Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. (Extended abstract)

Svante Janson[1†]

[1] *Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden*

Abstract. We consider conditioned Galton–Watson trees and show asymptotic normality of additive functionals that are defined by toll functions that are not too large. This includes, as a special case, asymptotic normality of the number of fringe subtrees isomorphic to any given tree, and joint asymptotic normality for several such subtree counts. The offspring distribution defining the random tree is assumed to have expectation 1 and finite variance; no further moment condition is assumed.

Keywords: Fringe subtrees; conditioned Galton–Watson trees; random trees; asymptotic normality; toll functions.

1 Introduction

Given a rooted tree T and a node v in T, let T_v be the subtree of T rooted at v, i.e., the subtree consisting of v and all its descendents. Such subtrees are called *fringe subtrees*. The *random fringe subtree* T_* is the random rooted tree obtained by taking the subtree T_v at a uniformly random node v in T, see Aldous [1].

We let, for $T, T' \in \mathfrak{T}$,

$$n_{T'}(T) := \bigl|\{v \in T' : T_v = T'\}\bigr|, \tag{1.1}$$

i.e., the number of subtrees of T that are equal (i.e., isomorphic to) to T'. Then the distribution of T_* is given by

$$\mathbb{P}(T_* = T') = n_{T'}(T)/|T|, \qquad T' \in \mathfrak{T}. \tag{1.2}$$

Thus, to study the distribution of T_* is equivalent to studying the numbers $n_{T'}(T)$.

A related point of view is to let f be a functional of rooted trees, i.e., a function $f : \mathfrak{T} \to \mathbb{R}$, and for a tree $T \in \mathfrak{T}$ consider the sum

$$F(T) = F(T; f) := \sum_{v \in T} f(T_v). \tag{1.3}$$

Thus,

$$F(T)/|T| = \mathbb{E}\, f(T_*). \tag{1.4}$$

† Partly supported by the Knut and Alice Wallenberg Foundation

One important example of this is to take $f(T) = \mathbf{1}\{T = T'\}$, the indicator function that T equals some given tree $T' \in \mathfrak{T}$; then $F(T) = n_{T'}(T)$ and (1.4) reduces to (1.2). Conversely, for any f,

$$F(T) = \sum_{T' \in \mathfrak{T}} f(T') n_{T'}(T); \tag{1.5}$$

hence any $F(T)$ can be written as a linear combination of the subtree counts $n_{T'}(T)$, so the two points of views are essentially equivalent.

Remark 1.1 Functionals F that can be written as (1.3) for some f are called *additive functionals*. The definition (1.3) can also be written recusively as

$$F(T) = f(T) + \sum_{i=1}^{d} F(T_i), \tag{1.6}$$

where $T_1, \dots, T_d$ are the branches (i.e., the subtrees rooted at the children of the root) of T. In this context, $f(T)$ is often called a *toll function*. (One often considers toll functions that depend only on the size $|T|$ of T, but that is not always the case. We emphasise that we allow more general functionals f.)

Note that when T is a random tree, as it was in [1] and will be in the present paper, $F(T)$ is a random variable. In particular, $n_{T'}(T)$ is a random variable for each $T' \in \mathfrak{T}$, and thus the distribution of T_*, which is given by (1.2), is a random probability distribution on $\mathfrak{T}$. Note that (1.2) now reads

$$\mathbb{P}\bigl(T_* = T' \mid T\bigr) = n_{T'}(T)/|T| \tag{1.7}$$

and that similarly (1.4) then has to be replaced by

$$F(T)/|T| = \mathbb{E}\bigl(f(T_*) \mid T\bigr). \tag{1.8}$$

The random trees that we consider in this paper are conditioned Galton–Watson trees. (Related results for some other random trees are given by Fill and Kapur [9, 10] (m-ary search trees under different models) and Holmgren and Janson [11] (random binary search trees and random recursive trees).) The Galton–Watson trees are defined using an offspring distribution ξ and we assume throughout the paper that $\mathbb{E}\,\xi = 1$ and $\sigma^2 := \operatorname{Var}\xi$ is finite (and non-zero). We denote the probability distribution of ξ by $(p_k)_0^\infty$, i.e., $p_k := \mathbb{P}(\xi = k)$.

The results in Aldous [1] focus on convergence (in probability), as $|T| \to \infty$, of the fringe subtree distribution for suitable classes of random trees T, which by (1.8) is equivalent to convergence of $F(T)/|T|$ or $\mathbb{E}\,F(T)/|T|$ for suitable functionals f. For the conditioned Galton–Watson trees studied here, this is stated in the following theorem from [15, Theorem 7.12], improving earlier results by Aldous [1] and Bennies and Kersting [2].

