ASYMPTOTIC NORMALITY OF FRINGE SUBTREES AND ADDITIVE FUNCTIONALS IN CONDITIONED GALTON–WATSON TREES

SVANTE JANSON

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. Another example is the number of protected nodes. The offspring distribution defining the random tree is assumed to have expectation 1 and finite variance; no further moment condition is assumed.

1. Introduction

All trees in this paper are *rooted* and *ordered* (= *plane*). (We assume that the trees are ordered, i.e., that the children of each node are ordered, for technical convenience. In applications, the ordering is often irrelevant, and we may then treat unordered trees too by using a random labelling.) We consider in the present paper only finite trees (except $\hat{\mathcal{T}}$ in Lemma 5.9), and denote the size (or order), i.e. the number of nodes, of a tree T by $|T|$. We let $\mathfrak{T}$ denote the countable set of all ordered rooted trees (where we identify trees that are isomorphic in the natural way, with an isomorphism preserving the root and the orderings of children); let further $\mathfrak{T}_n$ be the (finite) subset of all such trees of order n. (See further e.g. [15] and [30].)

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.* (By "subtree", we mean in the present paper always a fringe subtree, except in Example 2.4.) We are interested in the collection $\{T_v\}$ of all fringe subtrees of a given tree T.

One way to study this collection is to consider the *random fringe subtree* T_*, which is the random rooted tree obtained by taking the subtree T_v at a uniformly random node v in T. This was introduced and studied by Aldous [2], both in general and for many important examples. We let, for $T, T' \in \mathfrak{T}$,

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

Date: 4 December, 2013.

2010 *Mathematics Subject Classification.* 60C05; 05C05.

Partly supported by the Knut and Alice Wallenberg Foundation.

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,

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

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 [2] 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

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

Remark 1.2. This is the *quenched* version of the fringe subtree T_*, where we first select a realization of the random tree T, and then fix this realization and choose $v \in T$ uniformly at random, yielding a fringe subtree T_* with a distribution depending on T; this is thus a random distribution, as said above. The alternative is the *annealed* version where we take a random tree T and a uniformly random node v in it as a combined random event; this yields a random fringe subtree with a distribution that is the expectation of the random distribution (1.2) in the quenched version. When $|T|$ is fixed (as in the cases we study in the present paper), the annealed version thus corresponds to considering only the expectation $\mathbb{E} F(T) = |T| \mathbb{E} f(T_*)$ of the sum (1.3), or equivalently $\mathbb{E} f(T_*)$, while the quenched version corresponds to studying the conditional expectation (1.8).

The random trees that we consider in this paper are conditioned Galton–Watson trees, see Section 3 for definition and notation. (Related results for some other random trees are given by Fill and Kapur [19, 20] (m-ary search trees under different models) and Holmgren and Janson [24] (random binary search trees and random recursive trees).) The Galton–Watson trees are defined using an offspring distribution; we let ξ denote a random variable with this distribution and we assume throughout the paper that the mean $\mathbb{E}\xi = 1$ and (except in Theorem 1.3) that the variance $\sigma^2 := \operatorname{Var}\xi$ is finite (and non-zero). We recall the well-known fact that several standard examples of random trees can de defined in this way, for example uniform random ordered trees ($\xi \sim \mathrm{Ge}(1/2)$, $\sigma^2 = 2$), uniform random labelled trees ($\xi \sim \mathrm{Po}(1)$, $\sigma^2 = 1$) and uniform random binary trees ($\xi \sim \mathrm{Bi}(2, 1/2)$, $\sigma^2 = 1/2$), see e.g. Aldous [3], Devroye [12], Drmota [15], Janson [30].

The results in Aldous [2] 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. Part (i) was proved by Aldous [2], assuming $\operatorname{Var}\xi < \infty$ as we assume in the rest of the paper, and extended to more general ξ by Bennies and Kersting [8], and further by Janson [30]; the sharper version (ii) is proved in [30, Theorem 7.12].

Theorem 1.3 (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 $n \to \infty$:*

(i) *(Annealed version.) The fringe subtree $\mathcal{T}_{n,*}$ converges in distribution to the Galton–Watson tree $\mathcal{T}$. I.e., for every fixed tree T,*

$$\frac{\mathbb{E}\, n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_{n,*} = T) \to \mathbb{P}(\mathcal{T} = T). \tag{1.9}$$

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

$$\mathbb{E}\,\frac{F(\mathcal{T}_n)}{n} = \mathbb{E} f(\mathcal{T}_{n,*}) \to \mathbb{E} f(\mathcal{T}). \tag{1.10}$$

(ii) *(Quenched version.) The conditional distributions $\mathcal{L}(\mathcal{T}_{n,*} \mid \mathcal{T}_n)$ converge to the distribution of $\mathcal{T}$ in probability. I.e., for every fixed tree*

T,

$$\frac{n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_{n,*} = T \mid \mathcal{T}_n) \overset{\mathrm{p}}{\longrightarrow} \mathbb{P}(\mathcal{T} = T). \tag{1.11}$$

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

$$\frac{F(\mathcal{T}_n)}{n} = \mathbb{E}\, f\big(\mathcal{T}_{n,*} \mid \mathcal{T}_n\big) \overset{\mathrm{p}}{\longrightarrow} \mathbb{E}\, f(\mathcal{T}). \tag{1.12}$$

□

Remark 1.4. The statement in [30, Theorem 7.12] uses (1.9) and (1.11), here expanded using (1.7). The equivalences with (1.10) and (1.12) follow by standard properties of convergence in distribution, see e.g. [9, Theorem 2.1 and Section 4]. (Note that the set of finite ordered trees is a countable discrete set, which simplifies the situation and e.g. justifies that it is enough to consider point probabilities in (1.9) and (1.11). To show (1.12) it may be convenient to use the Skorohod representation theorem [32, Theorem 4.30] and assume that (1.11) holds a.s. for every T.)

The result is easily extended to include also unbounded f with suitable growth conditions, see for example Theorem 1.5(i),(ii) and Remark 5.3 below.

Theorem 1.3 is a law of large numbers for $F(\mathcal{T}_n)$. In the present paper we take the next step and study the fluctuations of $F(\mathcal{T}_n)$; we prove a central limit theorem, i.e., asymptotic normality of $F(\mathcal{T}_n)$ under suitable assumptions. This includes, as a special case, (joint) normal convergence of the subgraph counts $n_{T'}(T)$, see Corollary 1.8. Our main result is the following. (The proof of this and the following results is given in Section 8.)

Theorem 1.5. *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*

$$\mathbb{E}\, F(\mathcal{T}_n) = n\mu + o\big(\sqrt{n}\big). \tag{1.13}$$

(ii) *If*

$$\mathbb{E}\, f(\mathcal{T}_n)^2 \to 0 \tag{1.14}$$

as $n \to \infty$, and

$$\sum_{n=1}^{\infty} \frac{\sqrt{\mathbb{E}(f(\mathcal{T}_n)^2)}}{n} < \infty, \tag{1.15}$$

then

$$\operatorname{Var} F(\mathcal{T}_n) = n\gamma^2 + o(n) \tag{1.16}$$

where

$$\gamma^2 := 2\,\mathbb{E}\Bigl(f(\mathcal{T})\bigl(F(\mathcal{T})-|\mathcal{T}|\mu\bigr)\Bigr) - \operatorname{Var} f(\mathcal{T}) - \mu^2/\sigma^2 \tag{1.17}$$

is finite; moreover,

$$\frac{F(\mathcal{T}_n)-n\mu}{\sqrt{n}} \overset{\mathrm{d}}{\longrightarrow} N(0,\gamma^2). \tag{1.18}$$

By (1.13), we may replace $n\mu$ by the exact mean $\mathbb{E}\,F(\mathcal{T}_n)$ in (1.18).

Remark 1.6. By (4.13), the condition $\mathbb{E}\,|f(\mathcal{T})| < \infty$ is equivalent to $\sum_n n^{-3/2}\,\mathbb{E}\,|f(\mathcal{T}_n)| < \infty$; in particular, this holds if $\mathbb{E}\,|f(\mathcal{T}_n)| = O(1)$, and thus if (1.14) holds. (It is also implied by (1.15).)

