On general subtrees of a conditioned Galton–Watson tree

Svante Janson

Abstract

We show that the number of copies of a given rooted tree in a conditioned Galton–Watson tree satisfies a law of large numbers under a minimal moment condition on the offspring distribution.

Keywords: random trees; conditioned Galton-Watson tree; subtrees.

MSC2020 subject classifications: 60C05; 05C05.

1 Introduction

Let $T_n$ be a random conditioned Galton–Watson tree with $n$ nodes, defined by an offspring distribution $\xi$ with mean $E\xi = 1$, and let $t$ be a fixed ordered rooted tree. We are interested in the number of copies of $t$ as a (general) subtree of $T_n$, which we denote by $N_t(T_n)$. For details of these and other definitions, see Section 2. Note that we consider subtrees in a general sense. (Thus, e.g., not just fringe trees; for them, see similar results in [9, Theorem 7.12] and [10].)

The purpose of the present paper is to show the following law of large numbers under minimal moment assumptions. Let $n_t(T)$ be the number of rooted copies of $t$ in a tree $T$, i.e., copies with the root at the root of $T$. Further, let $\Delta(t)$ be the maximum outdegree in $t$.

**Theorem 1.1.** Let $t$ be a fixed ordered tree, and let $T_n$ be a conditioned Galton–Watson tree defined by an offspring distribution $\xi$ with $E\xi = 1$ and $E\xi^{\Delta(t)} < \infty$. Also, let $T$ be a Galton–Watson tree with the same offspring distribution. Then, as $n \to \infty$,

$$N_t(T_n)/n \overset{L}{\to} E n_t(T),$$

where the limit is finite and given explicitly by (3.2) below.

Equivalently,

$$N_t(T_n)/n \overset{P}{\to} E n_t(T),$$

and

$$E N_t(T_n)/n \to E n_t(T).$$

*Supported by the Knut and Alice Wallenberg Foundation.
†Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden.
E-mail: svante.janson@math.uu.se
On general subtrees of a conditioned Galton-Watson tree

**Remark 1.2.** $N_t(T_n)/n$ can be regarded as the average of the number of rooted copies of $t$ in a randomly chosen fringe tree of $T_n$, see (2.1). In other words, denoting the random fringe tree by $T_n^*$, we have $N_t(T_n)/n = E(n_t(T_n^*) | T_n)$ and thus $E N_t(T_n)/n = E n_t(T_n^*)$.

Aldous [1] showed (when $E x^2 < \infty$) that $T_n^*$ converges in distribution to $T$; moreover, this holds also conditioned on $T_n$, see [9, Theorem 7.12] for a precise formulation. Hence, (1.3) and (1.2) can be interpreted as convergence of an expectation in these results on convergence in distribution, which makes them very plausible. Nevertheless, since the variable $n_t(T_n)$ is unbounded, a rigorous proof along this path would require further estimates. (Our proof below is related but somewhat different.)

The fact that (1.1) is equivalent to (1.2)–(1.3) is an instance of the general fact that for any random variables, convergence in $L^1$ is equivalent to convergence in probability together with convergence of the means of the absolute values (i.e., in this case, with non-negative variables, the means); see e.g. [7, Theorem 5.5.4]. We nevertheless state both versions for convenience.

Chyzak, Drmota, Klausner and Kok [2] (see also [3, Section 3.3]) considered patterns in random trees; their patterns differ from the subgraph counts above in that some external vertices are added to $t$, and that one only considers copies of $t$ in a tree $T$ such that each internal vertex in the copy has the same degree in $T$ as in $t$ (counting also edges to external vertices); equivalently, each vertex in $T$ is equipped with a number, and one considers only copies of $t$ where the vertex degrees match these numbers. (Another difference is that [2] consider unrooted trees, but the proof proceeds by first considering rooted (planted) trees. Furthermore, only uniformly random labelled trees are considered in [2], but the proofs extend to suitable more general conditioned Galton-Watson trees, as remarked in [2] and shown explicitly in [12; 13].) It was shown in Chyzak, Drmota, Klausner and Kok [2] that the number of occurrences of such a pattern is asymptotically normal, with asymptotic mean and variance both of the order $n$ (except that the variance might be smaller in at least one exceptional degenerate case), which of course entails a law of large numbers. Moreover, [2] discuss briefly generalizations, including subtrees without further degree conditions as in the present paper; they expect asymptotic normality to hold in this case too, but it seems that their method, which is based on setting up and analyzing a system of functional equations for generating functions, in general would require extensions to infinite systems, which as far as we know has not been pursued. (See [5] for a related problem.) See further Section 5.

Our method is probabilistic, and quite different from the analysis of generating functions in [2].

## 2 Notation

All trees are rooted and ordered. The root of a tree $T$ is denoted $o = o_T$. The size $|T|$ of a tree $T$ is defined as the number of vertices in $T$.

