Functionals of the size and height on conditioned BGW trees

Michel Nassif
with R. Abraham and J.-F. Delmas

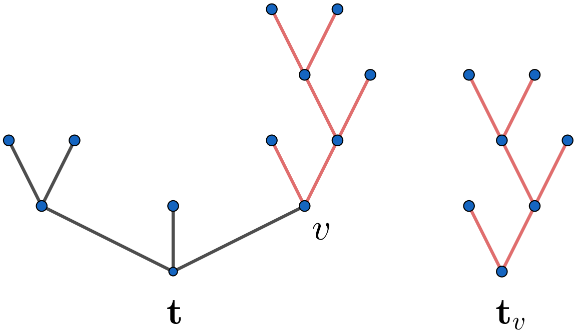
CERMICS (ENPC)

August 2020

Notations for finite trees

Let $\mathbf{t}$ be a finite rooted ordered tree.

- For every vertex $v \in \mathbf{t}$, $\mathbf{t}_v$ is the subtree of $\mathbf{t}$ above v.



- $|\mathbf{t}|$ is its size, $\mathfrak{h}(\mathbf{t})$ its height.
- $\mathbf{t}^\circ$ is the set of internal vertices of $\mathbf{t}$.

Framework

We consider τ^n a Bienaymé-Galton-Watson (BGW) tree conditioned to have size n with offspring distribution ξ satisfying:

- ξ is critical, i.e. $\mathbb{E}[\xi] = 1$,
- ξ belongs to the domain of attraction of a stable distribution with index $\gamma \in (1, 2]$, i.e. $\exists (b_n,\ n \geqslant 1)$ such that

$$\frac{1}{b_n}\left(\sum_{k=1}^{n} \xi_k - n\right) \xrightarrow[n\to\infty]{(d)} X,$$

  where $(\xi_n,\ n \geqslant 1)$ is a sequence of i.i.d. random variables with the same distribution as ξ and X has Laplace transform $\mathbb{E}[\exp(-\lambda X)] = \exp(\lambda^\gamma)$,
- the sequence $(b_n,\ n \geqslant 1)$ satisfies

$$\underline{b}\, n^{1/\gamma} \leqslant b_n \leqslant \overline{b}\, n^{1/\gamma}, \quad \forall n \geqslant 1,$$

  for some constants $0 < \underline{b} < \overline{b} < \infty$.

Additive functionals

A functional F on (finite rooted ordered) trees is called additive if it satisfies the recursion

$$F(\mathbf{t}) = \sum_{i=1}^{d} F(\mathbf{t}_i) + f(\mathbf{t}),$$

where $\mathbf{t}_1, \ldots, \mathbf{t}_d$ are the branches (= subtrees rooted at the children of the root) of $\mathbf{t}$ and f is a toll function.

Remark.

$$F(\mathbf{t}) = \sum_{w \in \mathbf{t}} f(\mathbf{t}_w),$$

where $\mathbf{t}_w$ is the subtree of $\mathbf{t}$ above w.

Examples of additive functionals

- Total path length

$$P(\mathbf{t}) = \sum_{v \in \mathbf{t}} d(\emptyset, v) = \sum_{w \in \mathbf{t}} |\mathbf{t}_w| - |\mathbf{t}|.$$

- Wiener index

$$W(\mathbf{t}) = \sum_{u,v \in \mathbf{t}} d(u, v) = 2|\mathbf{t}| \sum_{w \in \mathbf{t}} |\mathbf{t}_w| - 2 \sum_{w \in \mathbf{t}} |\mathbf{t}_w|^2.$$

- Shao and Sokal's B_1 index [Shao and Sokal (1990)] used to assess the balance of a phylogenetic tree

$$B_1(\mathbf{t}) = \sum_{\substack{w \in \mathbf{t}^\circ \\ w \neq \emptyset}} \frac{1}{\mathfrak{h}(\mathbf{t}_w)},$$

where $\mathbf{t}^\circ$ is the set of internal vertices of $\mathbf{t}$.

Known results in the global regime

- **Fill and Kapur** (2003) showed that for the Catalan model (i.e. $\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 2) = 1/2$) for $\alpha > 1/2$,

$$n^{-(\alpha+1/2)} \sum_{w \in \tau^n} |\tau_w^n|^\alpha \xrightarrow[n\to\infty]{(d)} Z_\alpha,$$

  where Z_α is characterized by its moments.
- **Fill and Janson** (2007) announced that Z_α can be represented as a functional of the Brownian excursion and later generalized the result for arbitrary offspring distribution with finite variance.
- When the offspring distribution has infinite variance but lies in the domain of attraction of a stable distribution, **Delmas, Dhersin and Sciauveau** (2018) proved convergence for $\alpha \geqslant 1$ towards a functional of the stable Lévy tree and conjectured a phase transition at $\alpha = 1/\gamma$.

