Sums of Powers of Subtree Sizes in Conditioned Critical Galton–Watson Trees

Jim Fill

Department of Applied Mathematics and Statistics
The Johns Hopkins University

AofA2023

Academia Sinica
Taipei, Taiwan

June 26, 2023

Origin of my work on conditioned critical Galton–Watson trees; Nevin Kapur

- The foundation for a significant portion of the work I will discuss today was built in joint work with PhD advisee Nevin Kapur (PhD 2003), especially in these three papers:

  F, Philippe Flajolet, and N. Kapur. Singularity analysis, Hadamard products, and tree recurrences. *Journal of Computational and Applied Mathematics*, (2004), **174/2**, 271–313;

  F and N. Kapur. Limiting distributions for additive functionals on Catalan trees. *Theoretical Computer Science*, (2004), **326**, 69–102;

  F and N. Kapur. A repertoire for additive functionals of uniformly distributed *m*-ary search trees. Refereed extended abstract: pages 105–114 in *2005 International Conference on the Analysis of Algorithms* (ed.: Conrado Martínez), *Discrete Mathematics and Theoretical Computer Science Proceedings*, **AD**, 2005.

Superb background reference and my latest trees research

- The background material I will present at the beginning of the talk is adapted from this comprehensive & beautiful survey:

  Svante Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys* (2012), **9**, 103–252.

  I won't review even the main result there—rather, just material that's needed as background for recent work.

- The later material is very recent (and some ongoing!) work:

  F and S. Janson. The sum of powers of subtree sizes for conditioned Galton–Watson trees. *Electron. J. Probab.*, (2022), **27**, Paper No. 114, 77 pages (with a corrigendum to appear);

  F, S. Janson, and Stephan Wagner. Conditioned Galton–Watson trees: The shape functional, and more on the sum of powers of subtree sizes and its mean. 50 pages. Submitted in January, 2023 to a special Analysis of Algorithms issue of *La Matematica*.

Outline of talk

Here's an outline for today's talk:

1 Background on trees
  1.1 (Conditioned) Galton–Watson trees and criticality
  1.2 Additive functionals
    1.2.1 Sums of powers of subtree sizes
    1.2.2 The "shape functional" (i.e., log-product of subtree sizes)

2 Distributional asymptotics for sums of powers of subtree sizes
  2.1 Limit laws
  2.2 Convergence of moments

3 Comparing means across Galton–Watson models:
  additive functionals with completely monotone toll

1.1. (Conditioned) Galton–Watson trees and criticality

- Galton–Watson trees are defined as the family trees of Galton–Watson processes: Given an offspring r.v. ξ with pmf $\mathbf{p} = (p_k)_{k \geq 0}$ on $\mathbb{Z}_{\geq 0}$, we build the tree $\mathcal{T}$ recursively,
starting with the root and giving each node a number of children that is an independent copy of ξ. In other words, the outdegrees $d^+(v)$ are i.i.d. with the distribution $\mathbf{p}$.
- Recall that the Galton–Watson process is called *subcritical*, *critical*, or *supercritical* as the expected number of children $\mathbb{E}\,\xi = \sum_{k=0}^{\infty} kp_k$ satisfies $\mathbb{E}\,\xi < 1$, $\mathbb{E}\,\xi = 1$, or $\mathbb{E}\,\xi > 1$.
- standard basic fact of branching process theory: $\mathcal{T}$ is finite a.s. if $\mathbb{E}\,\xi \leq 1$ (subcritical and critical cases), but $\mathcal{T}$ is infinite with positive probability if $\mathbb{E}\,\xi > 1$ (supercritical case).
- The GWTs have random sizes. We are mainly interested in random trees with a given size; we thus define $\mathcal{T}_n$ as $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$. These random trees $\mathcal{T}_n$ are called *conditioned Galton–Watson trees*, and are a (not very) special case of Meir–Moon simply generated trees.