The degree $d(v)$ of a vertex $v \in T$ always means the outdegree, i.e., the number of children of $v$. The degree sequence of $T$ is the sequence of all degrees $d(v)$, $v \in T$, for definiteness in depth first order. Let $\Delta(T) := \max_{v \in T} d(v)$ be the maximum (out)degree in $T$.

A (general) subtree $T'$ of a tree $T$ is a non-empty connected subgraph of $T$; we regard a subtree as a rooted tree in the obvious way, with the root being the vertex in $T'$ that is closest to the root in $T$. Note that for any vertex $v \in T'$, its set of children in $T'$ is a subset of its set of children in $T$; the order of the children of $v$ in $T'$ is (by definition) the same as their relative order in $T$.

If $v \in T$, the fringe subtree $T^v$ is the subtree of $T$ consisting of $v$ and all its descendants; this is thus a subtree with root $v$.
On general subtrees of a conditioned Galton–Watson tree

If \( t \) and \( T \) are ordered rooted tree, let \( N_t(T) \) be the number of (general) subtrees of \( T \) that are isomorphic to \( t \) (as ordered trees), and let \( n_t(T) \) be the number of such subtrees that furthermore have root \( o_T \). Then \( n_t(T^\circ) \) is the number of subtrees with root \( v \) isomorphic to \( t \), and thus

\[
N_t(T) = \sum_{v \in T} n_t(T^\circ).
\]

In other words, \( N_t(T) \) is an additive functional with toll function \( n_t(T) \), see e.g. [10].

Let \( T \) be a random Galton–Watson tree defined by an offspring distribution \( (p_i)_0^\infty \), and let \( T_n \) be the conditioned Galton–Watson tree defined as \( T \) conditioned on \( |T| = n \) (tacitly considering only \( n \) such that \( \mathbb{P}(|T| = n) > 0 \)); see e.g. [9] for a survey. We let \( \xi \) be a random variable with the distribution \( (p_i)_0^\infty \); we call both \( (p_i)_0^\infty \) and (with a minor abuse) \( \xi \) the offspring distribution. We will only consider offspring distributions with \( \mathbb{E} \xi = 1 \) (i.e., \( \xi \) is critical). (We often repeat this for emphasis.) Let \( \sigma^2 := \operatorname{Var} \xi \leq \infty \); we tacitly assume \( \sigma^2 > 0 \), but do not require \( \sigma^2 < \infty \) unless we say so.

\( C \) and \( c \) denote unspecified constants that may vary from one occurrence to the next. They may depend on parameters such as the offspring distribution or the fixed tree \( t \), but they never depend on \( n \).

Convergence in probability and distribution is denoted \( \overset{p}{\to} \) and \( \overset{d}{\to} \), respectively. Unspecified limits are as \( n \to \infty \).

## 3 Proof

We begin by finding the expectation of \( n_t \) for both unconditioned and conditioned Galton–Watson trees. Let

\[
S_n := \sum_{i=1}^n \xi_i,
\]

where \( \xi_1, \xi_2, \ldots \) are i.i.d. copies of \( \xi \).

**Lemma 3.1.** Let \( t \) be a fixed ordered tree with degree sequence \( d_1, \ldots, d_k \), where thus \( k = |T| \).

(i) Then

\[
\mathbb{E} n_t(T) = \prod_{i=1}^k \mathbb{E} \left( \frac{\xi_i}{d_i} \right) = \prod_{i=1}^k \sum_{m_i=0}^\infty p_{m_i} \left( \frac{m_i}{d_i} \right).
\]

(ii) If \( n > k \), then, with \( m := \sum_{i=1}^k m_i \),

\[
\mathbb{E} n_t(T_n) = \frac{n}{n-k} \sum_{m_1, \ldots, m_k \geq 0} \prod_{i=1}^k p_{m_i} \left( \frac{m_i}{d_i} \right) \cdot \frac{(m-k+1) \mathbb{P}(S_{n-k} = n-m-1)}{\mathbb{P}(S_n = n-1)}.
\]

**Proof.** (i): We try to construct a copy \( t' \) of \( t \) in \( T \), with the given root \( o \). Let \( m_1 \) be the root degree of \( T \). Then there are \( \binom{m_1}{d_1} \) ways to choose the \( d_1 \) children of the root that belong to \( t' \). Fix one of these choices, say \( v_{11}, \ldots, v_{1d_1} \).

Next, let \( m_2 \) be the number of children of \( v_{11} \) in \( T \). Given \( m_2 \), there are \( \binom{m_2}{d_2} \) ways to choose the \( d_2 \) children of \( v_{11} \) that belong to \( t' \). Fix one of these choices.

Continuing in the same way, taking the vertices of \( t' \) in depth first order, we find for every sequence \( m_1, \ldots, m_k \) of non-negative integers, a total of \( \prod_{i=1}^k \binom{m_i}{d_i} \) choices, and each

ECP 26 (2021), paper 23.  
https://www.imstat.org/ecp
of these gives a tree \( t' \cong t \) provided the selected vertices in \( T \) have degrees \( m_1, \ldots, m_k \), which occurs with probability \( \prod_{i=1}^{k} p_{m_i} \). Hence,

\[
E n_t(T) = \sum_{m_1, \ldots, m_k \geq 0} \prod_{i=1}^{k} p_{m_i} \prod_{i=1}^{k} \left( m_i \right) \cdot \prod_{i=1}^{k} \left( p_{m_i} \right) = \prod_{i=1}^{k} \sum_{m_i=0}^{\infty} p_{m_i} \left( m_i \right),
\]

and (3.2) follows.