Remark 1.7. It follows from (1.16) that $\gamma^2 \geqslant 0$. We do not know whether $\gamma^2 = 0$ is possible except in trivial cases when $F(\mathcal{T}_n)$ is deterministic for all n.

Special cases of Theorem 1.5 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 [37], see Example 2.1. (See also Aldous [2, Remark 7.5.3].) Wagner [51] considered random labelled trees (the case $\xi \sim \mathrm{Po}(1)$) and showed Theorem 1.5 (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 (1.14)–(1.15)).

Theorem 1.5 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 (i.e., by considering linear combinations); the asymptotic covariances follow from (1.17) by polarization in the usual way (i.e., using e.g. $\operatorname{Cov}(X,Y) = \frac{1}{4}(\operatorname{Var}(X+Y) - \operatorname{Var}(X-Y))$). One example is the following corollary for the subtree counts (1.1); by (1.7), this corollary shows that the conditional distribution $\mathcal{L}(\mathcal{T}_{n,*} \mid \mathcal{T}_n)$ of the fringe subtree $\mathcal{T}_{n,*}$ of $\mathcal{T}_n$ has asymptotically Gaussian fluctuations around the limit distribution given by Theorem 1.3.

Corollary 1.8. *The subtree counts $n_T(\mathcal{T}_n)$, $T \in \mathfrak{T}$, are asymptotically jointly normal. More precisely, let $\pi_T := \mathbb{P}(\mathcal{T} = T)$,*

$$\gamma_{T,T} := \pi_T - \bigl(2|T| - 1 + \sigma^{-2}\bigr)\pi_T^2, \tag{1.19}$$

and, for $T_1 \neq T_2$,

$$\gamma_{T_1,T_2} := n_{T_2}(T_1)\pi_{T_1} + n_{T_1}(T_2)\pi_{T_2} - \bigl(|T_1| + |T_2| - 1 + \sigma^{-2}\bigr)\pi_{T_1}\pi_{T_2}. \tag{1.20}$$

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

$$\mathbb{E}\,n_T(\mathcal{T}_n) = n\pi_T + o\bigl(\sqrt{n}\bigr), \tag{1.21}$$

$$\operatorname{Var} n_T(\mathcal{T}_n) = n\gamma_{T,T} + o(n), \tag{1.22}$$

$$\operatorname{Cov}\bigl(n_{T_1}(\mathcal{T}_n), n_{T_2}(\mathcal{T}_n)\bigr) = n\gamma_{T_1,T_2} + o(n), \tag{1.23}$$

$$\frac{n_T(\mathcal{T}_n) - n\pi_T}{\sqrt{n}} \overset{\mathrm{d}}{\longrightarrow} Z_T, \tag{1.24}$$

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

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 1.8 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 (1.17) or, equivalently, follows from (1.19)–(1.20).

For functionals with finite support, we can show that $\gamma^2 > 0$ except in trivial cases, cf. Remark 1.7.

Theorem 1.9. *Suppose that* f *is a functional on* $\mathfrak{T}$ *with finite support, and that* $\gamma^2 = 0$.

- (i) *If the support* $\{k : p_k > 0\}$ *of* ξ *contains at least two positive integers, then* $f(\mathcal{T}) = F(\mathcal{T}) = F(\mathcal{T}_n) = 0$ *a.s.*
- (ii) *Otherwise, i.e., if* $\{k : p_k > 0\} = \{0, r\}$ *for some* $r > 1$*, then* $f(\mathcal{T}) = a\mathbf{1}\{|\mathcal{T}| = 1\}$ *for some real* a *and* $F(\mathcal{T}) = a(n - (n-1)/r)$ *is deterministic.*

Equivalently, in (i), the matrix $(\gamma_{T_1,T_2})_{|T_1|,|T_2|\leqslant M}$ (where we only consider trees T_1, T_2 with $\mathbb{P}(\mathcal{T} = T_j) > 0$) is positive definite for every M; in (ii) the submatrix $(\gamma_{T_1,T_2})_{2\leqslant|T_1|,|T_2|\leqslant M}$ is positive definite.

Remark 1.10. The condition (1.14) in Theorem 1.5(ii) is equivalent to $\mathbb{E} f(\mathcal{T}_n) \to 0$ together with $\operatorname{Var} f(\mathcal{T}_n) \to 0$, and it implies $\mathbb{E}|f(\mathcal{T}_n)| \to 0$ as assumed in (i). Both this condition and (1.15) 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 $|T| > 1$). If we assume better integrability of ξ, we can weaken the condition a little, see Remark 6.9, but not by much. 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 (1.17) vanish, so $\gamma^2 = -\sigma^{-2} < 0$, which is absurd for an asymptotic variance, and (1.16) and (1.18) fail. Nevertheless, in this trivial counterexample, it is only the value of γ^2 that is wrong; (1.16) and (1.18) trivially hold with $\gamma^2 = 0$. Example 6.14 is a more complicated counterexample where f is bounded (and $f(T) \to 0$ as $|T| \to \infty$ so (1.14) holds) but at least one of (1.16) and (1.18) fails (for any finite γ^2); we conjecture that both fail in this example. Example 6.13 is a related example where (1.14) holds but $\operatorname{Var} F(\mathcal{T}_n)/n \to \infty$.

Remark 1.11. If we go further and allow $f(T)$ that grow with the size $|T|$, we cannot expect the results to hold. Fill and Kapur [18] have made an interesting and illustrative study (for certain f) of the case of binary trees, which is the case $\xi \sim \text{Bin}(2, 1/2)$ 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. We conjecture that similar results hold for general conditioned Galton–Watson trees and increasing f, but the precise limits presumably depend on the offspring distribution ξ.

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 1.12. 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 [26] (binary search tree, f depends on $|T|$ only), Fill and Kapur [19] (m-ary search tree, f depends on $|T|$ only), Holmgren and Janson [24] (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 1.3 is the distribution of $|\mathcal{T}|$, decays rather slowly, with $\mathbb{P}(|\mathcal{T}| = n) \asymp n^{-3/2}$, see (4.13), 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 [2]. Cf. also the related results in Fill, Flajolet and Kapur [17, 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.

The counterexamples in Examples 6.13–6.14 are constructed to have rather large correlations between $f(T_v)$ and $f(T_w)$ for different subtrees T_v and T_w. In typical applications, this is not the case, and we expect Theorem 1.5 to hold also for nice functions f that do not quite satisfy (1.14) and (1.15). A simple example is the number of nodes of outdegree r, for some fixed $r \geqslant 0$ (with $p_r > 0$). This equals $F(T)$ if we let $f(T) := \mathbf{1}\{\text{the root of } T \text{ has degree } r\}$. In this case, Kolchin [37, Theorem 2.3.1] has proved asymptotic normality, see further Examples 2.1–2.2. We can extend this as follows.

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 1.13. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree as in Theorem 1.5. Suppose that $f : \mathfrak{T} \to \mathbb{R}$ is a bounded and local functional. Then the conclusions* (1.13)*,* (1.16) *and* (1.18) *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 $\sum_n |\mathbb{E} f(\mathcal{T}_n)|/n < \infty$.

The proof in Section 8 shows also that the asymptotic variance γ^2 equals $\lim_{N\to\infty}(\gamma^{(N)})^2$, where $(\gamma^{(N)})^2$ is given by (8.5) or, in the case of a bounded local functional, (8.5) applied to $f(T) - \mathbb{E} f(\hat{\mathcal{T}})$, with $\hat{\mathcal{T}}$ defined in (5.39).

We give some examples in Section 2. Sections 3–4 contain preliminaries. The expectation $\mathbb{E} F(\mathcal{T}_n)$ is studied in Section 5, and Theorem 1.5(i) is proved. Section 6 establishes bounds for the variance $\operatorname{Var} F(\mathcal{T}_n)$, and proves the asymptotic (1.16) in the special case of a functional f with finite support. Section 7 shows asymptotic normality for functionals f with finite support. Finally, in Section 8, the variance bounds in Section 6 and a truncation argument are used to extend the latter results to more general f, completing the proofs of the theorems above.

Remark 1.14. In Theorem 1.13, γ^2 is not always given by (1.17) because $\mathbb{E}\big(f(\mathcal{T})\big(F(\mathcal{T}) - |\mathcal{T}|\mu\big)\big)$ does not necessarily exist (in the usual sense, as an absolutely convergent integral), see Example 2.2; thus we in general take limits using truncations.