Theorem 1.2 (Aldous, et al.) *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree with n nodes, defined by an offspring distribution ξ with $\mathbb{E}\,\xi = 1$, and let $\mathcal{T}$ be the corresponding unconditioned Galton–Watson tree. Then, as $n \to \infty$: For every fixed tree T,*

$$\frac{n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_{n,*} = T \mid \mathcal{T}_n) \xrightarrow{\mathrm{p}} \mathbb{P}(\mathcal{T} = T). \tag{1.9}$$

Equivalently, for any bounded functional f on $\mathfrak{T}$,

$$\frac{F(\mathcal{T}_n)}{n} = \mathbb{E}\, f\bigl(\mathcal{T}_{n,*} \mid \mathcal{T}_n\bigr) \xrightarrow{\mathrm{p}} \mathbb{E}\, f(\mathcal{T}). \tag{1.10}$$

Theorem 1.2 is a law of large numbers for $F(\mathcal{T}_n)$. In the present paper we take the next step and give a central limit theorem. This includes, as a special case, (joint) normal convergence of the subgraph counts $n_{T'}(T)$, see Corollary 1.4.

Theorem 1.3 *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of order n with offspring distribution ξ, where $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 := \operatorname{Var}\xi < \infty$, and let $\mathcal{T}$ be the corresponding unconditioned Galton–Watson tree. Suppose that $f : \mathfrak{T} \to \mathbb{R}$ is a functional of rooted trees such that $\mathbb{E}\,|f(\mathcal{T})| < \infty$, and let $\mu := \mathbb{E}\, f(\mathcal{T})$.*

(i) *If $\mathbb{E}\, f(\mathcal{T}_n) \to 0$ as $n \to \infty$, then*

$$\mathbb{E}\, F(\mathcal{T}_n) = n\mu + o\bigl(\sqrt{n}\bigr). \tag{1.11}$$

(ii) *If*

$$\mathbb{E}\, f(\mathcal{T}_n)^2 \to 0 \tag{1.12}$$

as $n \to \infty$, and

$$\sum_{n=1}^{\infty} \frac{\sqrt{\mathbb{E}(f(\mathcal{T}_n)^2)}}{n} < \infty, \tag{1.13}$$

then

$$\operatorname{Var} F(\mathcal{T}_n) = n\gamma^2 + o(n) \tag{1.14}$$

where

$$\gamma^2 := 2\,\mathbb{E}\Bigl(f(\mathcal{T})\bigl(F(\mathcal{T}) - |\mathcal{T}|\mu\bigr)\Bigr) - \operatorname{Var} f(\mathcal{T}) - \mu^2/\sigma^2 \tag{1.15}$$

is finite; moreover,

$$\frac{F(\mathcal{T}_n) - n\mu}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0, \gamma^2). \tag{1.16}$$

By (1.11), we may replace $n\mu$ by the exact mean $\mathbb{E}\, F(\mathcal{T}_n)$ in (1.16).

Special cases of Theorem 1.3 have been proved before, by various methods. A simple example is the number of leaves in $\mathcal{T}_n$, shown to be normal by Kolchin [19], see Example 2.1. (See also Aldous [1, Remark 7.5.3].) Wagner [26] considered random labelled trees (the case $\xi \sim \operatorname{Po}(1)$) and showed Theorem 1.3 (and convergence of all moments) for this case, assuming further that f is bounded and $\mathbb{E}\,|f(\mathcal{T}_n)| = O(c^n)$ for some $c < 1$ (a stronger assumption that our (1.12)–(1.13)).

Theorem 1.3 is stated for a single functional F, but joint convergence for several different F (each satisfying the conditions in the theorem) follows immediately by the Cramér–Wold device. One example is the following corollary for the subtree counts (1.1). (We state the result as asymptotic normality for the infinite family of all subtree counts; by definition, this is the same as asymptotic normality for any finite subfamily.)

Corollary 1.4 *The subtree counts $n_T(\mathcal{T}_n)$, $T \in \mathfrak{T}$, are asymptotically jointly normal. More precisely, let $\pi_T := \mathbb{P}(\mathcal{T} = T)$,*

$$\gamma_{T,T} := \pi_T - \big(2|T| - 1 + \sigma^{-2}\big)\pi_T^2, \tag{1.17}$$

and, for $T_1 \neq T_2$,

$$\gamma_{T_1,T_2} := n_{T_2}(T_1)\pi_{T_1} + n_{T_1}(T_2)\pi_{T_2} - \big(|T_1| + |T_2| - 1 + \sigma^{-2}\big)\pi_{T_1}\pi_{T_2}. \tag{1.18}$$

Then, for any trees $T, T_1, T_2 \in \mathfrak{T}$,

$$\mathbb{E}\, n_T(\mathcal{T}_n) = n\pi_T + o\big(\sqrt{n}\big), \tag{1.19}$$

$$\operatorname{Cov}\big(n_{T_1}(\mathcal{T}_n), n_{T_2}(\mathcal{T}_n)\big) = n\gamma_{T_1,T_2} + o(n), \tag{1.20}$$

$$\frac{n_T(\mathcal{T}_n) - n\pi_T}{\sqrt{n}} \xrightarrow{\mathrm{d}} Z_T, \tag{1.21}$$

the latter jointly for all $T \in \mathfrak{T}$, where Z_T are jointly normal with mean $\mathbb{E}\, Z_T = 0$ and covariances $\operatorname{Cov}\big(Z_{T_1}, Z_{T_2}\big) = \gamma_{T_1,T_2}$.