(ii): Consider again \( T \). We have just shown that each sequence \( m_1, \ldots, m_k \) gives \( \prod_{i=1}^{k} ( m_i ) \) choices of possible subtrees \( t' \cong t \) in \( T \), where the vertices of \( t' \) are supposed to have degrees \( m_1, \ldots, m_k \) in \( T \). This gives a total of \( m = \sum_{i=1}^{k} m_i \) children, of which \( k - 1 \) are the non-root vertices in \( t' \), and thus \( m - (k - 1) \) are unaccounted children. Then, \( |T| = n \) if and only if these \( m - k + 1 \) children and their descendants yield exactly \( n - k \) vertices.

Condition on \( m_1, \ldots, m_k \) and one of the corresponding choices of \( t' \). The probability that the \( m - k + 1 \) children above and their descendants are \( n - k \) vertices is the probability that a Galton–Watson process (with offspring distribution \( \xi \)) started with \( m - k + 1 \) individuals has total progeny \( n - k \), which by the Otter–Dwass formula [6] (see also [18] and the further references there) is given by

\[
\frac{m - k + 1}{n - k} P(S_{n-k} = n - k - (m - k + 1)).
\]

Multiplying with \( \prod_{i=1}^{k} p_{m_i} \), the probability that the vertices in \( t' \) have the right degrees in \( T \), and summing over all possibilities, we obtain

\[
E \left[ n_t(T) \right] P(|T| = n) = E \left[ n_t(T) \mid |T| = n \right] P(|T| = n) = E \left[ n_t(T) 1 \{ |T| = n \} \right]
\]

\[
= \sum_{m_1, \ldots, m_k \geq 0} \prod_{i=1}^{k} p_{m_i} \left( m_i \right) \cdot \frac{m - k + 1}{n - k} P(S_{n-k} = n - m - 1).
\]

By the Otter–Dwass formula again (this time the original case in [16]),

\[
P(|T| = n) = \frac{1}{n} P(S_n = n - 1)
\]

and (3.3) follows. (Cf. [9, Lemma 15.9] for a related result.)

We need estimates of the probabilities \( P(S_n = n - m) \). The estimate (3.8) below is standard; we expect that also (3.9) is known, but we have not found a reference, so we give a proof. (It is related to more difficult estimates in e.g. [17] assuming more moments, see Remark 3.3 below.)

**Lemma 3.2.** Suppose that \( E \xi = 1 \) and \( E \xi^2 < \infty \). Then, uniformly for all \( n \geq 1 \) and \( m \in \mathbb{Z} \),

\[
P(S_n = n - m) \leq C n^{-1/2}, \tag{3.8}
\]

\[
P(S_n = n - m) \leq C |m|^{-1}. \tag{3.9}
\]

**Proof.** (3.8): This is well-known. In fact, the classical local limit theorem, see e.g. [17, Theorem VII.1], gives the much more precise result that, uniformly in \( m \in \mathbb{Z} \) as \( n \to \infty \),

\[
P(S_n = n - m) = \frac{1}{\sigma \sqrt{2 \pi}} e^{-m^2/2 \sigma^2 n} + o(1). \tag{3.10}
\]
On general subtrees of a conditioned Galton–Watson tree

where $h$ is the span of the offspring distribution. (Provided $h|(n - m)$; otherwise the probability is 0.)

(3.9): Let $\varphi(t) := E e^{it(\xi - 1)}$ be the characteristic function of $\xi - 1 = \xi - E \xi$; note that $\varphi(t)$ is twice differentiable because $E \xi^2 < \infty$. Then, by Fourier inversion,

$$P(S_n = n - m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{int} \varphi(t)^n \, dt. \quad (3.11)$$

Hence, using an integration by parts,

$$2\pi i m P(S_n = n - m) = \int_{-\pi}^{\pi} \left( \frac{d}{dt} e^{int} \right) \varphi(t)^n \, dt = -\int_{-\pi}^{\pi} e^{int} \frac{d}{dt} (\varphi(t)^n) \, dt \quad (3.12)$$

and thus

$$|m| P(S_n = n - m) \leq \int_{-\pi}^{\pi} \left| \frac{d}{dt} (\varphi(t)^n) \right| \, dt = n \int_{-\pi}^{\pi} |\varphi'(t)| |\varphi(t)|^{n-1} \, dt. \quad (3.13)$$