1.1.1 Why tilt to *critical* Galton–Watson trees?

- Thus we can generate G–W trees to sample from conditioned G–W trees (simply generated trees)—by conditioning on size!
- For subcritical/supercritical G–W trees, one can show that the tails of the distribution of $|\mathcal{T}|$ (given $|\mathcal{T}| < \infty$) decay exponentially. (So the conditioning probability is exponentially small, which is bad.)
- For critical G–W trees we'll see that the tails of the distribution of $|\mathcal{T}|$ decay like $\mathbb{P}(|\mathcal{T}| = n) = \Theta(n^{-3/2})$ whenever the offspring distribution has finite variance. This is a nice fat tail: We even have $\mathbb{E}\,|\mathcal{T}| = \infty$.
- Supercritical G–W trees can *always* be "exponentially tilted" to critical, and the same is usually true for subcritical.
- For G–W trees, we usually write $q_n = \mathbb{P}(|\mathcal{T}| = n)$ and $y(z) := \sum_n q_n z^n$ for the corresponding pgf. (awkward but standard!)

1.1.1. Why tilt to *critical* Galton–Watson trees? (cont.)

Consider a critical Galton–Watson tree, with offspring r.v. ξ having pgf Φ. How do we see that $q_n = \mathbb{P}(|\mathcal{T}| = n) = \Theta(n^{-3/2})$ if $\sigma^2 := \operatorname{Var} \xi < \infty$? A key equation, resulting from the recursive nature of $\mathcal{T}$, is

$$y(z) = \sum_{n=1}^{\infty} q_n z^n = z\Phi(y(z)).$$

That is, y is the inverse of the function $t \mapsto t/\Phi(t)$. By Lagrange inversion,

$$q_n = n^{-1} [t^{n-1}]\Phi(t)^n = n^{-1} \mathbb{P}(S_n = n - 1),$$

where S_n is the sum of n i.i.d. copies of the offspring r.v. ξ. By the local CLT, this implies

$$q_n \sim (2\pi\sigma^2)^{-1/2} n^{-3/2},$$

assuming for simplicity that ξ has span 1. (It's easy to adjust otherwise.)

Standardizing to *critical* Galton–Watson trees also allows fair comparisons across offspring distributions!

Many important examples of uniform distributions over various classes of trees can be formulated as critGWTs.

Examples:

- (a) $\xi \sim \text{Geometric}\left(\frac{1}{2}\right)$ (with support $\{0, 1, \dots\}$).

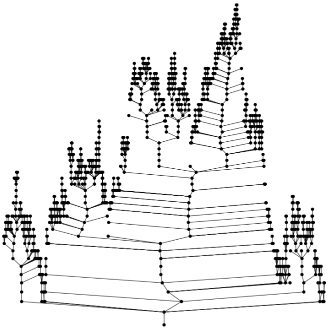
  Then $\mathcal{T}_n$ is uniformly distributed over ordered rooted trees (ORTs) with n nodes.

Here's a uniformly random ORT of size 1000 (root is at bottom):

ALEA, Lat. Am. J. Probab. Math. Stat. **16**, 561–604 (2019)
DOI: 10.30757/ALEA.v16-21

Inference for conditioned Galton-Watson trees from their Harris path

Romain Azaïs, Alexandre Genadot and Benoit Henry



1.1.2. Examples of critGWTs (continued)

- (b) $\xi \sim$ Poisson(1).

  If we generate $\mathcal{T}_n$ this way, label the n nodes randomly with $\{1, \ldots, n\}$, and ignore birth order, then the resulting distribution is uniform over labelled unordered rooted trees (LURTs) with n nodes.

  Remark: The uniform distribution on UURTs (*unlabelled* unordered rooted trees) can't be formulated in terms of GWTs.

1.1.2. Examples of critGWTs (conclusion)

- (c) Full binary trees: $\xi \sim 2\,\text{Bernoulli}(\frac{1}{2})$.

  This example can be generalized to full m-ary trees, with
  $$\xi \sim m\,\text{Bernoulli}\left(\frac{1}{m}\right).$$

- (d) Binary trees (distinguishing L & R only-children):
  $$\xi \sim \text{Binomial}(2, \tfrac{1}{2}).$$

  This example can be generalized to m-ary trees, with
  $$\xi \sim \text{Binomial}\left(m, \frac{1}{m}\right).$$

1.2. Additive functionals; sums of powers of subtree sizes

- *Additive functionals* are important to study for random trees.
- These are functionals of rooted trees of the type

$$F(T) := \sum_{v \in T} f(T_v),$$

where T_v is the *fringe subtree* consisting of v and all its descendants, and f is a given functional on trees, often called the *toll function*.