We say that the functional f has *finite support* if $f(T) \neq 0$ only for finitely many trees $T \in \mathfrak{T}$; equivalently, there exists a constant K such that $f(T) = 0$ unless $|T| \leqslant K$. Note that a functional with finite support necessarily is bounded. By (1.5), the additive functionals F that arise from functionals f with finite support are exactly the finite linear combinations of subgraph counts $n_{T'}(T)$. Hence Corollary 1.4 is equivalent to asymptotic normality (with convergence of mean and variance) for $F(\mathcal{T}_n)$ whenever f has finite support. The asymptotic variance $\gamma^2 = \lim_{n\to\infty} \operatorname{Var} F(\mathcal{T}_n)/n$ is given by (1.15) or, equivalently, follows from (1.17)–(1.18).

Remark 1.5 The condition (1.12) in Theorem 1.3(ii) is equivalent to $\mathbb{E}\, f(\mathcal{T}_n) \to 0$ and $\operatorname{Var} f(\mathcal{T}_n) \to 0$, and it implies $\mathbb{E}\,|f(\mathcal{T}_n)| \to 0$ as assumed in (i). Both this condition and (1.13) say that $f(T)$ is (on the average, at least) decreasing as $|T| \to \infty$, but a rather slow decrease is sufficient; for example, the theorem applies when $f(T) = 1/\log^2|T|$ (for $|T| > 1$). In particular, it is *not* enough to assume that f is a bounded functional. For a trivial example, let $f(T) = 1$ for all trees T; then $F(T) = |T|$ so $F(\mathcal{T}_n) = n$ is constant, with mean n and variance 0. However, the first two terms on the right-hand side of (1.15) vanish, so $\gamma^2 = -\sigma^{-2} < 0$, which is absurd for an asymptotic variance, and (1.14) and (1.16) fail.

Remark 1.6 If we go further and allow $f(T)$ that grow with the size $|T|$, we cannot expect the results to hold. Fill and Kapur [8] have made an interesting and illustrative study (for certain f) of the case of binary trees, which is the case $\xi \sim \operatorname{Bin}(2, 1/2)$ of conditioned Galton–Watson tree, and presumably typical for other conditioned Galton–Watson trees as well. They show that for $f(T) = \log|T|$, $F(\mathcal{T}_n)$ is asymptotically normal, but with a variance of the order $n \log n$. And if $f(T)$ increases more rapidly, with $f(T) = |T|^\alpha$ for some $\alpha > 0$, then the variance is of order $n^{1+2\alpha}$, and $F(\mathcal{T}_n)$ has, after normalization, a non-normal limiting distribution.

Intuitively, our conditions are such that the sum (1.3) is dominated by the many small subtrees T_v; since different parts of our trees are only weakly dependent on each other, this makes asymptotic normality plausible. For a toll function f that grows too rapidly with the size of T, the sum (1.3) will on the contrary be dominated by large subtrees, which are more strongly dependent, and then other limit distributions will appear.

Remark 1.7 For the m-ary search tree ($2 \leqslant m \leqslant 26$) and random recursive tree a similar theorem holds, but there $f(T)$ may grow almost as $|T|^{1/2}$, see Hwang and Neininger [13] (binary search tree, f depends on $|T|$ only), Fill and Kapur [9] (m-ary search tree, f depends on $|T|$ only), Holmgren and Janson [11] (binary search tree and random recursive tree, general f). A reason for this difference is that for a conditioned Galton–Watson tree, the limit distribution of the size of the fringe subtree, which by Theorem 1.2 is the distribution of $|\mathcal{T}|$, decays rather slowly, with $\mathbb{P}(|\mathcal{T}| = n) \asymp n^{-3/2}$. while the corresponding limit distribution for fringe subtrees in a binary search tree or random recursive tree decays somewhat faster, as n^{-2}, see Aldous [1]. Cf. also the related results in Fill, Flajolet and Kapur [7, Theorem 13 and 14], showing a similar contrast (but at orders $n^{1/2}$ and n) between uniform binary trees (an example of a conditioned Galton–Watson tree) and binary search trees for the asymptotic expectation of an additive functional.

We can weaken the conditions on the size of f if we assume that f is "nice". We say that a functional $f(T)$ on $\mathfrak{T}$ is *local* (with *cut-off* M) if it depends only on the first M generations of T, for some $M < \infty$, i.e., if we let $T^{(M)}$ denote T truncated at height M, then $f(T) = f(T^{(M)})$. More generally, we say that f is *weakly local* (with *cut-off* M) if $f(T)$ depends on $|T|$ and $T^{(M)}$ for some M.

Theorem 1.8 *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree as in Theorem 1.3. Suppose that $f : \mathfrak{T} \to \mathbb{R}$ is a bounded and local functional. Then the conclusions* (1.11), (1.14) *and* (1.16) *hold for some $\gamma^2 < \infty$.*

More generally, the same holds if f is a bounded and weakly local functional such that $\mathbb{E}\, f(\mathcal{T}_n) \to 0$ and $\sum_n |\mathbb{E}\, f(\mathcal{T}_n)|/n < \infty$.