The assumptions yield $\varphi'(0) = E(\xi - 1) = 0$ and $\sup |\varphi''(t)| = |\varphi''(0) = \text{Var} \xi = C < \infty$, and thus

$$|\varphi'(t)| \leq Ct. \quad (3.14)$$

Assume for simplicity that the span of $\xi$ is 1 (the general case is similar, with standard modifications). Then, as is well-known, it is easy to see that there exist $c > 0$ such that

$$|\varphi(t)| \leq e^{-ct^2}, \quad |t| \leq \pi. \quad (3.15)$$

Using (3.14) and (3.15) in (3.13) we obtain

$$|m| P(S_n = n - m) \leq nC \int_{-\pi}^{\pi} |t| e^{-ct(n-1)t^2} \, dt \leq Cn \int_{0}^{\infty} te^{-cnt^2} \, dt = C, \quad (3.16)$$

which proves (3.9). \hfill \Box

**Remark 3.3.** In the same way, taking two derivatives inside (3.11), one obtains

$$P(S_n = n - m) \leq Cn^{1/2} m^{-2}, \quad (3.17)$$

which is stronger for large $m$; note that (3.8) and (3.17) imply (3.9). Furthermore, even stronger estimates hold if we assume more moments; see [17, Theorem VII.16] for a precise asymptotic estimate assuming $E \xi^k < \infty$ for some $k \geq 3$. In fact, [17, Theorem VII.16] holds for $k = 2$ too, which can be seen by refining the argument above; this is an asymptotic estimate that is more precise than (3.17) (and implies it). \hfill \Box

**Lemma 3.4.** Let $t$ be a fixed ordered tree and suppose that $E \xi = 1$, $E \xi^2 < \infty$ and $E \xi^{\Delta(t)} < \infty$. Then $E n_t(T_n) = o(n^{1/2})$.

**Proof.** Let again the degree sequence of $t$ be $d_1, \ldots, d_k$. For a vector $m = (m_1, \ldots, m_k) \in Z_+^k$, let

$$a_m := \prod_{i=1}^{k} p_{m_i}(m_i / d_i). \quad (3.18)$$

Then, (3.2)–(3.3) and the assumption $E \xi^{\Delta(t)} < \infty$ yield

$$\sum_m a_m = E n_t(T) < \infty \quad (3.19)$$
On general subtrees of a conditioned Galton–Watson tree

and for \( n > k \), with as above \( m := \sum_i m_i =: \vert m \vert \) (and \( C = 1 \), actually),

\[
E n_t(T_n) \leq C \sum_m a_m \frac{m \mathbb{P}(S_{n-k} = n - m - 1)}{\mathbb{P}(S_n = n - 1)}.
\]  

(3.20)

Denote the summand in (3.20) by \( b_{m,n} \). By the local limit theorem (3.10), as is well-known,

\[
\mathbb{P}(S_n = n - 1) \sim cn^{-1/2},
\]  

(3.21)

and thus

\[
b_{m,n}/n^{1/2} \leq C ma_m \mathbb{P}(S_{n-k} = n - m - 1).
\]  

(3.22)

Hence, (3.8) implies that for every fixed \( m \), as \( n \to \infty \),

\[
b_{m,n}/n^{1/2} \leq C ma_m n^{-1/2} \to 0.
\]  

(3.23)

Furthermore, (3.22) and (3.9) yield

\[
b_{m,n}/n^{1/2} \leq Ca_m,
\]  

(3.24)

which is summable by (3.19). Consequently, dominated convergence shows that

\[
n^{-1/2} \sum_m b_{m,n} = \sum_m b_{m,n}/n^{1/2} \to 0,
\]  

(3.25)

which together with (3.20) yields the result \( n^{-1/2} E n_t(T_n) \to 0 \).

We will see in Example 4.4 below, that the estimate \( o(n^{1/2}) \) in Lemma 3.4 is best possible in general. However, if we assume another moment on \( \xi \), we can improve the estimate to \( O(1) \), and furthermore show that \( E n_t(T_n) \) converges. We next show this, although it is not required for our main result.

**Lemma 3.5.** Let \( t \) be a fixed tree with degree sequence \( d_1, \ldots, d_k \), and suppose that \( E \xi = 1 \). Then, as \( n \to \infty \),

\[
E n_t(T_n) \to \sum_{i=1}^{k} (d_i + 1) E \left( \xi / d_i + 1 \right) \prod_{j \neq i} E \left( \xi / d_j \right).
\]  

(3.26)

In particular, \( E n_t(T_n) = O(1) \) if \( E \xi^{\Delta(t)+1} < \infty \), while \( E n_t(T_n) \to \infty \) if \( E \xi^{\Delta(t)+1} = \infty \).

