trees into the Brownian excursion e (1)



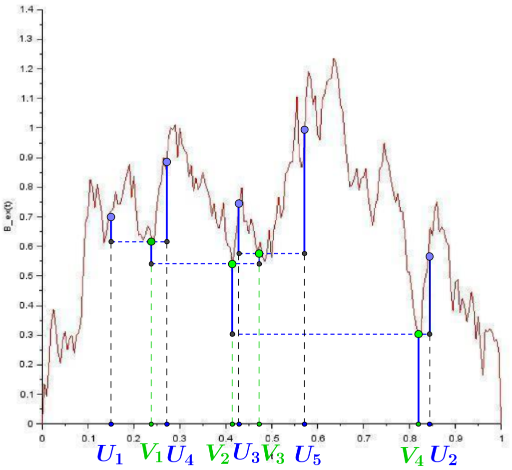
- Normalized Brownian excursion: e

Natural embedding of binary trees into the Brownian excursion e (1)



- Normalized Brownian excursion: e
- $(U_i)_{1 \leq i \leq 5}$ i.i.d. uniform on $[0, 1]$ and indep. of e

Natural embedding of binary trees into the Brownian excursion e (1)



- Normalized Brownian excursion: e
- $(U_i)_{1\leq i\leq 5}$ i.i.d. uniform on $[0,1]$ and indep. of e
- $(V_i)_{1\leq i\leq 4}$ such that

$$e(V_i) = \min_{u\in[U_{(i)},U_{(i+1)}]} e(u)$$

Natural embedding of binary trees into the Brownian excursion e (2)

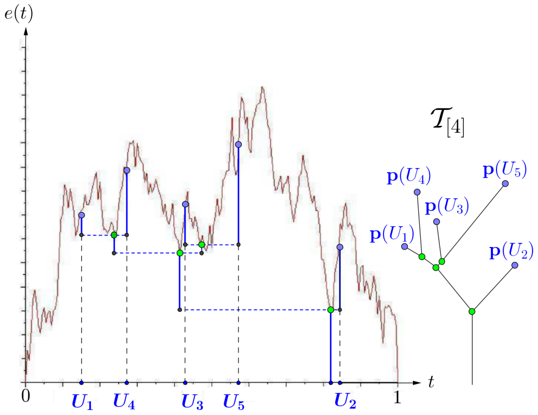


Figure: The Brownian excursion and $\mathcal{T}_{[n]}$ for $n = 4$

Natural embedding of binary trees into the Brownian excursion e (2)

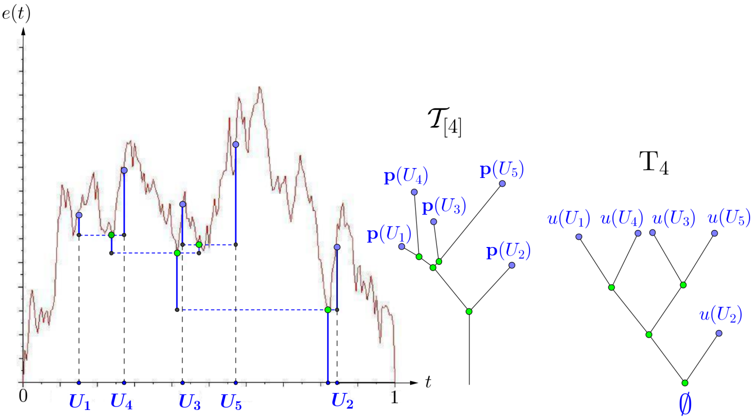
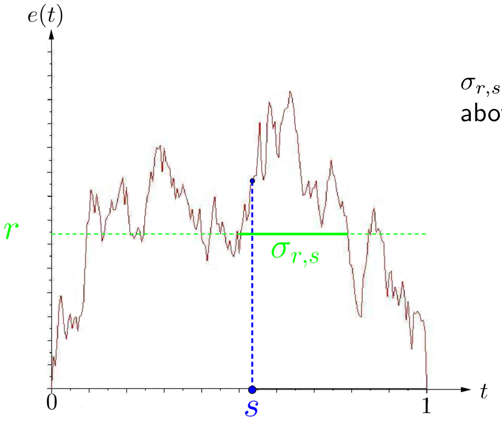