Remark 1.9 The asymptotic variance γ^2 equals 0 in two trivial cases:

(i) $f(\mathcal{T}) = F(\mathcal{T}) = F(\mathcal{T}_n) = 0$ a.s.;

(ii) $\{k : p_k > 0\} = \{0, r\}$ for some $r > 1$ and $f(\mathcal{T}) = a\mathbf{1}\{|\mathcal{T}| = 1\}$ for some real a; then $F(\mathcal{T}_n) = a(n - (n-1)/r)$ is deterministic.

We can show, using [17], that if f has finite support, then $\gamma^2 > 0$ except in these trivial cases. For general f, we do not know whether $\gamma^2 = 0$ is possible except in such trivial cases. (See Example 2.2 for another trivial case. For some f, it may be possible to use the simple criterion in [6] to show $\gamma^2 > 0$, but in general, our f is not of the type studied there so more research is needed.)

This is an extended abstract of [16], where proofs and further details are given.

2 Examples

Example 2.1 The perhaps simplest non-trivial example is to take $f(T) = \mathbf{1}\{|T| = 1\}$. Then $F(T)$ is the number of leaves in T. We have $\mathbb{E}\, f(\mathcal{T}) = \mathbb{P}(|\mathcal{T}| = 1) = \mathbb{P}(\xi = 0) = p_0$.

Theorems 1.3 and 1.8 both apply and show asymptotic normality of $F(\mathcal{T}_n)$, and so does Corollary 1.4 since $F(T) = n_\bullet(T)$, where $\bullet$ is the tree of order 1; (1.15) yields

$$\gamma^2 = 2p_0(1-p_0) - p_0(1-p_0) - p_0^2/\sigma^2 = p_0 - (1+\sigma^{-2})p_0^2, \tag{2.1}$$

which also is seen directly from (1.17). The asymptotic normality in this case (and a local limit theorem) was proved by Kolchin [19, Theorem 2.3.1]. By Remark 1.9, or by a simple calculation directly from (2.1), $\gamma^2 > 0$ except in the case $p_r = 1 - p_0 = 1/r$ for some $r \geqslant 2$ when all nodes in $\mathcal{T}_n$ have 0 or r children (full r-ary trees) and $n_\bullet(\mathcal{T}_n) = n - (n-1)/r$ is deterministic.

Example 2.2 A natural extension is to consider the number of nodes of outdegree r, for some given integer $r \geqslant 1$; we denote this by $n_r(T)$. Then $n_r(T) = F(T)$ with $f(T) = 1$ if the root of T has degree r, and $f(T) = 0$ otherwise. Asymptotic normality of $n_r(\mathcal{T}_n)$ too was proved by Kolchin [19, Theorem 2.3.1], with

$$n^{-1/2}\bigl(F(\mathcal{T}_n) - np_r\bigr) \xrightarrow{\mathrm{d}} N\bigl(0, \gamma_r^2\bigr) \tag{2.2}$$

where

$$\gamma_r^2 = p_r(1-p_r) - (r-1)^2 p_r^2/\sigma^2, \tag{2.3}$$

see also Janson [14] (joint convergence and moment convergence, assuming at least $\mathbb{E}\,\xi^3 < \infty$), Minami [23] and Drmota [5, Section 3.2.1] (both assuming an exponential moment) for different proofs.

It is easily checked that for $r > 0$, $\gamma_r > 0$ except in the two trivial cases $p_r = 0$, when $n_r(\mathcal{T}_n) = 0$, and $p_r = 1 - p_0 = 1/r$, when all nodes have 0 or r children (full r-ary trees) and $n_r(\mathcal{T}_n) = (n-1)/r$ is deterministic.

In this example,

$$\mathbb{E}\, f(\mathcal{T}_n) = \mathbb{P}(\text{the root of } \mathcal{T}_n \text{ has degree } r) \to rp_r. \tag{2.4}$$

see [18] and [15, Theorem 7.10]. Hence (1.12) and (1.13) both fail, and we cannot apply Theorem 1.3. (It does not help to subtract a constant, since $f(\mathcal{T}_n)$ is an indicator variable.) However, f is a bounded local functional. Hence Theorem 1.8 applies and yields (2.2), together with convergence of mean and variance, for some γ_r. It is immediate from the definition of the Galton–Watson tree $\mathcal{T}$ that

$$\mu := \mathbb{E}\, f(\mathcal{T}) = \mathbb{P}(\text{the root of } \mathcal{T} \text{ has degree } r) = p_r. \tag{2.5}$$

Similarly, we obtain joint convergence for different r by Theorem 1.8 and the Cramér–Wold device. (It seems that joint convergence has not been proved before without assuming at least $\mathbb{E}\,\xi^3 < \infty$.)