- Equivalently, additive functionals may be defined by the recursion

$$F(T) := f(T) + \sum_{i=1}^{d} F(T_{v(i)}),$$

where d is the degree of the root o of T and $v(1), \ldots, v(d)$ are the children of o. Note that, due to $\sum$, sibling order no longer matters!

- main interest: asymptotics for the distribution of $F(\mathcal{T}_n)$ as $n \to \infty$

1.2.1. Sums of powers of subtree sizes

- today's talk: critGWTs with $0 < \sigma^2 = \operatorname{Var} \xi < \infty$ (sometimes silently assuming $\mathbb{E}\, \xi^{2+\delta} < \infty$ for some $\delta > 0$) and (mainly) tolls $f_\alpha(T) := |T|^\alpha$ for some constant $\alpha \in \mathbb{C}$
- We denote the corresponding additive functional by F_α.
- important special cases: Note that for $\alpha = 0$, we trivially have $F_0(\mathcal{T}_n) = n$. The case $\alpha = 1$ yields the *total pathlength*, important in computer science, and from $\alpha = 1$ and $\alpha = 2$ (jointly) we can relate results to the so-called *Wiener index*[1], important in chemistry.
- Our results have interesting connections to the continuum random tree of David Aldous and to Brownian excursion.
- These particular functionals serves as prototypes for functionals favoring small ($\operatorname{Re} \alpha < 0$) and large ($\operatorname{Re} \alpha > 0$) subtrees.

[1] named for the Austrian–American chemist Harry Wiener (1924–1998)—not for Norbert Wiener!

1.2.2. The "shape functional"

- f_0 and F_0 are trivial! However, ...
- $f(T) := \ln|T|$ and $F(T) := \sum_v \ln|T_v|$ are the respective derivatives of f_α and F_α at $\alpha = 0$ and are nontrivial.

  We call this F the *shape functional*.
- I introduced this functional in the context of a different (non-SGT) model of random trees, namely, (`QuickSort`-related) binary search trees under the random permutation model, in F (1996, RSA). There the functional is entropy-related and serves as a crude measure of the "shape" of the random tree.
- Meir and Moon (1998) started the study of F for simply generated trees by finding asymptotics for the mean and variance of $F(\mathcal{T}_n)$.

2.1. Limit Laws

Previous papers have studied limiting distributions in various special cases when α is real, but we consider these variables for arbitrary complex α.

This is advantageous, even for the study of real α, since it allows us to use powerful results from the theory of analytic functions in the proofs.

We also find new phenomena for non-real α.

There's not nearly enough time for even an outline of how things turn out for various values of α.

However, as one illustration, I will state the normal limit-law result for $\operatorname{Re}\alpha < 0$ on the next slide.

2.1.2. Normal limit law for $\operatorname{Re}\alpha < 0$

Theorem 1.1 in F and Janson (2022)

Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree defined by an offspring distribution ξ with $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 := \operatorname{Var}\xi < \infty$. Then there exists a family of centered complex normal random variables $X(\alpha)$, $\operatorname{Re}\alpha < 0$, such that, as $n \to \infty$,

$$\frac{F_\alpha(\mathcal{T}_n) - \mathbb{E}\,F_\alpha(\mathcal{T}_n)}{\sqrt{n}} \xrightarrow{\mathrm{d}} X(\alpha), \qquad \operatorname{Re}\alpha < 0. \tag{1}$$

Moreover, $X(\alpha)$ is a (random) analytic function of α, and the convergence (1) holds in the space $\mathcal{H}(H_-)$ of analytic functions in the left half-plane $H_- := \{\alpha : \operatorname{Re}\alpha < 0\}$. Furthermore,

$$\overline{X(\alpha)} = X(\overline{\alpha}), \qquad \alpha \in H_-.$$

The covariance function $\mathbb{E}\left[X(\alpha)X(\beta)\right]$ is an analytic function of two variables $\alpha, \beta \in H_-$, and, as $n \to \infty$,

$$n^{-1}\operatorname{Cov}\bigl(F_\alpha(\mathcal{T}_n), F_\beta(\mathcal{T}_n)\bigr) \to \mathbb{E}\bigl[X(\alpha)X(\beta)\bigr], \qquad \alpha, \beta \in H_-. \quad \square$$

2.2. Convergence of moments for $\operatorname{Re}\alpha < 0$, and more

In FJW (2023), we further complete the picture for $\operatorname{Re}\alpha < 0$ by

(i) proving convergence of moments in the above theorem, notably using the complex-analytic technique of singularity analysis, especially including the results on Hadamard products of power series by F, Flajolet, and Kapur (2004);

(ii) establishing asymptotic normality for the additive "shape functional", with toll $f(T) = \log |T|$;

(iii) establishing asymptotic normality for imaginary $\alpha \neq 0$ (as is true for $\operatorname{Re}\alpha < 0$), with asymptotic variance that depends only on σ^2 (as is true for $\operatorname{Re}\alpha > 0$).