**Proof.** Define again \( a_m \) by (3.18), and denote the summand in (3.3) by \( b'_{m,n} \), where as above \( m = (m_1, \ldots, m_k) \in Z_{\geq 0}^k \). It follows from the local limit theorem (3.10) that for every fixed \( m \), as \( n \to \infty \),

\[
\frac{\mathbb{P}(S_{n-k} = n - m - 1)}{\mathbb{P}(S_n = n - 1)} = \frac{h(2\pi\sigma^2(n-k))^{-1/2}(1 + o(1))}{h(2\pi\sigma^2n)^{-1/2}(1 + o(1))} \to 1.
\]  

(3.27)

(This holds also if the span \( h > 1 \), assuming as we may that all \( m_i > 0 \), so \( h \vert m \).) Hence,

\[
b'_{m,n} \to ma_m(m - k + 1).
\]  

(3.28)

Furthermore, by (3.8) and (3.21),

\[
\frac{\mathbb{P}(S_{n-k} = n - m - 1)}{\mathbb{P}(S_n = n - 1)} \leq \frac{Cn^{-1/2}}{cn^{-1/2}} = C,
\]  

(3.29)
and thus
\[ b'_{m,n} \leq Ce_m(m - k + 1). \tag{3.30} \]
Consequently, if \( \sum_m a_m(m - k + 1) < \infty \), then
\[ \sum_m b'_{m,n} \to \sum_m a_m(m - k + 1) \tag{3.31} \]
by (3.28), (3.30) and dominated convergence. On the other hand, if \( \sum_m a_m(m - k + 1) = \infty \), then \( \sum_m b'_{m,n} \to \infty \) by (3.28) and Fatou’s lemma, and thus (3.31) holds in this case too. Recalling (3.3), this shows that in any case,
\[ \mathbb{E} n_t(T_n) \to \sum_m a_m(m - k + 1), \tag{3.32} \]
and it remains only to evaluate the limit.

Since \( t \) is a tree, we have \( \sum_{i=1}^{k} d_i = k - 1 \), and thus \( m - k + 1 = \sum_{i=1}^{k} (m_i - d_i) \).
Recalling the definition (3.18) of \( a_m \), we thus have
\[
\sum_m a_m(m - k + 1) = \sum_{i=1}^{k} (m_i - d_i) p_{m_i} \left( \frac{m_j}{d_j} \right) \prod_{j \neq i} p_m \left( \frac{m_j}{d_j} \right) 
= \sum_{i=1}^{k} \sum_{m_i=0}^{\infty} \sum_{d_i=0}^{\infty} (m_i - d_i) \left( \frac{d_j}{m_j} \right) \prod_{j \neq i} \sum_{m_j=0}^{\infty} p_m \left( \frac{m_j}{d_j} \right), \tag{3.33}
\]
which equals the right-hand side of (3.26) because \( (m_i - d_i) \left( \frac{m_j}{d_j} \right) = (d_i + 1) \left( \frac{m_j}{d_j} \right) \). This completes the proof by (3.32).

\[ \Box \]

**Remark 3.6.** Assume only \( \mathbb{E} \xi = 1 \). If \( \hat{T} \) is the infinite size-biased Galton–Watson tree defined by Kesten [11], see also [9, Section 5], then \( T_n \xrightarrow{d} \hat{T} \) in a local topology (i.e., close to the root), see [9, Theorem 7.1], and it follows that
\[ n_t(T_n) \xrightarrow{d} n_t(\hat{T}). \tag{3.34} \]
It is not difficult to see that \( \mathbb{E} n_t(\hat{T}) \) equals the right-hand side of (3.26), which thus says that \( \mathbb{E} n_t(T_n) \to \mathbb{E} n_t(\hat{T}) \). (This could presumably be used to give an alternative proof of Lemma 3.5, but we prefer the direct proof above.)

In particular, if \( \mathbb{E} \xi^{\Delta(t)+1} = \infty \), then \( \mathbb{E} n_t(\hat{T}) = \infty \), and thus (3.34) and Fatou’s lemma yield \( \mathbb{E} n_t(T_n) \to \infty \). Hence, the last sentence in Lemma 3.5 holds also without the assumption \( \mathbb{E} \xi^2 < \infty \). \[ \Box \]

We proceed to the proof of Theorem 1.1. The case \( \Delta(t) \leq 1 \) is special, since we then do not assume \( \mathbb{E} \xi^2 < \infty \), but on the other hand this case is simple and rather trivial, so we discuss it separately in the following example.

**Example 3.7.** Consider the case \( \Delta(t) \leq 1 \). This means that \( t \) is a path \( P_k \) with \( k \geq 1 \) vertices, and thus length \( k - 1 \). A copy of \( t \) in a tree \( T \) is thus a path consisting of \( k \) vertices \( v_1, \ldots, v_k \) such that \( v_{i+1} \) is a child of \( v_i \); such a path is determined by its endpoint \( v_k \), and every vertex of depth (= distance from the root) at least \( k - 1 \) is the endpoint of a copy of \( t \). Hence, if \( \nu_i(T) \) is the number of vertices in \( T \) of depth \( i \), then
\[ N_{P_k}(T) = \sum_{i \geq k-1} \nu_i(T) = |T| - \sum_{i=0}^{k-2} \nu_i(T). \tag{3.35} \]
In particular, \( N_{P_k}(T_n) = n \) and \( N_{P_{k+1}}(T_n) = n - 1 \) are deterministic; these are trivially just the numbers of vertices and edges.