Remark 1.15. Although the Galton–Watson tree $\mathcal{T}$ is finite a.s., its expected size $\mathbb{E}|\mathcal{T}| = \infty$, as is seen from (4.13) or directly from the definition. Since the random fringe tree $\mathcal{T}_{n,*} \xrightarrow{\mathrm{d}} \mathcal{T}$ by Theorem 1.3, it follows that $\mathbb{E}|\mathcal{T}_{n,*}| \to \infty$; similarly, Theorem 1.3(ii) implies $\mathbb{E}\big(|\mathcal{T}_{n,*}| \mid \mathcal{T}_n\big) \xrightarrow{\mathrm{p}} \infty$.

In fact, for any tree T with $|T| = n$, letting $d(v)$ be the depth of v and defining a partial order on the nodes of T by $v \leqslant w$ if v is on the path from the root to w,

$$\mathbb{E}|T_*| = \frac{1}{n}\sum_{v\in T}|T_v| = \frac{1}{n}\sum_{v,w\in T}\mathbf{1}\{v \leqslant w\} = \frac{1}{n}\sum_{w\in T}\big(d(w)+1\big) = 1+\frac{1}{n}\sum_{w\in T}d(w),$$

i.e., 1 plus the average path length. Well-known results on the average path length in a conditioned Galton–Watson tree, see Aldous [3, 4], thus imply

$$n^{-1/2}\,\mathbb{E}\big(|\mathcal{T}_{n,*}| \mid \mathcal{T}_n\big) \xrightarrow{\mathrm{d}} \sigma^{-1}\hat{\xi}, \tag{1.25}$$

where $\hat{\xi}$ is twice the Brownian excursion area. Hence, although the distribution of the size of a random fringe tree is tight, so the size is bounded in probability, the average fringe tree size is of the order $n^{1/2}$. Similarly,

$$\mathbb{E}|T_*|^2 = \frac{1}{n}\sum_{v,w,u\in T}\mathbf{1}\{v \leqslant w,\, v \leqslant u\} = \frac{1}{n}\sum_{u,w\in T}\big(d(w\wedge u)+1\big)$$

and by [28, Theorem 3.1],

$$n^{-3/2}\,\mathbb{E}\bigl(|\mathcal{T}_{n,*}|^2 \mid \mathcal{T}_n\bigr) \overset{\mathrm{d}}{\longrightarrow} \sigma^{-1}\eta, \tag{1.26}$$

for a certain positive random variable η. Hence the average of the square of the fringe tree size is of the order $n^{3/2}$.

1.1. **Some notation.** All unspecified limits are as $n \to \infty$.

We let $C_1, C_2, \dots$ and $c_1, c_2, \dots$ denote unspecified positive constants (possibly depending on f and ξ, but not on n and other variables, and possibly different at different occurences). (We use C_i for large constants and c_i for small.) We also use standard O and o notation (with the limit in o as $n \to \infty$ unless otherwise said). Moreover, we sometimes use the less common notation (for $a_n, b_n \geqslant 0$) $a_n \ll b_n$ for $a_n = O(b_n)$ (or, equivalently, $a_n \leqslant C_1 b_n$).

The *outdegree* of a node in a tree is its number of children. (This is, except for the root, the degree minus 1.) The *degree sequence* of a tree $T \in \mathfrak{T}_n$ is the sequence $(d_1, \dots, d_n)$ of outdegrees of the nodes taken in depth-first order, i.e., starting with the (out)degree d_1 of the root and then taking the degree sequences of the branches T_{v_i} one by one, where $v_1, \dots, v_{d_1}$ are the children of the root, in order. It is easily seen that a sequence $(d_1, \dots, d_n) \in \mathbb{N}^n$ (where $\mathbb{N} := \{0, 1, 2, \dots\}$) is the degree sequence of a tree $T \in \mathfrak{T}_n$ if and only if

$$\begin{cases} \sum_{i=1}^{j} d_i \geqslant j, & 1 \leqslant j < n, \\ \sum_{i=1}^{n} d_i = n-1, \end{cases} \tag{1.27}$$

see e.g. [30, Lemma 15.2]. Note also that a tree in $\mathfrak{T}$ is uniquely determined by its degree sequence.

The *depth* of a node v in a tree is its distance to the root; we denote it by $d(v)$.

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 1.5 and 1.13 both apply and show asymptotic normality of $F(\mathcal{T}_n)$, and so does Corollary 1.8 since $F(T) = n_\bullet(T)$, where $\bullet$ is the tree of order 1; (1.17) 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 (1.19). The asymptotic normality in this case (and a local limit theorem) was proved by Kolchin [37, Theorem 2.3.1]. By Theorem 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 [37, Theorem 2.3.1], with

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

where

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

see also Janson [27] (joint convergence and moment convergence, assuming at least $\mathbb{E}\,\xi^3 < \infty$), Minami [45] and Drmota [15, 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 [35] and [30, Theorem 7.10]. Hence (1.14) and (1.15) both fail, and we cannot apply Theorem 1.5. (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 1.13 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 1.13 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 1.13 shows existence of γ^2 but the formula (given by the proof) as a limit of (8.5) 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 [27] provided $\mathbb{E}\,\xi^3 < \infty$); we give this proof in Section 8.

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 (4.13),

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

which shows that the expectation in (1.17) does not exist, so γ^2 is not given by (1.17).

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 [11] and Mansour [42], and convergence in probability of the fraction of protected nodes is proved for general conditioned Galton–Watson trees by Devroye and Janson [13].

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 1.13 applies and shows asymptotic normality of $F(\mathcal{T}_n)$.

The asymptotic mean $\mu = \mathbb{E}\, f(\mathcal{T})$ is easily calculated, see [13] 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 (8.5) (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, perhaps related to our proof of (2.6) in Section 8, 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 [51] 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 [51] used this and the special case of Theorem 1.5 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 1.5, 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 (1.14)–(1.15) hold. Consequently,

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

for some $\mu = \mathbb{E} \log\bigl(1 + s_1(\mathcal{T})^{-1}\bigr)$ and γ^2 given by (1.17) (both depending on the distribution of ξ); Wagner [51] makes a numerical calculation of μ and σ^2 for his case.

Furthermore, as noted in [51], $s_1(T) \leqslant s(T) \leqslant |T|s_1(T)$ 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 [51, 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 $\sim \gamma^2$ 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 [51] for the case considered there. (Note that a uniformly random arbitrary subtree thus is much larger than a uniformly random fringe subtree, see Remark 1.15.)

Example 2.5. Another example by Wagner [51] is the number of nodes whose children all are leaves (i.e., no grandchildren; cf. Example 2.3). This is $F(T)$ with $f(T) := \mathbf{1}\{T \text{ has no nodes of depth} > 1\}$. This is a bounded local functional, so Theorem 1.13 applies and shows asymptotic normality of $F(\mathcal{T}_n)$ for any conditioned Galton–Watson tree with $\mathbb{E}\,\xi = 1$ and $\sigma^2 < \infty$, generalizing the result by Wagner [51]. Moreover, $f(T) = 1$ only if T is a star, and thus $\mathbb{E}\,f(\mathcal{T}_n) = p_{n-1}p_0^{n-1}/\,\mathbb{P}(|\mathcal{T}| = n) = O\bigl(n^{3/2}p_0^n\bigr)$ so (1.14)–(1.15) hold and Theorem 1.5 applies too.

3. Galton–Watson trees

Given a random nonnegative integer-valued random variable ξ, with distribution $\mathcal{L}(\xi)$, the *Galton–Watson tree* $\mathcal{T}$ with offspring distribution $\mathcal{L}(\xi)$ is constructed recursively by starting with a root and giving each node a number of children that is a new copy of ξ, independent of the numbers of children of the other nodes. (This is thus the family tree of a Galton–Watson process, see e.g. [7].) Obviously, only the distribution of ξ matters; we sometimes abuse language and say that $\mathcal{T}$ has offspring distribution ξ. We assume that $\mathbb{P}(\xi = 0) > 0$ (otherwise the tree is a.s. infinite).