Nevertheless, this result is a bit disappointing, since we do not obtain the explicit formula (2.3) for the variance. Theorem 1.8 shows existence of γ^2 but the formula (3.19) given by the proof is rather involved, and we do not know any way to derive (2.3) from it. In this example, because of the simple structure of f, we can use a special argument and derive both (2.3) and the asymptotic covariance γ_{rs} for two different outdegrees $r, s \geqslant 0$:

$$\gamma_{rs} = -p_r p_s - (r-1)(s-1)p_r p_s/\sigma^2, \qquad r \neq s, \tag{2.6}$$

(as proved by [14] provided $\mathbb{E}\,\xi^3 < \infty$).

Note that by (2.2), $\liminf_{n\to\infty} n^{-1/2}\,\mathbb{E}\,|F(\mathcal{T}_n) - n\mu| \geqslant (2/\pi)^{1/2}\gamma_r$, so assuming $\gamma_r > 0$, $\mathbb{E}\,|F(\mathcal{T}_n) - n\mu| \geqslant c_1 n^{1/2}$, at least for large n. It is easily seen that also $\mathbb{E}\, f(\mathcal{T}_n)|F(\mathcal{T}_n) - n\mu| \geqslant c_2 n^{1/2}$, at least for large n; hence, using $\mathbb{P}(|\mathcal{T}| = n) \sim cn^{-3/2}$,

$$\mathbb{E}\bigl|f(\mathcal{T})\bigl(F(\mathcal{T}) - |T|\mu\bigr)\bigr| = \sum_{n=1}^{\infty} \mathbb{P}(|\mathcal{T}| = n)\,\mathbb{E}\bigl|f(\mathcal{T}_n)\bigl(F(\mathcal{T}_n) - n\mu\bigr)\bigr| = \infty,$$

which shows that the expectation in (1.15) does not exist, so γ^2 is not given by (1.15).

Example 2.3 A node in a (rooted) tree is said to be *protected* if it is neither a leaf nor the parent of a leaf. Asymptotics for the expected number of protected nodes in various random trees, including several examples of conditioned Galton–Watson trees, have been given by e.g. Cheon and Shapiro [3] and Mansour

[22], and convergence in probability of the fraction of protected nodes is proved for general conditioned Galton–Watson trees by Devroye and Janson [4].

We can now extend this to asymptotic normality of the number of protected nodes, in any conditioned Galton–Watson tree $\mathcal{T}_n$ with $\mathbb{E}\,\xi = 1$ and $\sigma^2 < \infty$. We define $f(T) := \mathbf{1}\{\text{the root of } T \text{ is protected}\}$, and then $F(T)$ is the number of protected nodes in T. Since f is a bounded and local functional, Theorem 1.8 applies and shows asymptotic normality of $F(\mathcal{T}_n)$.

The asymptotic mean $\mu = \mathbb{E}\, f(\mathcal{T})$ is easily calculated, see [4] where also explicit values are given for several examples of conditioned Galton–Watson trees. However, as in Example 2.2, we do not see how to find an explicit value of γ^2 from (3.19) (although it ought to be possible to use these for numerical calculation for a specific offspring distribution). It seems possible that there is some other argument to find γ^2, but we have not pursued this and we leave it as an open problem to find the asymptotic variance γ^2, for example for uniform labelled trees or uniform binary trees.

Example 2.4 Wagner [26] studied the number $s(T)$ of arbitrary subtrees (not necessarily fringe subtrees) of the tree T, and the number $s_1(T)$ of such subtrees that contain the root. He noted that if T has branches $T_1, \dots, T_d$, then $s_1(T) = \prod_{i=1}^{d}(1 + s_1(T_i))$ and thus

$$\log\bigl(1 + s_1(T)\bigr) = \log\bigl(1 + s_1(T)^{-1}\bigr) + \sum_{i=1}^{d} \log\bigl(1 + s_1(T_i)\bigr), \tag{2.7}$$

so $\log\bigl(1 + s_1(T)\bigr)$ is an additive functional with toll function $f(T) = \log\bigl(1 + s_1(T)^{-1}\bigr)$, see (1.6). Wagner [26] used this and the special case of Theorem 1.3 shown by him to show asymptotic normality of $\log\bigl(1 + s_1(\mathcal{T}_n)\bigr)$ (and thus of $\log s_1(\mathcal{T}_n)$) for the case of uniform random labelled trees (which is $\mathcal{T}_n$ with $\xi \sim \mathrm{Po}(1)$). We can generalize this to arbitrary conditioned Galton–Watson trees with $\mathbb{E}\,\xi = 1$ and $\mathbb{E}\,\xi^2 < \infty$ by Theorem 1.3, noting that $|f(\mathcal{T}_n)| \leqslant s_1(\mathcal{T}_n)^{-1} \leqslant n^{-1}$ (since $s_1(T) \geqslant |T|$ by considering only paths from the root); hence (1.12)–(1.13) hold. Consequently,

$$\bigl(\log s_1(\mathcal{T}_n) - n\mu\bigr)/\sqrt{n} \xrightarrow{\mathrm{d}} N(0, \gamma^2) \tag{2.8}$$

for some $\mu = \mathbb{E} \log\bigl(1 + s_1(\mathcal{T})^{-1}\bigr)$ and γ^2 given by (1.15) (both depending on the distribution of ξ); Wagner [26] makes a numerical calculation of μ and σ^2 for his case.