Scaling limits of BGW trees

The stable Lévy tree with index $\gamma \in (1, 2]$ is the real tree coded by the height process with index γ, see Le Gall and Le Jan (1998), Duquesne and Le Gall (2002).

Assume that the offspring distribution is critical and lies in the domain of attraction of a stable distribution with index $\gamma \in (1, 2]$.

Theorem (Aldous (1991), Duquesne (2003))

Let τ^n be a BGW(ξ) conditioned to have n vertices. We have the convergence in distribution

$$\frac{b_n}{n}\tau^n \xrightarrow[n\to\infty]{(d)} \mathcal{T}$$

for the Gromov-Hausdorff-Prokhorov distance, where $\mathcal{T}$ is the stable Lévy tree with index γ.

Convergence of polynomial functionals of the mass and height

Assume that ξ is critical and lies in the domain of attraction of a stable distribution with index $\gamma \in (1,2]$. Furthermore, assume that $\underline{b}n^{1/\gamma} \leqslant b_n \leqslant \overline{b}n^{1/\gamma}$.

Theorem

- *If* $\gamma\alpha + (\gamma-1)\beta > 1$, *we have the convergence in distribution and of the first moment*

$$\frac{b_n^{1+\beta}}{n^{1+\alpha+\beta}} \sum_{w\in\tau^{n,\circ}} |\tau_w^n|^\alpha \mathfrak{h}(\tau_w^n)^\beta \xrightarrow[n\to\infty]{(d)+\text{mean}} Z_{\alpha,\beta}.$$

- *If* $\gamma\alpha + (\gamma-1)\beta \leqslant 1$, *we have*

$$\frac{b_n^{1+\beta}}{n^{1+\alpha+\beta}} \sum_{w\in\tau^{n,\circ}} |\tau_w^n|^\alpha \mathfrak{h}(\tau_w^n)^\beta \xrightarrow[n\to\infty]{(d)+\text{mean}} Z_{\alpha,\beta} = \infty.$$

The limit

Remark.

- Phase transition at $\gamma\alpha + (\gamma - 1)\beta = 1$ & compensation.
- One can add leaves when possible ($\beta \geqslant 0$).

$Z_{\alpha,\beta}$ has the following representation in terms of the stable Lévy tree $\mathcal{T}$ with index γ:

$$Z_{\alpha,\beta} = \int_{\mathcal{T}} \mathfrak{m}(\mathcal{T}_y)^{\alpha} \mathfrak{h}(\mathcal{T}_y)^{\beta} \, \ell(\mathrm{d}y),$$

where ℓ is the length measure on $\mathcal{T}$. Here $\mathcal{T}_y$ is the subtree above y, $\mathfrak{m}(\mathcal{T}_y)$ is its mass and $\mathfrak{h}(\mathcal{T}_y)$ its height.

First moment of $Z_{\alpha,\beta}$

Proposition

For every $\alpha, \beta \in \mathbb{R}$ such that $\gamma\alpha + (\gamma - 1)\beta > 1$, we have

$$\mathbb{E}\left[Z_{\alpha,\beta}\right] = \frac{1}{|\Gamma(-1/\gamma)|} \mathrm{B}(\alpha + \beta(1 - 1/\gamma) - 1/\gamma, 1 - 1/\gamma)\, \mathbb{E}\left[\mathfrak{h}(\mathcal{T})^{\beta}\right].$$

In particular, in the Brownian case ($\gamma = 2$), we have

$$\mathbb{E}\left[Z_{\alpha,\beta}\right] = \pi^{(\beta-1)/2}\xi(\beta)\, \mathrm{B}\left(\alpha + \beta/2 - 1/2, 1/2\right),$$

where ξ is the Riemann xi function defined by

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s), \quad \forall s \in \mathbb{C}.$$

Recap of the finite variance case

Assume that ξ is critical and has finite variance. Asymptotic behavior of

$$\sum_{w \in \tau^{n,\circ}} |\tau_w^n|^\alpha \mathfrak{h}(\tau_w^n)^\beta$$

- $2\alpha + \beta > 1$: converges in distribution to $Z_{\alpha,\beta}$ (functional of the Brownian excursion) after rescaling by $n^{\alpha+(1+\beta)/2}$.
- $2\alpha + \beta < 0$: CLT after recentering and rescaling by $\sqrt{n}$ using Janson (2016) & Addario-Berry, Devroye and Janson (2013).
- $0 \leqslant 2\alpha + \beta \leqslant 1$: only known for $\beta = 0$.

Remark. Shao and Sokal's B_1 index ($\alpha = 0, \beta = -1$) lies in the region $2\alpha + \beta < 0$.