The asymptotic variance for the shape functional and for imaginary α is (not *too* surprisingly) of order $n \log n$
(rather than order n for $\operatorname{Re}\alpha < 0$ and order $n^{1+2\operatorname{Re}\alpha}$ for $\operatorname{Re}\alpha > 0$).

3. Comparing means across Galton–Watson models

Due to lack of time, I will be even briefer in this last section of the talk.

- Recall that for $\operatorname{Re}\alpha < 0$, to get a limiting Gaussian process we need to center $F_\alpha(\mathcal{T}_n)$ by its mean and then divide by $\sqrt{n}$.
- With $\mu(\alpha) := \mathbb{E}\,|\mathcal{T}|^\alpha$, one can show that
$$\mathbb{E}\,F_\alpha(\mathcal{T}_n) = \mu(\alpha)n + o(\sqrt{n}),$$
so we can center instead by $\mu(\alpha)n$.
- This leads to study of $\mu(\alpha)$, which for general α with $\operatorname{Re}\alpha < 0$ can be computed only numerically.
- However, when α is a negative integer, it was shown in FJ (2022) that $\mu(\alpha)$ can be computed explicitly for certain examples of critGWTs.
- We noticed that, for any two of our four examples, we had $\mu_1(\alpha) < \mu_2(\alpha)$ for all such α and some ordering $(1, 2)$ of the two examples and wondered why.
- We built a comparison theory to explain this phenomenon.

3.1 Comparing means across Galton–Watson models: a whiff of the theory

- We built a comparison theory to explain this phenomenon. The full story? I don't have enough time to explain.
- The essence is this theorem: Given two critical Galton–Watson offspring distributions ξ_1 and ξ_2 with respective probability generating functions Φ_1 and Φ_2, if $\Phi_1(t) < \Phi_2(t)$ for **all** real $t \in (0, 1)$, then $\mu_1(\alpha) < \mu_2(\alpha)$ for all real $\alpha < 0$.
- Remarks:
  (1) This result extends from power-function tolls to completely monotone tolls, and there is a converse of sorts.
  (2) If $\Phi_1 \prec \Phi_2$ are ordered this way, then the means for the shape functional are reverse-ordered.
  (3) Although this is only a partial order on critGWT pgfs, surprisingly many important pgfs are comparable!

3.2. Comparing means for additive functionals: Examples of ordering across Galton–Watson models

Examples of critical Galton–Watson models with ordered pgfs:

Recall these examples:

m-ary trees:	$\xi_{1,m} \sim \mathrm{Bi}(m, \frac{1}{m})$	$(m \geq 2)$;
LURTs:	$\xi_2 \sim \mathrm{Po}(1)$;	
full binary trees:	$\xi_3 \sim 2\,\mathrm{Bi}(1, \frac{1}{2})$;	
ORTs:	$\xi_4 \sim \mathrm{Ge}(\frac{1}{2})$;	
full m-ary trees:	$\xi_{5,m} \sim m\,\mathrm{Bi}(1, \frac{1}{m})$	$(m \geq 3)$.

3.2. Comparison ordering examples (continued)

These examples are all ordered by $\prec$!

Proposition

For every $t \in (0, 1)$ we have

$$\text{m-ary:} \quad \Phi_{1,m}(t) \uparrow \text{ strictly as } m \uparrow,$$
$$\text{full m-ary:} \quad \Phi_{5,m}(t) \uparrow \text{ strictly as } m \uparrow;$$

and, for any $m \geq 2$,

$$\Phi_{1,m} \prec \Phi_2 \text{ (LURTs)} \prec \Phi_3 \text{ (full binary)} \prec \Phi_4 \text{ (ORTs)} \prec \Phi_{5,3}.$$

Proof.

The proof is a collection of simple exercises in calculus. □