Furthermore, as noted in [26], $s_1(T) \leqslant s(T) \leqslant |T| s_1(T)$ for any tree (an arbitrary subtree is a fringe subtree of some subtree containing the root), and thus the asymptotic normality (2.8) holds for $\log s(\mathcal{T}_n)$ too.

Similarly, the example by Wagner [26, pp. 78–79] on the average size of a subtree containg the root generalizes to arbitrary conditioned Galton–Watson trees (with $\mathbb{E}\,\xi^2 < \infty$), showing that the average size is asymptotically normal with expectation $\sim \mu n$ and variance $\sim \gamma^2 n$ for some $\mu > 0$ and γ^2; we omit the details. We conjecture that the same is true for the average size of an arbitrary subtree, as shown in [26] for the case considered there. (Note that a uniformly random arbitrary subtree thus is much larger than a uniformly random fringe subtree.)

3 Sketch of proofs

We let $\xi_1, \xi_2, \dots$ be a sequence of independent copies of ξ, and let

$$S_n := \sum_{i=1}^{n} \xi_i. \tag{3.1}$$

3.1 A useful representation

A tree in $\mathfrak{T}$ is uniquely described by its degree sequence $(d_1, \dots, d_n)$. We may thus define the functional f also on finite nonnegative integer sequences $(d_1, \dots, d_n)$, $n \geqslant 1$, by

$$f(d_1, \dots, d_n) := \begin{cases} f(T), & (d_1, \dots, d_n) \text{ is the degree sequence of a tree } T, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

If T has degree sequence $(d_1, \dots, d_n)$, and its nodes are numbered $v_1, \dots, v_n$ in depth-first order so d_i is the degree of v_i, then the subtree T_{v_i} has degree sequence $(d_i, d_{i+1}, \dots, d_{i+k-1})$, where $k \leqslant n - i + 1$ is the unique index such that $(d_i, \dots, d_{i+k-1})$ is a degree sequence of a tree. By the definition (3.2), we thus can write (1.3) as

$$F(T) = \sum_{1 \leqslant i \leqslant j \leqslant n} f(d_i, \dots, d_j) = \sum_{k=1}^{n} \sum_{i=1}^{n-k+1} f(d_i, \dots, d_{i+k-1}). \tag{3.3}$$

Moreover, if we regard $(d_1, \dots, d_n)$ as a cyclic sequence and allow wrapping around by defining $d_{n+i} := d_i$, we also have the more symmetric formula

$$F(T) = \sum_{k=1}^{n} \sum_{i=1}^{n} f(d_i, \dots, d_{i+k-1}). \tag{3.4}$$

The difference from (3.3) is that we have added some terms $f(d_i, \dots, d_{i+k-1-n})$ where the indices wrap around, but these terms all vanish by definition because $(d_i, \dots, d_{i+k-1-n})$ is never a degree sequence. (The subtree with root v_i is completed at the latest by v_n.)

It is a well-known fact, see e.g. [15, Corollary 15.4], that up to a cyclic shift, the degree sequence $(d_1, \dots, d_n)$ of the conditioned Galton–Watson tree $\mathcal{T}_n$ has the same distribution as $\bigl((\xi_1, \dots, \xi_n) \mid S_n = n-1\bigr)$. Since (3.4) is invariant under cyclic shifts of $(d_1, \dots, d_n)$, it follows that, recalling (3.1),

$$F(\mathcal{T}_n) \overset{\mathrm{d}}{=} \left(\sum_{k=1}^{n} \sum_{i=1}^{n} f(\xi_i, \dots, \xi_{i+k-1 \bmod n}) \;\middle|\; S_n = n-1 \right), \tag{3.5}$$

where $j \bmod n$ denotes the index in $\{1, \dots, n\}$ that is congruent to j modulo n.

3.2 More notation

We let, for $k \geqslant 1$, f_k be f restricted to $\mathfrak{T}_k$; more precisely, we define f_k for all trees $T \in \mathfrak{T}$ by $f_k(T) := f(T)$ if $|T| = k$ and $f_k(T) := 0$ otherwise. In other words,

$$f_k(T) := f(T) \cdot \mathbf{1}\{|T| = k\}. \tag{3.6}$$

Extended to integer sequences as in (3.2), this means that

$$f_k(d_1, \dots, d_n) = f(d_1, \dots, d_n) \cdot \mathbf{1}\{n = k\}. \tag{3.7}$$

Note that $\mathfrak{T}_k$ is a finite set; thus f_k is always a bounded function for each k.

We further let, for $k \geqslant 1$ and any tree T, with degree sequence $(d_1, \dots, d_n)$,

$$F_k(T) := F(T; f_k) = \sum_{i=1}^{n-k+1} f_k(d_i, \dots, d_{i+k-1}). \tag{3.8}$$

(We can also let the sum extend to n, wrapping around d_i as in (3.4).) Obviously,

$$f(T) = \sum_{k=1}^{\infty} f_k(T) \qquad \text{and} \qquad F(T) = \sum_{k=1}^{\infty} F_k(T) \tag{3.9}$$

for any tree T, where in both sums it suffices to consider $k \leqslant |T|$ since the summands vanish for $k > |T|$.