Figure: The Brownian excursion, $\mathcal{T}_{[n]}$ (for $n = 4$) and T_n

Length of a subexcursion



$\sigma_{r,s}$ = length of the excursion of e above level r straddling s

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

Invariance principle

Let

$$A_n(f) = |\mathrm{T}_n|^{-\frac{3}{2}} \sum_{v \in \mathrm{T}_n} |\mathrm{T}_{n,v}| f\left(\frac{|\mathrm{T}_{n,v}|}{|\mathrm{T}_n|}\right)$$

and

$$\Phi_e(f) = \int_0^1 ds \int_0^{e_s} dr \ f(\sigma_{r,s})$$

Theorem

A.s., $\forall f \in \mathcal{C}^0((0,1])$ *s.t.* $\lim_{x\downarrow 0^+} x^a f(x) = 0$ *for some* $0 \le a < \frac{1}{2}$*, we have:*

$$\lim_{n\to+\infty} A_n(f) = 2\,\Phi_e(f)$$

Application for $f(x) = x^{\beta-1}$

For$\beta > 0$ and $n \in \mathbb{N}^*$, we set:

$$Z_\beta = \int_0^1 ds \int_0^{e_s} dr\, \sigma_{r,s}^{\beta-1} \quad \text{and} \quad Z_\beta^{(n)} = |\mathrm{T}_n|^{-(\beta+\frac{1}{2})} \sum_{v\in\mathrm{T}_n} |\mathrm{T}_{n,v}|^\beta$$

Theorem

We have a.s., $\forall\beta > 0$,

$$\lim_{n\to+\infty} Z_\beta^{(n)} = 2\, Z_\beta$$

Application for $f(x) = x^{\beta-1}$

For $\beta > 0$ and $n \in \mathbb{N}^*$, we set:

$$Z_\beta = \int_0^1 ds \int_0^{e_s} dr\, \sigma_{r,s}^{\beta-1} \quad \text{and} \quad Z_\beta^{(n)} = |\mathrm{T}_n|^{-(\beta+\frac{1}{2})} \sum_{v \in \mathrm{T}_n} |\mathrm{T}_{n,v}|^\beta$$

Theorem

We have a.s., $\forall \beta > 0$,

$$\lim_{n \to +\infty} Z_\beta^{(n)} = 2\, Z_\beta$$

Lemma

If $\beta > \frac{1}{2}$,

$$\textit{a.s. } Z_\beta < +\infty \textit{ and } \mathbb{E}[Z_\beta] < +\infty$$

Otherwise,

$$\textit{a.s. } Z_\beta = +\infty$$

Fluctuations of the invariance principle

Theorem

Let $f \in \mathcal{C}([0,1])$ *be locally Lipschitz continuous on* $(0,1]$ *with* $\|x^a f^{'}\|_{essup} < +\infty$ *for some* $a \in (0,1)$*. We have*

$$\left(\underbrace{|T_n|^{1/4}}_{speed\ of\ CV} (A_n - 2\Phi_e)(f),\ A_n \right) \xrightarrow[n\to\infty]{\mathcal{L}} \left(\sqrt{2}\sqrt{\Phi_e(xf^2)}\, G, 2\Phi_e \right),$$

where $G \sim \mathcal{N}(0,1)$ *and is independent of the excursion* e

Results and ongoing work

- **Results**:
  - invariance principle for more general additive functional
  - recover some classical results on additive functional (e.g. total size, total path length ...)
  - fluctuations coming from the approximation of the branch lengths by their mean.
- **Ongoing work**: study asymetric cost functionals depending on the cardinal of the left and right sub-tree of each nodes.

Thank you for your attention !