Moreover, as said in Remark 3.6, assuming \( E \xi = 1 \), the random tree \( T_n \) converges locally in distribution as \( n \to \infty \), see [9, Theorem 7.1]; in particular each \( \nu_i(T_n) \) converges in distribution (to \( \nu_i(T) \)) and thus \( \nu_i(T_n) = O_p(1) \) (i.e., is bounded in probability). Hence, for every \( k \geq 1 \), (3.35) implies

\[
N_{P_k}(T_n) = n + O_p(1).
\]

(3.36)

In particular, \( N_{P_k}(T_n) \) is more strongly concentrated than the dispersion of order \( n^{1/2} \) typically seen in similar statistics, see e.g. Example 4.2 and Section 5.

Proof of Theorem 1.1. Suppose first \( \Delta(t) \leq 1 \). Then \( t = \nu_k \) for some \( k \geq 1 \) and Example 3.7 shows that (3.36) holds, and thus \( N_{P_k}(T_n)/n \to P \). Furthermore, (3.2) yields

\[
E n_{P_k}(T) = (E \xi)^{k-1} = 1,
\]

(3.37)

and thus (1.2) holds. Moreover, \( N_{P_k}(T_n)/n \leq 1 \) by (3.35), and thus dominated convergence applies to (1.2) and yields (1.3) and (1.1), see e.g. [7, Theorems 5.5.4 and 5.5.5].

In the remainder of the proof we may thus assume \( \Delta(t) \geq 1 \), and thus, in particular, \( E \xi^2 < \infty \). (The arguments below use \( E \xi^2 < \infty \), but apply to any \( \Delta(t) \).)

Lemma 3.1(i) and the assumption \( E \xi^{\Delta(t)} < \infty \) show that \( E n_{\nu_k}(T) < \infty \), and Lemma 3.4 shows \( E n_{\nu_k}(T_n) = o(n^{1/2}) \). Hence (1.2) and (1.3) follow by [10, Remark 5.3]. However, since only a sketch of the proof is given in that remark, let us add some details.

First, (1.3) follows by the argument in the proof of [10, Theorem 1.5(i)], adding the factor \( n^{1/2} \) at some places.

Next, define for \( M > 0 \) the truncation \( \nu^M_k(T) := n_k(T) \wedge M \) and let \( N^M_k(T) := \sum_{v \in T} \nu^M_k(T^v) \) be the corresponding additive functional, cf. (2.1). Let \( \epsilon > 0 \). Since \( \nu^M_k(T) \geq n_k(T) \) as \( M \to \infty \), we can by monotone convergence, and \( E n_k(T) < \infty \), choose \( M \) such that

\[
E n_k(T) - E \nu^M_k(T) < \epsilon^2.
\]

(3.38)

We have proved (1.3), and similarly \( E N^M_k(T_n)/n \to E \nu^M_k(T) \) by [10, Theorem 1.3], since \( \nu^M_k \) is bounded. Hence, (3.38) implies that for all sufficiently large \( n \),

\[
E|N_k(T_n)/n - N^M_k(T_n)/n| = E N_k(T_n)/n - E N^M_k(T_n)/n < \epsilon^2.
\]

(3.39)

Furthermore, [10, Theorem 1.3] also yields \( N^M_k(T_n)/n \to P E \nu^M_k(T) \). Consequently, using also (3.38) again, (3.39) and Markov’s inequality, if \( n \) is large,

\[
\begin{align*}
P \left( \left| N_k(T_n)/n - E n_k(T) \right| > 3\epsilon \right) & \leq P \left( \left| N_k(T_n)/n - N^M_k(T_n)/n \right| > \epsilon \right) + P \left( \left| N^M_k(T_n)/n - E \nu^M_k(T) \right| > \epsilon \right) \\
& \leq 2\epsilon.
\end{align*}
\]

(3.40)

Hence, (1.2) holds.

Finally, as said earlier, (1.2) and (1.3) are together equivalent to the \( L^1 \) convergence (1.1).

4 Examples

We give some simple but illuminating examples. Recall also Example 3.7.

ECP 26 (2021), paper 23. https://www.imstat.org/ecp

Page 8/12
Example 4.1. Let $t = t_{q,r}$ consist of two paths with $q + 1$ and $r + 1$ vertices, joined at the root; here $q, r \geq 1$. We have $k = 1 + q + r$ and $d_1 = 2$ while $d_i = 1$ for $i > 1$; thus $\Delta(t) = 2$. Since $E \xi = 1$, (3.2) yields

$$E n_{t_{q,r}}(T) = E \left( \xi \right)^2 = \frac{E \xi^2 - 1}{2} = \frac{\sigma^2}{2}.$$  
(4.1)

Hence, Theorem 1.1 yields, for any $q, r \geq 1,$

$$N_{t_{q,r}}(T)/n \xrightarrow{L^1} \sigma^2/2.$$  
(4.2)

Example 4.2. Consider the special case $q = r = 1$ of Example 4.1. Then $t_{1,1}$ is a cherry, i.e., a root with two children. If a vertex $v$ in a tree $T$ has degree $d(v)$, then the number of cherries rooted at $v$ is $\binom{d(v)}{2}$, and thus

$$N_{t_{1,1}}(T) = \sum_{v \in T} \binom{d(v)}{2} = \sum_{r=1}^{\infty} \binom{r}{2} X_r(T),$$  
(4.3)

where $X_r(T)$ is the number of vertices of degree $r$ in $T$.