Furthermore, let $\mathcal{T}_n$ be $\mathcal{T}$ conditioned on having exactly n nodes; this is called a *conditioned Galton–Watson tree.* (We consider only n such that $\mathbb{P}(|\mathcal{T}| = n) > 0$.)

Remark 3.1. It is well-known that the Galton–Watson tree $\mathcal{T}$ is a.s. finite if and only if $\mathbb{E}\,\xi \leqslant 1$, see [7]. We will in this paper assume that $\mathbb{E}\,\xi = 1$, the *critical* case. In most cases, but not all, a conditioned Galton–Watson tree with an offspring distribution ξ' with an expectation $\mathbb{E}\,\xi' \neq 1$ is equivalent to a conditioned Galton–Watson tree with another offspring distribution ξ satisfying $\mathbb{E}\,\xi = 1$, so this is only a minor restriction. See e.g. [30] for details.

Remark 3.2. The degree sequence of the Galton–Watson tree $\mathcal{T}$ (when finite) equals a sequence $\xi_1, \xi_2, \dots$ of independent copies of ξ, truncated at the unique place making the sequence a degree sequence of a tree, cf. (1.27). This follows immediately from (and is equivalent to) the definition of $\mathcal{T}$. The degree sequence of the conditioned Galton–Watson tree $\mathcal{T}_n$ is more complicated and will be described in Section 4.

Remark 3.3. For any given n, $\mathfrak{T}_n$ is finite, so there is only a finite number of possible realizations of $\mathcal{T}_n$. Hence, for any functional f, the random variables $f(\mathcal{T}_n)$ and $F(\mathcal{T}_n)$ are bounded for each n; in particular they always have finite expectations and higher moments.

Conditioned Galton–Watson trees are also known as (a special case of) *simply generated random trees*, see e.g. [30].

4. Preliminaries and more notation

We assume throughout the paper that $f : \mathfrak{T} \to \mathbb{R}$ is a given functional on trees, and that F is the corresponding subtree sum given by (1.3). We assume further that $\mathcal{T}$ $[\mathcal{T}_n]$ is a [conditioned] Galton–Watson tree with a given offspring distribution ξ, with $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 := \operatorname{Var}\xi < \infty$. We let $\xi_1, \xi_2, \dots$ be a sequence of independent copies of ξ, and let

$$S_n := \sum_{i=1}^{n} \xi_i. \tag{4.1}$$

We denote the probability distribution of ξ by $(p_k)_0^\infty$, i.e., $p_k := \mathbb{P}(\xi = k)$.

4.1. We recall the local limit theorem, see e.g. [37, Theorem 1.4.2] or [48, Theorem VII.1], which in our setting can be stated as follows. Recall that the *span* of an integer-valued random variable ξ is the largest integer h such that $\xi \in a + h\mathbb{Z}$ a.s. for some $a \in \mathbb{Z}$; we will only consider ξ with $\mathbb{P}(\xi = 0) > 0$ and then the span is the largest integer h such that $\xi/h \in \mathbb{Z}$ a.s., i.e., the greatest common divisor of $\{n : \mathbb{P}(\xi = n) > 0\}$. (Typically, $h = 1$, but we have for example $h = 2$ in the case of full binary trees, with $p_0 = p_2 = 1/2$.) In the case we are interested in, the local limit theorem can be stated as follows.

Lemma 4.1. *Suppose that ξ is an integer-valued random variable with $\mathbb{P}(\xi = 0) > 0$, $\mathbb{E}\,\xi = 1$, $0 < \sigma^2 := \operatorname{Var}\xi < \infty$ and span h. Then, as $n \to \infty$, uniformly in all $m \in h\mathbb{Z}$,*

$$\mathbb{P}(S_n = m) = \frac{h}{\sqrt{2\pi\sigma^2 n}}\Bigl(e^{-(m-n)^2/(2n\sigma^2)} + o(1)\Bigr). \tag{4.2}$$

$\square$

In particular, which we will use repeatedly, as $n \to \infty$ with $n \equiv 1 \pmod{h}$,

$$\mathbb{P}(S_n = n-1) \sim \frac{h}{\sqrt{2\pi\sigma^2}}\, n^{-1/2}. \tag{4.3}$$

4.2. As said above, 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

$$f(d_1,\dots,d_n) := \begin{cases} f(T), & (d_1,\dots,d_n) \text{ is the degree sequence of a tree } T, \\ 0, & \text{otherwise (i.e., (1.27) is not satisfied).} \end{cases} \tag{4.4}$$

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, i.e., satisfies (1.27). By the definition (4.4), we thus can write (1.3) as

$$F(T) = \sum_{1\leqslant i\leqslant j\leqslant n} f(d_i,\dots,d_j) = \sum_{k=1}^{n} \sum_{i=1}^{n-k+1} f(d_i,\dots,d_{i+k-1}). \tag{4.5}$$

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

$$F(T) = \sum_{k=1}^{n} \sum_{i=1}^{n} f(d_i,\dots,d_{i+k-1}). \tag{4.6}$$

The difference from (4.5) 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; this also follows from (1.27).)

It is a well-known fact, see e.g. [30, 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 $\bigl((\xi_1,\dots,\xi_n) \mid \xi_1+\dots+\xi_n = n-1\bigr)$. Since (4.6) is invariant under cyclic shifts of $(d_1,\dots,d_n)$, it follows that, recalling (4.1),

$$F(\mathcal{T}_n) \overset{\mathrm{d}}{=} \left(\sum_{k=1}^{n} \sum_{i=1}^{n} f(\xi_i,\dots,\xi_{i+k-1 \bmod n}) \;\middle|\; S_n = n-1 \right), \tag{4.7}$$

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

4.3. 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,

$$f_k(T) := f(T)\cdot \mathbf{1}\{|T| = k\}. \tag{4.8}$$

Extended to integer sequences as in (4.4), this means that

$$f_k(d_1,\dots,d_n) = f(d_1,\dots,d_n)\cdot \mathbf{1}\{n = k\}. \tag{4.9}$$

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 $(d_1, \dots, d_n)$,

$$F_k(T) := F(T; f_k) = \sum_{i=1}^{n-k+1} f_k(d_i, \dots, d_{i+k-1}). \tag{4.10}$$

(We can also let the sum extend to n, wrapping around d_i as in (4.6).) Obviously,

$$f(T) = \sum_{k=1}^{\infty} f_k(T) \qquad \text{and} \qquad F(T) = \sum_{k=1}^{\infty} F_k(T) \tag{4.11}$$

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

4.4. It is well-known (see Otter [47], or [30, Theorem 15.5] and the further references given there) that for any $k \geqslant 1$,

$$\mathbb{P}(|\mathcal{T}| = n) = \frac{1}{n} \mathbb{P}(S_n = n - 1). \tag{4.12}$$

Hence, by (4.3), as $n \to \infty$ with $n \equiv 1 \pmod h$, see Kolchin [37],

$$\mathbb{P}(|\mathcal{T}| = n) \sim \frac{h}{\sqrt{2\pi\sigma^2}} n^{-3/2}. \tag{4.13}$$

In particular, $\mathbb{P}(|\mathcal{T}| = n) > 0$ for all large n with $n \equiv 1 \pmod h$,

We sometimes use the notation

$$\pi_n := \mathbb{P}(|\mathcal{T}| = n), \tag{4.14}$$

recalling (4.12)–(4.13).

5. Expectations

We begin the proof of Theorem 1.5 by calculating the expectation $\mathbb{E}\, F(\mathcal{T}_n)$, using (4.7) 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 [52, 53, 54], Swensen [50] and Janson [27]. In particular, the results (and methods) of Zabell [54] are closely related and partly overlapping (but the setup there is somewhat different).

We assume throughout the paper that $\xi = 1$ and $0 < \sigma^2 = \operatorname{Var} \xi < \infty$. For simplicity we also assume in some proofs in the sequel that the span h of the offspring distribution is 1, omitting the minor (and standard) modifications in the general case. All statements are true also for $h > 1$. (Note that when $h > 1$, the Galton–Watson tree $\mathcal{T}$ always has order $|\mathcal{T}| \equiv 1 \pmod h$, and thus we only consider $k, n \equiv 1 \pmod h$. The modifications when $h > 1$ consist in using the periodicity of the characteristic function $\varphi(t)$ and integrating only over $|t| < \pi/h$ in, for example, (5.11); we leave the details to the reader.) We assume further tacitly that n is so large that $\mathbb{P}(|\mathcal{T}| = n) > 0$, cf. (4.13).