3.3 Expectations

We calculate the expectation $\mathbb{E}\, F(\mathcal{T}_n)$ using (3.5), which converts this into a problem on expectations of functionals of a sequence of i.i.d. variables conditioned on their sum. Results of this type have been studied before under various conditions, see for example Zabell [27, 28, 29], Swensen [25] and Janson [14].

By (3.5) and symmetry,

$$\mathbb{E}\, F(\mathcal{T}_n) = n \sum_{k=1}^{n} \mathbb{E}\big(f(\xi_1, \dots, \xi_k) \mid S_n = n-1\big). \tag{3.10}$$

We consider first the expectation of each $F_k(\mathcal{T}_n)$ separately, recalling (3.9). Note that each f_k is bounded, and thus trivially $\mathbb{E}\,|f_k(\mathcal{T})| < \infty$. A simple argument yields the following.

Lemma 3.1 *If $1 \leqslant k \leqslant n$, then*

$$\mathbb{E}\, F_k(\mathcal{T}_n) = n \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)}\, \mathbb{E}\, f_k(\mathcal{T}). \tag{3.11}$$

We use this in combination with the following estimates, which are shown using the local limit theorem and the methods used to prove it. (Cf. [24, Theorem VII.13].)

Lemma 3.2 (i) *Uniformly for all k with $1 \leqslant k \leqslant n/2$, as $n \to \infty$,*

$$\frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} = 1 + O\left(\frac{k}{n}\right) + o\big(n^{-1/2}\big). \tag{3.12}$$

(ii) *If $n/2 < k \leqslant n$, then*

$$\frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} = O\left(\frac{n^{1/2}}{(n-k+1)^{1/2}}\right). \tag{3.13}$$

3.4 Variances and covariances

We next consider the variance of $F(\mathcal{T}_n)$. As in Section 3.3, we consider first the different $F_k(\mathcal{T}_n)$ separately; thus we study variances and covariances of these sums. We begin with an exact formula (omitted here), corresponding to Lemma 3.1; then the terms in it are estimated similarly to Lemma 3.2, but with more care since there typically is important cancellation between different terms. In particular, this leads to a simple asymptotic result for fixed k and m.

Lemma 3.3 *For any fixed k and m with $k \geqslant m$, as $n \to \infty$,*

$$\frac{1}{n}\operatorname{Cov}(F_k(\mathcal{T}_n), F_m(\mathcal{T}_n)) \to \mathbb{E}\bigl(f_k(\mathcal{T})F_m(\mathcal{T})\bigr) - (k+m-1+\sigma^{-2})\,\mathbb{E}\,f_k(\mathcal{T})\,\mathbb{E}\,f_m(\mathcal{T}).$$

Corollary 3.4 *Suppose that f has finite support. Then, as $n \to \infty$,*

$$\frac{1}{n}\operatorname{Var} F(\mathcal{T}_n) \to \mathbb{E}\Bigl(f(\mathcal{T})\bigl(2F(\mathcal{T}) - f(\mathcal{T})\bigr)\Bigr) - 2\,\mathbb{E}\bigl(|\mathcal{T}|f(\mathcal{T})\bigr)\,\mathbb{E}\,f(\mathcal{T}) + \bigl(1-\sigma^{-2}\bigr)\bigl(\mathbb{E}\,f(\mathcal{T})\bigr)^2.$$

For general f, we first show a uniform bound valid for all n.

Theorem 3.5 *For any functional $f : \mathfrak{T} \to \mathbb{R}$,*

$$\operatorname{Var}\bigl(F(\mathcal{T}_n)\bigr)^{1/2} \leqslant C_1 n^{1/2}\Biggl(\sup_k \sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2} + \sum_{k=1}^{\infty}\frac{\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2}}{k}\Biggr), \tag{3.14}$$

with C_1 independent of f.

Using this bound, (1.14)–(1.15) follow easily.

3.5 Asymptotic normality

To prove asymptotic normality, we first consider functionals f with finite support. We use the representation (3.5), where now it suffices to sum over $k \leqslant m$ for some $m < \infty$. We define

$$g(x_1,\dots,x_m) := \sum_{k=1}^{m} f(x_1,\dots,x_k) = \sum_{k=1}^{m} f_k(x_1,\dots,x_k). \tag{3.15}$$

Then (3.5) can be written (assuming $n \geqslant m$)

$$F(\mathcal{T}_n) \overset{\mathrm{d}}{=} \left(\sum_{i=1}^{n} g(\xi_i,\dots,\xi_{i+m-1 \bmod n}) \;\middle|\; S_n = n-1\right). \tag{3.16}$$

Asymptotic normality now follows by a method by Le Cam [21] and Holst [12], see also Kudlaev [20].

For general f in Theorems 1.3 and 1.8 we define the truncation

$$f^{(N)}(T) := \sum_{k=1}^{N} f_k(T) = f(T)\mathbf{1}\{|T| \leqslant N\} \tag{3.17}$$

and the corresponding sum $F^{(N)}(T)$. The result follow from the case of finite support, using Theorem 3.5 and a similar but sharper estimate for weakly local functionals.