It is known that $X_r(T)/n \xrightarrow{P} p_r$, see e.g. [9, Theorem 7.11]. Hence, (4.2) (with $q = r = 1$) is what we would get by dividing (4.3) by $n$ and taking the limit inside the sum; if the degree distribution is bounded, the sum is finite so this is rigorous and (4.2) (still with $q = r = 1$) follows from (4.3).

In this case we can say much more than (4.2). It was proved in [14], see also [4], that $X_r(T)/n$ is asymptotically normal, with

$$\frac{X_r(T_n) - np_r}{\sqrt{n}} \xrightarrow{d} N(0, \gamma_r^2)$$  
(4.4)

for some explicit $\gamma_r^2$. This was extended to joint convergence for all $r$ in [8], provided $E \xi^3 < \infty$. Hence, at least if $\xi$ is bounded, it follows from (4.3) that $N_{t_{1,1}}(T_n)$ is asymptotically normal, with

$$\frac{N_{t_{1,1}}(T_n) - n\sigma^2/2}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2)$$  
(4.5)

for some explicit $\gamma^2 > 0$. There are degenerate cases where $\gamma^2 = 0$. For example, for full binary trees ($P(\xi = 2) = P(\xi = 0) = \frac{1}{2}$), all degrees are 0 or 2, and then each $X_r(T)$ is a deterministic function of $|T|$; hence (4.3) shows that $N_{t_{1,1}}(T_n)$ is deterministic. More generally, the same happens for full $m$-ary trees, with $\xi \in \{0, m\}$ a.s., for any $m \geq 2$. But it can be seen from the covariances given in [8] that $\gamma^2 > 0$ in all other cases with bounded $\xi$. See further Section 5. □

Example 4.3. Let $\ell \geq 1$, and let $\omega_{\ell}(T)$ be the number of (undirected) paths of length $\ell$ in $T$. For definiteness, we count undirected paths, so this equals the number of unordered pairs $(v, w)$ of vertices of distance $\ell$. There are two cases:

(i) $v$ is an ancestor of $w$, or conversely; the number of such pairs is $N_{p_\ell}(T)$.

(ii) Neither $v$ nor $w$ is an ancestor of the other. Then $v$ and $w$ are the two leaves in a copy of $t_{q,r}$ with $q, r \geq 1$ and $q + r = \ell$. For given $q$ and $r$, the number of such pairs equals $N_{t_{q,r}}(T)$.
Consequently,
\[ \varpi_l(T) = n_{p_l}(T) + \sum_{q=1}^{\ell-1} N_{n_{p_{\ell-q}}}(T). \]  
(4.6)

Hence, Examples 3.7 and 4.1 yield
\[ \varpi_l(T_n)/n \xrightarrow{L_1} 1 + (\ell - 1)\sigma^2/2. \]  
(4.7)

For example, taking \( \xi \sim \text{Po}(1) \) we obtain (forgetting the ordering) a uniformly random unordered labelled tree; we have \( \sigma^2 = 1 \) and thus (4.7) yields
\[ \varpi_l(T_n) \xrightarrow{L_1} (\ell + 1)/2. \]  
(4.8)

Similarly, taking \( \xi \sim \text{Ge}(1/2) \) we obtain a uniformly random ordered tree; we have \( \sigma^2 = 2 \) and thus (4.7) then yields
\[ \varpi_l(T_n) \xrightarrow{L_1} \ell. \]  
(4.9)

Taking \( \xi \sim \text{Bi}(2,1/2) \) we obtain a uniformly random binary tree; we have \( \sigma^2 = 1/2 \) and thus (4.7) now yields
\[ \varpi_l(T_n) \xrightarrow{L_1} (\ell + 3)/4. \]  
(4.10)

The following example shows that the estimate \( o(n^{1/2}) \) in Lemma 3.4 is best possible.

**Example 4.4.** For simplicity, let the tree \( t \) be a star, where the root has degree \( \Delta \geq 2 \) and its children are leaves with degree 0. (The argument is easily modified to any tree \( t \) with \( \Delta(t) \geq 2 \)) Thus \( k := |t| = \Delta + 1 \). Assume that the span of \( \xi \) is 1.