By (4.7) and symmetry,

$$\mathbb{E}\,F(\mathcal{T}_n) = n\sum_{k=1}^{n}\mathbb{E}\big(f(\xi_1,\dots,\xi_k)\mid S_n=n-1\big). \tag{5.1}$$

We consider first the expectation of each $F_k(\mathcal{T}_n)$ separately, recalling (4.11). Note that each f_k is bounded, and thus trivially $\mathbb{E}\,|f_k(\mathcal{T})|<\infty$.

Lemma 5.1. *If $1\leqslant k\leqslant n$, then*

$$\mathbb{E}\,F_k(\mathcal{T}_n) = n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\,f_k(\mathcal{T}). \tag{5.2}$$

Proof. If $f_k(\xi_1,\dots,\xi_k)\neq 0$, then $S_k:=\xi_1+\dots+\xi_k=k-1$ by (1.27). Consequently, for every $n\geqslant k$, by (5.1), and the fact that the ξ_i are i.i.d.,

$$\begin{aligned}\mathbb{E}\,F_k(\mathcal{T}_n) &= n\,\mathbb{E}\big(f_k(\xi_1,\dots,\xi_k)\mid S_n=n-1\big)\\ &= n\frac{\mathbb{E}\big(f_k(\xi_1,\dots,\xi_k)\cdot\mathbf{1}\{S_n=n-1\}\big)}{\mathbb{P}(S_n=n-1)}\\ &= n\frac{\mathbb{E}\big(f_k(\xi_1,\dots,\xi_k)\cdot\mathbf{1}\{S_n-S_k=n-k\}\big)}{\mathbb{P}(S_n=n-1)}\\ &= n\frac{\mathbb{E}\,f_k(\xi_1,\dots,\xi_k)\cdot\mathbb{P}(S_n-S_k=n-k)}{\mathbb{P}(S_n=n-1)}\\ &= n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\,f_k(\xi_1,\dots,\xi_k).\end{aligned} \tag{5.3}$$

The result (5.2) now follows by Remark 3.2, which implies that, recalling again that by definition $f_k(T)=0$ unless $|T|=k$,

$$\mathbb{E}\,f_k(\xi_1,\dots,\xi_k)=\mathbb{E}\,f_k(\mathcal{T}). \tag{5.4}$$

For future use, we give also an alternative derivation of (5.4). By taking $n=k$ in (5.3), we obtain

$$\mathbb{E}\,F_k(\mathcal{T}_k)=\frac{k}{\mathbb{P}(S_k=k-1)}\,\mathbb{E}\,f_k(\xi_1,\dots,\xi_k). \tag{5.5}$$

(We may assume that $\mathbb{P}(|\mathcal{T}|=k)>0$, since the result trivially is true if $\mathbb{P}(|\mathcal{T}|=k)=0$, when $f_k(\mathcal{T})=0$ a.s.) Furthermore, since $f_k(T)=0$ unless $|T|=k$, (1.3) yields $F_k(\mathcal{T}_k)=f_k(\mathcal{T}_k)$ and

$$\mathbb{E}\,F_k(\mathcal{T}_k)=\mathbb{E}\,f_k(\mathcal{T}_k)=\mathbb{E}\big(f_k(\mathcal{T})\mid|\mathcal{T}|=k\big)=\frac{\mathbb{E}\,f_k(\mathcal{T})}{\mathbb{P}(|\mathcal{T}|=k)}. \tag{5.6}$$

Finally, recalling (4.12) (which also follows by taking $f_k(T)=1$ in (5.5)), (5.5)–(5.6) yield (5.4). ☐

Lemma 5.2. (i) *Uniformly for all k with $1\leqslant k\leqslant n/2$, as $n\to\infty$,*

$$\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}=1+O\left(\frac{k}{n}\right)+o\big(n^{-1/2}\big). \tag{5.7}$$

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

$$\frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} = O\left(\frac{n^{1/2}}{(n-k+1)^{1/2}}\right). \tag{5.8}$$

If ξ has a finite third moment, this follows easily from the refined local limit theorem in [48, Theorem VII.13]. Since we do not assume this, we have to work harder and take advantage of some cancellation.

Proof. (i): We let $\varphi(t) := \mathbb{E}\, e^{\mathrm{i}t\xi}$ be the characteristic function of ξ, and $\tilde{\varphi}(t) := e^{-\mathrm{i}t}\varphi(t)$ the characteristic function of the centred variable $\tilde{\xi} := \xi - \mathbb{E}\,\xi = \xi - 1$.

We begin with a standard estimate. Since $\mathbb{E}\,\tilde{\xi} = 0$ and $\operatorname{Var}(\tilde{\xi}) = \sigma^2 < \infty$, we have

$$\tilde{\varphi}(t) = 1 - \tfrac{1}{2}\sigma^2 t^2 + o(t^2) \qquad \text{as } |t| \to 0. \tag{5.9}$$

It follows that $|\varphi(t)| = |\tilde{\varphi}(t)| < e^{-\sigma^2 t^2/3}$ for $|t| \leqslant c_3$. Furthermore, assuming that ξ has span $h = 1$, $|\varphi(t)| < 1$ for $0 < |t| \leqslant \pi$, so by compactness, $|\varphi(t)| \leqslant 1 - c_4$ for $c_3 \leqslant |t| \leqslant \pi$. It follows that

$$|\varphi(t)| = |\tilde{\varphi}(t)| \leqslant e^{-c_5 t^2}, \qquad |t| \leqslant \pi. \tag{5.10}$$

To estimate the ratio in (5.7), we note first that by (4.3), it suffices to estimate the difference $\mathbb{P}(S_{n-k} = n-k) - \mathbb{P}(S_n = n-1)$. We do this in two steps.

First, consider the difference $\mathbb{P}(S_{n-k} = n-k) - \mathbb{P}(S_{n-1} = n-1)$. By Fourier inversion,

$$\begin{aligned}\mathbb{P}(S_{n-k} = n-k) - \mathbb{P}(S_{n-1} = n-1) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(\tilde{\varphi}^{n-k}(t) - \tilde{\varphi}^{n-1}(t)\Bigr)\,\mathrm{d}t \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(1 - \tilde{\varphi}^{k-1}(t)\Bigr)\tilde{\varphi}^{n-k}(t)\,\mathrm{d}t. \end{aligned} \tag{5.11}$$

By (5.9), $\tilde{\varphi}(t) = 1 + O(t^2)$, and thus, using also $|\tilde{\varphi}(t)| \leqslant 1$, for all $j \geqslant 0$ and $t \in \mathbb{R}$,

$$|\tilde{\varphi}^j(t) - 1| = O(jt^2). \tag{5.12}$$

Furthermore, by (5.10) and $k \leqslant n/2$, for $|t| \leqslant \pi$,

$$|\tilde{\varphi}^{n-k}(t)| \leqslant \exp(-c_5(n-k)t^2) \leqslant \exp(-c_6 n t^2). \tag{5.13}$$

Consequently, (5.11) yields

$$\begin{aligned}&\bigl|\mathbb{P}(S_{n-k} = n-k) - \mathbb{P}(S_{n-1} = n-1)\bigr| \\ &\qquad\ll \frac{1}{2\pi}\int_{-\pi}^{\pi} k t^2 e^{-c_6 n t^2}\,\mathrm{d}t \ll k n^{-3/2}.\end{aligned} \tag{5.14}$$

Next, consider $\mathbb{P}(S_{n-1}=n-1)-\mathbb{P}(S_n=n-1)$. By Fourier inversion,

$$\mathbb{P}(S_{n-1}=n-1)-\mathbb{P}(S_n=n-1)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-\mathrm{i}(n-1)t}\Bigl(\varphi^{n-1}(t)-\varphi^{n}(t)\Bigr)\,\mathrm{d}t$$
$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(1-\varphi(t)\Bigr)\tilde{\varphi}^{n-1}(t)\,\mathrm{d}t$$
$$=\frac{-\mathrm{i}}{2\pi}\int_{-\pi}^{\pi}t\tilde{\varphi}^{n-1}(t)\,\mathrm{d}t+\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(1+\mathrm{i}t-\varphi(t)\Bigr)\tilde{\varphi}^{n-1}(t)\,\mathrm{d}t. \tag{5.15}$$