More precisely, with $\mu^{(N)} := \mathbb{E}\, f^{(N)}(\mathcal{T})$ and

$$(\gamma^{(N)})^2 := 2\,\mathbb{E}\Big(f^{(N)}(\mathcal{T})\big(F^{(N)}(\mathcal{T}) - |\mathcal{T}|\mu^{(N)}\big)\Big) - \operatorname{Var} f^{(N)}(\mathcal{T}) - (\mu^{(N)})^2/\sigma^2, \tag{3.18}$$

this proof yields asymptotic normality with the asymptotic variance

$$\gamma^2 := \lim_{N\to\infty} (\gamma^{(N)})^2. \tag{3.19}$$

In Theorem 1.3, this leads to (1.15). In Theorem 1.8, we do not know any simple general formula for γ^2.

References

[1] David Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266.

[2] Jürgen Bennies & Götz Kersting, A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** (2000), no. 3, 777–803.

[3] Gi-Sang Cheon & Louis W. Shapiro, Protected points in ordered trees. *Applied Mathematics Letters* **21** (2008), no. 5, 516–520.

[4] Luc Devroye & Svante Janson, Protected nodes and fringe subtrees in some random trees. `arXiv:1310.0665`

[5] Michael Drmota, *Random Trees*, Springer, Vienna, 2009.

[6] Michael Drmota, Éric Fusy, Mihyun Kang, Veronika Kraus & Juanjo Rué, Asymptotic study of subcritical graph classes. *SIAM J. Discrete Math.* **25** (2011), no. 4, 1615–1651.

[7] James Allen Fill, Philippe Flajolet & Nevin Kapur, Singularity analysis, Hadamard products, and tree recurrences. *J. Comput. Appl. Math.* **174** (2005), no. 2, 271–313.

[8] James Allen Fill & Nevin Kapur, Limiting distributions for additive functionals on Catalan trees. *Theoret. Comput. Sci.* **326** (2004), no. 1-3, 69–102.

[9] James Allen Fill & Nevin Kapur, Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Structures Algorithms* **26** (2005), no. 4, 359–391.

[10] James Allen Fill & Nevin Kapur, A repertoire for additive functionals of uniformly distributed m-ary search trees (extended abstract). *Discrete Math. Theor. Comput. Sci. Proc.* **AD** (2005), 105–114.

[11] Cecilia Holmgren & Svante Janson, Limit laws for functions of fringe trees for binary search trees and recursive trees. In preparation.

[12] Lars Holst, Some conditional limit theorems in exponential families. *Ann. Probab.* **9** (1981), no. 5, 818–830.

[13] Hsien-Kuei Hwang & Ralph Neininger, Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31** (2002), no. 6, 1687–1722.

[14] Svante Janson, Moment convergence in conditional limit theorems. *J. Appl. Probab.* **38** (2001), no. 2, 421–437.

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

[16] Svante Janson, Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. Preprint, 2013. `arXiv:1312.1211`

[17] Svante Janson, On degenerate sums of m-dependent variables. Preprint, 2013.

[18] Douglas P. Kennedy, The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12** (1975), 800–806.

[19] Valentin F. Kolchin, *Random Mappings*. Nauka, Moscow, 1984. (Russian.) English transl.: Optimization Software, New York, 1986.

[20] E. M. Kudlaev, Decomposable statistics. (Russian.) *Itogi Nauki i Tekhniki* **26** (1988), 97–149. English transl.: *J. Soviet Math.* **50** (1990), no. 3, 1601–1642.

[21] Lucien Le Cam, Un théorème sur la division d'un intervalle par des points pris au hasard. *Publ. Inst. Statist. Univ. Paris* **7** (1958), no. 3/4, 7–16.

[22] Toufik Mansour, Protected points in k-ary trees. *Applied Mathematics Letters* **24** (2011), no. 4, 478–480.

[23] Nariyuki Minami, On the number of vertices with a given degree in a Galton–Watson tree. *Adv. Appl. Probab.* **37** (2005), no. 1, 229–264.

[24] Valentin V. Petrov, *Sums of Independent Random Variables*. Springer-Verlag, Berlin, 1975.

[25] Anders Rygh Swensen, A note on convergence of distributions of conditional moments. *Scand. J. Statist.* **10** (1983), no. 1, 41–44.

[26] Stephan Wagner, Additive tree functionals with small toll functions and subtrees of random trees. *Discrete Math. Theor. Comput. Sci. Proc.*, **AQ** (2012), 67–80.

[27] Sandy Lew Zabell, *A limit theorem for conditional expectations with applications to probability theory and statistical mechanics*. Ph.D. thesis, Harvard University, 1975.

[28] Sandy L. Zabell, Rates of convergence for conditional expectations. *Ann. Probab.* **8** (1980), no. 5, 928–941.

[29] S. L. Zabell, A limit theorem for expectations conditional on a sum. *J. Theoret. Probab.* **6** (1993), no. 2, 267–283.