The local limit theorem (3.10) implies that if \( n \) is large and \( m \leq n^{1/2} \), then
\[ P(S_{n-k} = n - m - 1) \geq cn^{-1/2}, \]  
(4.11)

and thus, using (3.21),
\[ P(S_{n-k} = n - m - 1)/P(S_n = n - 1) \geq c. \]  
(4.12)

Hence, by (3.3) and considering there only terms with \( m_2 = \cdots = m_k = 0 \),
\[ E n_k(T_n) \geq c \quad \sum_{\Delta < m_1 \leq n^{1/2}} p_{m_1} \left( \frac{m_1}{\Delta} \right) m_1 \geq c \quad \sum_{\Delta < m \leq n^{1/2}} m \Delta^{-1}. \]  
(4.13)

If \( \epsilon > 0 \), and we let \( p_m = m^{-\Delta-1-\epsilon} \) for large \( m \), then \( E \xi^\Delta < \infty \), and (4.13) yields, for large \( n \),
\[ E n_k(T_n) \geq c \quad \sum_{\Delta < m \leq n^{1/2}} m^{-\epsilon} \geq cn^{(1-\epsilon)/2}. \]  
(4.14)

Hence, for any \( \epsilon > 0 \), \( E n_k(T_n) \) can grow faster than \( n^{1/2-\epsilon} \).

Similarly, we can find an offspring distribution \( (p_m)_{m=0}^\infty \) satisfying the conditions such that \( E n_k(T_n) = n^{1/2-o(1)} \); we omit the details. Moreover, for any given sequence \( \delta(n) \searrow 0 \), we can find \( (p_m)_{m=0}^\infty \) such that \( E n_k(T_n) \geq \delta(n)n^{1/2} \), at least for a subsequence. To see this, take an increasing sequence \( (\delta_j)_{j=1}^\infty \) with \( \sum_{j=1}^\infty j \delta(m_j^2) < 1 \). Let \( p_{m_j} := j \delta(m_j^2)m_j^{-\Delta} \), and \( p_m = 0 \) for all other \( m \geq 2 \), choosing \( \delta_0 \) and \( \delta_1 \) such that \( \sum_{i} p_i = \sum_{i} \delta_i = 1 \). Also, let \( n_j := m_j^2 \). Then (4.13) implies that, for large \( j \),
\[ E n_k(T_{n_j}) \geq c p_{m_j} m_j^{\Delta+1} = c j m_j \delta(m_j^2) \geq m_j \delta(m_j^2) = n_j^{1/2} \delta(n_j). \]  
(4.15)

\[ \square \]
5 Asymptotic normality?

We showed in Example 4.2 that if \( \xi \) is bounded, then \( N_{t,x}(T_n) \) is asymptotically normal in the sense that (4.5) holds (although \( \gamma^2 = 0 \) is possible). In fact, this holds for any fixed tree \( t \).

**Proposition 5.1.** Assume that \( \xi \) is bounded. Then, for any fixed tree \( t \),

\[
\frac{N_t(T_n) - n\mu_t}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2_t),
\]

for \( \mu_t := E(n_t(T)) \) and some \( \gamma^2_t \geq 0 \).

**Proof.** This follows from the result by Chyzak, Drmota, Klausner and Kok [2] on patterns discussed in Section 1 (extended to conditioned Galton–Watson trees [2; 12; 13]); the assumption on \( \xi \) means that vertex degrees are bounded by some constant, and thus there is a finite number of patterns that correspond to subtrees isomorphic to \( t \); hence \( N_t(T_n) \) is a linear combination of pattern counts, and the result follows from the joint asymptotic normality of the latter. (See also [15] for a special case.)

Alternatively, this is an application of [10, Theorem 1.13]: the functional \( n_t \) is local (as defined in [10]) and for trees with degrees bounded by some constant \( K \), \( n_t \) is bounded. Hence (5.1) follows from [10, Theorem 1.13].

We conjecture that this behaviour is typical, and that Proposition 5.1 holds for every \( \xi \) with \( E\xi = 1 \) that satisfies a suitable moment condition. However, it seems that substantial additional work would be required to show this. As said in the introduction, this was briefly discussed in [2], but it seems that the method there requires extensions to infinite systems of functional equations. (As suggested by a referee, it is possible that the methods of [5] might be applicable, at least when \( \xi \) has a finite exponential moment, but we have not pursued this.)

Similarly, the application of [10, Theorem 1.13] requires \( n_t(T_n) \) to be bounded, which is not the case when \( \xi \) is unbounded. It is possible that this may be overcome by truncations and some variance estimates, but again more work is needed. (The extension in [19] applies to the case when \( t \) is a star with root degree \( \Delta \) (including Example 4.2 with \( \Delta = 2 \)) and \( E\xi^{2\Delta+1} < \infty \); this might suggest further extensions.) This problem is thus left for future research.

Note also that there are degenerate cases when the asymptotic variance in (5.1) \( \gamma^2_t = 0 \); see Examples 3.7 and 4.2. (Then (5.1) does not give asymptotic normality; only a concentration result.) However, we conjecture that this is an exception, occurring only in a few special cases.

**References**

On general subtrees of a conditioned Galton–Watson tree


**Acknowledgments.** I thank Stephan Wagner and an anonymous referee for helpful comments.