Since $\varphi(t)=1+\mathrm{i}t+O(t^2)$, the second integral in (5.15) is $O(n^{-3/2})$ by the argument in (5.14). For the first integral, we make the change of variable $t=x/\sqrt{n}$:

$$\int_{-\pi}^{\pi}t\tilde{\varphi}^{n-1}(t)\,\mathrm{d}t=\frac{1}{n}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}x\tilde{\varphi}^{n-1}\left(\frac{x}{\sqrt{n}}\right)\mathrm{d}x. \tag{5.16}$$

We have $\tilde{\varphi}^{n-1}(x/\sqrt{n})\to e^{-\sigma^2x^2/2}$ as $n\to\infty$ for every x by (5.9), and thus by dominated convergence (justified by (5.10)),

$$\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}x\tilde{\varphi}^{n-1}\left(\frac{x}{\sqrt{n}}\right)\mathrm{d}x\to\int_{-\infty}^{\infty}xe^{-\sigma^2x^2/2}\,\mathrm{d}x=0. \tag{5.17}$$

Consequently, the expressions in (5.16) are $o(1/n)$, and (5.15) yields

$$\mathbb{P}(S_{n-1}=n-1)-\mathbb{P}(S_n=n-1)=o\bigl(n^{-1}\bigr). \tag{5.18}$$

This and (5.14) yield, together with (4.3),

$$\frac{|\,\mathbb{P}(S_{n-k}=n-k)-\mathbb{P}(S_n=n-1)|}{\mathbb{P}(S_n=n-1)}\ll\frac{kn^{-3/2}+o(n^{-1})}{n^{-1/2}}, \tag{5.19}$$

and the result follows.

(ii): We use (4.3) together with the similar estimate, also from Lemma 4.1, $\mathbb{P}(S_{n-k}=n-k)=O\bigl((n-k+1)^{-1/2}\bigr)$. ☐

Proof of Theorem 1.5(i). Let $a_k:=|\,\mathbb{E}\,f(\mathcal{T}_k)|=|\,\mathbb{E}\,f_k(\mathcal{T}_k)|$; thus by assumption $a_k\to0$ as $k\to\infty$. Moreover, by (5.6) and (4.13),

$$|\,\mathbb{E}\,f_k(\mathcal{T})|=|\,\mathbb{E}\,f_k(\mathcal{T}_k)|\,\mathbb{P}(|\mathcal{T}|=k)=a_k\,\mathbb{P}(|\mathcal{T}|=k)=O\bigl(a_kk^{-3/2}\bigr). \tag{5.20}$$

By (4.11) and Lemma 5.1,

$$\frac{1}{n}\,\mathbb{E}\,F(\mathcal{T}_n)-\mathbb{E}\,f(\mathcal{T})=\sum_{k=1}^{\infty}\Bigl(\frac{1}{n}\,\mathbb{E}\,F_k(\mathcal{T}_n)-\mathbb{E}\,f_k(\mathcal{T})\Bigr)$$
$$=\sum_{k=1}^{n}\left(\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}-1\right)\mathbb{E}\,f_k(\mathcal{T})-\sum_{k=n+1}^{\infty}\mathbb{E}\,f_k(\mathcal{T}). \tag{5.21}$$

We split the expression in (5.21) into three parts. First, for $k \leqslant n/2$ we use Lemma 5.2(i) and obtain, using (5.20) and $a_k \to 0$ as $k \to \infty$,

$$\sum_{k\leqslant n/2}\left|\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}-1\right|\bigl|\mathbb{E}\,f_k(\mathcal{T})\bigr| \ll \sum_{k\leqslant n/2}\left(\frac{k}{n}+o\bigl(n^{-1/2}\bigr)\right)a_k k^{-3/2}$$
$$\ll n^{-1}\sum_{k\leqslant n}a_k k^{-1/2}+o\bigl(n^{-1/2}\bigr)=o\bigl(n^{-1/2}\bigr). \tag{5.22}$$

For $n/2 < k \leqslant n$, we use Lemma 5.2(ii), yielding

$$\sum_{n/2<k\leqslant n}\left|\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}-1\right|\bigl|\mathbb{E}\,f_k(\mathcal{T})\bigr| \ll \sum_{n/2<k\leqslant n}\frac{n^{1/2}}{(n-k+1)^{1/2}}a_k k^{-3/2}$$
$$\ll n^{-1}\max_{k\geqslant n/2}a_k\sum_{n/2<k\leqslant n}\frac{1}{(n-k+1)^{1/2}}=o\bigl(n^{-1/2}\bigr). \tag{5.23}$$

Finally, for $k > n$ we have by (5.20)

$$\sum_{k>n}\bigl|\mathbb{E}\,f_k(\mathcal{T})\bigr| \ll \max_{k>n}a_k\sum_{k>n}k^{-3/2}=o(1)\cdot\sum_{k>n}k^{-3/2}=o\bigl(n^{-1/2}\bigr). \tag{5.24}$$

The result follows by (5.21)–(5.24). ∎

Remark 5.3. Trivial modifications in the proof above show that if $\mathbb{E}\,|f(\mathcal{T})| < \infty$ and $|\,\mathbb{E}\,f(\mathcal{T}_k)| = o(k^{1/2})$, then $\mathbb{E}\,F(\mathcal{T}_n) = n\mu + o(n)$, so (1.10) holds. Moreover, the quenched version (1.12) holds too; this follows easily from (1.10) together with (1.12) applied to truncations of f. (We omit the details.)

If we assume further moment conditions on ξ, we can improve the error term in (5.7) and thus in (1.13). (Cf. Zabell [54, Theorem 4].)

Lemma 5.4. *If* $\mathbb{E}\,\xi^{2+\delta} < \infty$ *with* $0 < \delta \leqslant 1$, *then, uniformly for all* k *and* n *with* $1 \leqslant k \leqslant n/2$,

$$\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)} = 1+O\left(\frac{k}{n}\right)+O\bigl(n^{-(1+\delta)/2}\bigr). \tag{5.25}$$

Proof. This follows by minor modifications in the proof of Lemma 5.2(i). We now have

$$\tilde{\varphi}(t) = 1-\tfrac{1}{2}\sigma^2t^2+O(|t|^{2+\delta}) \tag{5.26}$$

which leads to

$$x\tilde{\varphi}^{n-1}\left(\frac{x}{\sqrt{n}}\right) = xe^{-\sigma^2x^2/2}\left(1+O\left(\frac{|x|^{2+\delta}}{n^{\delta/2}}\right)+O\left(\frac{x^2}{n}\right)\right), \tag{5.27}$$

for $|x| \leqslant n^{\delta/6}$, at least. It follows (using (5.13) for $x > n^{\delta/6}$) that the first integral in (5.17) is $O\bigl(n^{-\delta/2}\bigr)$. The rest is as before. ∎

Theorem 5.5. *Suppose, in addition to the assumptions of Theorem 1.5*(i)*, that* $0 < \delta < 1$ *and that* $\mathbb{E}\,\xi^{2+\delta} < \infty$ *and* $\mathbb{E}\, f(\mathcal{T}_n) = O(n^{-\delta/2})$. *Then*

$$\mathbb{E}\, F(\mathcal{T}_n) = n\mu + O\big(n^{(1-\delta)/2}\big). \tag{5.28}$$

Similarly, if $\mathbb{E}\,\xi^3 < \infty$, $\mathbb{E}\, f(\mathcal{T}_n) = O(n^{-1/2})$ *and* $\sum_{n=1}^{\infty} |\,\mathbb{E}\, f(\mathcal{T}_n)|n^{-1/2} < \infty$, *then*

$$\mathbb{E}\, F(\mathcal{T}_n) = n\mu + O(1). \tag{5.29}$$

Proof. As the proof of Theorem 1.5(i) above, using (5.25) and the assumptions on $\mathbb{E}\, f(\mathcal{T}_n)$. We omit the details. □

Remark 5.6. In fact, if $\mathbb{E}\,\xi^3 < \infty$, by including the next terms explicitly in the calculations in the proof of Lemma 5.2(i), it is easily shown that for every fixed k, as $n \to \infty$,

$$\frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} = 1 + \frac{1}{2n}\big(k + \sigma^{-2} - \varkappa_3\sigma^{-4}\big) + o\big(n^{-1}\big), \tag{5.30}$$

where $\varkappa_3 = \mathbb{E}(\xi - \mathbb{E}\,\xi)^3$ is the third cumulant of ξ. (If $\mathbb{E}\,\xi^4 < \infty$, this also follows easily from [48, Theorem VII.13].) Hence, if for simplicity f has finite support, (5.21) yields, with $\mu := \mathbb{E}\, f(\mathcal{T})$ as above,

$$\mathbb{E}\, F(\mathcal{T}_n) = n\mu + \tfrac{1}{2}\,\mathbb{E}\big(|\mathcal{T}| f(\mathcal{T})\big) + \tfrac{1}{2}\big(\sigma^{-2} - \varkappa_3\sigma^{-4}\big)\mu + o(1). \tag{5.31}$$

We leave it to the reader to find more general conditions on f for (5.31) to hold.

The following example (adapted from [54]) shows that the sharper results in Lemma 5.4 and Theorem 5.5 do not hold without the extra moment assumption on ξ.

Example 5.7. This is a discrete version of [54, Examples 5–6]. Consider, as in Example 2.1, $f(T) = \mathbf{1}\{|T| = 1\}$. Suppose that $p_k = \mathbb{P}(\xi = k) = ak^{-\alpha}$ for $k \geqslant 2$ for some $a > 0$ and $\alpha \in (3,4)$. (With p_0, p_1 adjusted so that $\sum_k p_k = 1$ and $\mathbb{E}\,\xi = 1$; this is obviously possible if a is small.) Then $\mathbb{E}\,\xi^r < \infty \iff r < \alpha - 1$; in particular, $\mathbb{E}\,\xi^2 < \infty$ but $\mathbb{E}\,\xi^3 = \infty$. It can be verified that

$$\varphi(t) = 1 + \mathrm{i}t - \tfrac{1}{2}\,\mathbb{E}\,\xi^2 t^2 + a\Gamma(1-\alpha)(-\mathrm{i}t)^{\alpha-1} + O(t^3), \tag{5.32}$$

see e.g. [46, 25.12.12] or [21, Theorem VI.7]. It follows that, for $|t| \leqslant c_7$,

$$\log\tilde{\varphi}(t) = -\tfrac{1}{2}\sigma^2 t^2 + a\Gamma(1-\alpha)(-\mathrm{i}t)^{\alpha-1} + O(t^3), \tag{5.33}$$

and hence, for $|x| \leqslant n^{(\alpha-3)/6}$, by a simple calculation,

$$\tilde{\varphi}^{n-1}\left(\frac{x}{\sqrt{n}}\right) = e^{-\sigma^2 x^2/2}\Bigg(1 + a\Gamma(1-\alpha)(-\mathrm{i}x)^{\alpha-1} n^{(3-\alpha)/2} + O\Big(\frac{x^2 + |x|^3}{n^{1/2}}\Big) + O\Big(\frac{|x|^{2\alpha-2}}{n^{\alpha-3}}\Big)\Bigg).$$

Using this in (5.16), it is easy to obtain

$$\begin{aligned}
&-\mathrm{i}\int_{-\pi}^{\pi} t\tilde{\varphi}^{n-1}(t)\,\mathrm{d}t \\
&\quad= \int_{-\infty}^{\infty} e^{-\sigma^2x^2/2}\Bigl(\frac{-\mathrm{i}x}{n} + a\Gamma(1-\alpha)(-\mathrm{i}x)^{\alpha} n^{(1-\alpha)/2}\Bigr)\,\mathrm{d}x + O\bigl(n^{-3/2}+n^{2-\alpha}\bigr) \\
&\quad= 2a\Gamma(1-\alpha)\cos\frac{\pi\alpha}{2} n^{(1-\alpha)/2}\int_0^{\infty} x^{\alpha}e^{-\sigma^2x^2/2}\,\mathrm{d}x + O\bigl(n^{-3/2}+n^{2-\alpha}\bigr) \\
&\quad= bn^{(1-\alpha)/2} + o\bigl(n^{(1-\alpha)/2}\bigr)
\end{aligned}$$

for some $b\neq 0$. Using this, instead of (5.16)–(5.17), in the proof of Lemma 5.2, leads to the estimate, for some $c\neq 0$,

$$\frac{\mathbb{P}(S_{n-1}=n-1)}{\mathbb{P}(S_n=n-1)} = 1 + cn^{1-\alpha/2} + o\bigl(n^{1-\alpha/2}\bigr), \tag{5.34}$$

and thus by Lemma 5.1

$$\mathbb{E}\,F(\mathcal{T}_n) = np_0 + cp_0n^{2-\alpha/2} + o\bigl(n^{2-\alpha/2}\bigr), \tag{5.35}$$

showing that without further assumptions, the error term $o(n^{-1/2})$ in Theorem 1.5(i) is essentially best possible.

We can also take $\alpha=4$ in this example; then

$$\varphi(t) = 1 + \mathrm{i}t - \tfrac{1}{2}\,\mathbb{E}\,\xi^2t^2 + \tfrac{\mathrm{i}}{6}at^3\log|t| + O(t^3), \tag{5.36}$$

and similar calculations lead to, for some $c\neq 0$,

$$\mathbb{E}\,F(\mathcal{T}_n) = np_0 + cp_0\log n + O(1). \tag{5.37}$$

We end this section with a result on the expectation $\mathbb{E}\,f(\mathcal{T}_n)$ in the case of a local functional f. We first state an estimate similar to Lemma 5.2 (but somewhat coarser and simpler); it can be refined but the present version is enough for our needs. (If ξ has a finite third moment, it too, and more, follows easily from the refined local limit theorem in [48, Theorem VII.13].)

Lemma 5.8. *For any integers $w,z\geqslant 0$ with $z\leqslant n/2$,*

$$\frac{\mathbb{P}(S_{n-z}=n-z-w)}{\mathbb{P}(S_n=n-1)} = 1 + O\Bigl(\frac{w+z+1}{n^{1/2}}\Bigr). \tag{5.38}$$

Proof. We argue as in the proof of Lemma 5.2 and obtain, recalling (5.12) and (5.13),

$$\mathbb{P}(S_{n-z}=n-z-w)-\mathbb{P}(S_n=n-1)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(\tilde\varphi^{n-z}(t)e^{\mathrm{i}wt}-\tilde\varphi^{n}(t)e^{\mathrm{i}t}\Bigr)\,\mathrm{d}t$$
$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-z}(t)\Bigl(e^{\mathrm{i}wt}-\tilde\varphi^{z}(t)e^{\mathrm{i}t}\Bigr)\,\mathrm{d}t$$
$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-z}(t)\Bigl(1+O\bigl(w|t|\bigr)-\bigl(1+O\bigl(zt^2\bigr)+O\bigl(|t|\bigr)\bigr)\Bigr)\,\mathrm{d}t$$
$$=\int_{-\pi}^{\pi}e^{-c_6nt^2}O\bigl(w|t|+z|t|+|t|\bigr)\,\mathrm{d}t=O\Bigl(\frac{w+z+1}{n}\Bigr).$$

The result follows by division by $\mathbb{P}(S_n=n-1)$, using (4.3). □

Let $\hat{\mathcal{T}}$ be the *size-biased* Galton–Watson tree defined by Kesten [36], see also Aldous [3], Aldous and Pitman [5], Lyons, Pemantle and Peres [41] and Janson [30]; this is a random infinite tree, whose distribution can be described in terms of the truncations $\hat{\mathcal{T}}^{(M)}$ by

$$\mathbb{P}(\hat{\mathcal{T}}^{(M)}=T)=w_M(T)\,\mathbb{P}(\mathcal{T}^{(M)}=T),\qquad T\in\mathfrak{T},\tag{5.39}$$

where $w_M(T)$ denotes the number of nodes of depth M in T. Then $\mathcal{T}_n\xrightarrow{\mathrm{d}}\hat{\mathcal{T}}$ as $n\to\infty$ in the appropriate (local) topology, as shown by Kennedy [35] and Aldous and Pitman [5], see also Janson [30] for details and generalizations. This means that $\mathcal{T}_n^{(M)}\xrightarrow{\mathrm{d}}\hat{\mathcal{T}}^{(M)}$ for every fixed M. If f is a local functional with cut-off M, so $f(T)=f(T^{(M)})$ for every finite tree T, then we define f also for the infinite tree $\hat{\mathcal{T}}$ by $f(\hat{\mathcal{T}}):=f(\hat{\mathcal{T}}^{(M)})$. It follows that $f(\mathcal{T}_n)=f(\mathcal{T}_n^{(M)})\xrightarrow{\mathrm{d}}f(\hat{\mathcal{T}}^{(M)})=f(\hat{\mathcal{T}})$, and thus, if f furthermore is bounded, that $\mathbb{E}\,f(\mathcal{T}_n)\to\mathbb{E}\,f(\hat{\mathcal{T}})$. We establish an upper bound on the rate of this convergence. (Note that we do not impose any further moment condition on ξ beyond finite variance.)

Lemma 5.9. *If $f(T)$ is a bounded local functional on $\mathfrak{T}$, then*

$$\mathbb{E}\,f(\mathcal{T}_n)=\mathbb{E}\,f(\hat{\mathcal{T}})+O\bigl(n^{-1/2}\bigr).\tag{5.40}$$

Proof. Let as above $M\geqslant 1$ be the cut-off of f. Let T be a tree with height $\leqslant M$ and condition on the event that $\mathcal{T}^{(M)}=T$. Then the rest of the tree, more precisely $\mathcal{T}\setminus\mathcal{T}^{(M-1)}$, is a random forest consisting of $w:=w_M(T)$ independent copies of $\mathcal{T}$; denote this random forest by $\mathcal{F}_w$. By an extension of (4.12) due to Dwass [16], see also Kemperman [33, 34] and Pitman [49],

$$\mathbb{P}(|\mathcal{F}_w|=n)=\frac{w}{n}\,\mathbb{P}(S_n=n-w).\tag{5.41}$$

Let $z_k=z_k(T):=\sum_{j=0}^{k}w_j(T)$, the number of nodes in the first k generations of T. Let further $\pi_T^{(M)}:=\mathbb{P}\bigl(\mathcal{T}^{(M)}=T\bigr)$ and, using (5.39), $\hat\pi_T^{(M)}:=$

$\mathbb{P}\big(\hat{\mathcal{T}}^{(M)} = T\big) = w\pi_T^{(M)}$. It follows that, for $n \geqslant z_M$,

$$\begin{aligned}\mathbb{P}\big(\mathcal{T}^{(M)} = T \text{ and } |\mathcal{T}| = n\big) &= \mathbb{P}\big(\mathcal{T}^{(M)} = T\big)\,\mathbb{P}\big(|\mathcal{F}_w| = n - z_{M-1}\big)\\ &= \pi_T^{(M)} \frac{w}{n - z_{M-1}}\,\mathbb{P}\big(S_{n-z_{M-1}} = n - z_{M-1} - w\big)\\ &= \hat{\pi}_T^{(M)} \frac{\mathbb{P}\big(S_{n-z_{M-1}} = n - z_{M-1} - w\big)}{n - z_{M-1}}.\end{aligned}$$

Hence, recalling (4.12),

$$\begin{aligned}\mathbb{P}(\mathcal{T}_n^{(M)} = T) &= \mathbb{P}\big(\mathcal{T}^{(M)} = T \mid |\mathcal{T}| = n\big)\\ &= \frac{\mathbb{P}\big(\mathcal{T}^{(M)} = T \text{ and } |\mathcal{T}| = n\big)}{\mathbb{P}(|\mathcal{T}| = n)}\\ &= \hat{\pi}_T^{(M)} \frac{n}{n - z_{M-1}} \frac{\mathbb{P}\big(S_{n-z_{M-1}} = n - z_{M-1} - w\big)}{\mathbb{P}\big(S_n = n - 1\big)}\end{aligned}$$

If $z_{M-1} \leqslant n/2$, we thus obtain by Lemma 5.8,

$$\mathbb{P}(\mathcal{T}_n^{(M)} = T) = \hat{\pi}_T^{(M)}\Big(1 + O\Big(\frac{z_{M-1} + w}{n^{1/2}}\Big)\Big) = \hat{\pi}_T^{(M)}\Big(1 + O\Big(\frac{z_M}{n^{1/2}}\Big)\Big). \tag{5.42}$$

(Incidentally, this proves $\mathbb{P}(\mathcal{T}_n^{(M)} = T) \to \hat{\pi}_T^{(M)} = \mathbb{P}(\hat{\mathcal{T}}^{(M)} = T)$, i.e., $\mathcal{T}_n^{(M)} \xrightarrow{\mathrm{d}} \hat{\mathcal{T}}^{(M)}$ as asserted above.)

Consequently, since f is bounded, using (5.42) when $|T| \leqslant n/2$,

$$\begin{aligned}&\big|\mathbb{E} f(\mathcal{T}_n) - \mathbb{E} f(\hat{\mathcal{T}})\big| = \big|\mathbb{E} f(\mathcal{T}_n^{(M)}) - \mathbb{E}(\hat{\mathcal{T}}^{(M)})\big|\\ &\quad= \Big|\sum_T f(T)\,\mathbb{P}(\mathcal{T}_n^{(M)} = T) - \sum_T f(T)\,\mathbb{P}(\hat{\mathcal{T}}^{(M)} = T)\Big|\\ &\quad\ll \sum_T \big|\mathbb{P}(\mathcal{T}_n^{(M)} = T) - \mathbb{P}(\hat{\mathcal{T}}^{(M)} = T)\big|\\ &\quad\ll \sum_T \mathbb{P}(\hat{\mathcal{T}}^{(M)} = T)\frac{z_M(T)}{n^{1/2}} + \sum_{|T| > n/2}\big(\mathbb{P}(\mathcal{T}_n^{(M)} = T) + \mathbb{P}(\hat{\mathcal{T}}^{(M)} = T)\big)\\ &\quad= n^{-1/2}\,\mathbb{E}\,|\hat{\mathcal{T}}^{(M)}| + \mathbb{P}\big(|\mathcal{T}_n^{(M)}| > n/2\big) + \mathbb{P}\big(|\hat{\mathcal{T}}^{(M)}| > n/2\big)\\ &\quad\ll n^{-1/2}\,\mathbb{E}\,|\hat{\mathcal{T}}^{(M)}| + n^{-1}\,\mathbb{E}\,|\mathcal{T}_n^{(M)}|,\end{aligned} \tag{5.43}$$

using Markov's inequality at the final step. Finally, we observe that

$$\mathbb{E}\,|\mathcal{T}_n^{(M)}| = \sum_{k=0}^{M} \mathbb{E}\, w_k(\mathcal{T}_n) = O(1) \tag{5.44}$$

since $\mathbb{E}\, w_j(\mathcal{T}_k) = O(j)$ for each j, see [44] (assuming that ξ has an exponential moment) and [29, Theorem 1.13] (general ξ with $\mathbb{E}\,\xi^2 < \infty$; in fact it is shown that the estimate holds uniformly in j); see also [1] for further results.

Similarly, or as a consequence,

$$\mathbb{E}\,|\hat{\mathcal{T}}^{(M)}| = \sum_{k=0}^{M} \mathbb{E}\, w_k(\hat{\mathcal{T}}) = \sum_{k=0}^{M} \mathbb{E}\, w_k(\mathcal{T})^2 < \infty. \tag{5.45}$$

Hence (5.43) yields the estimate $O(n^{-1/2})$. □

6. Variances and covariances

We next consider the variance of $F(\mathcal{T}_n)$. As in Section 5, 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, corresponding to Lemma 5.1.

Lemma 6.1. *If $m \leqslant k$ and $n \geqslant k+m-1$, then*

$$\begin{aligned}\operatorname{Cov}(F_k(\mathcal{T}_n), F_m(\mathcal{T}_n)) &= n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\big(f_k(\mathcal{T})F_m(\mathcal{T})\big)\\ &- n(k+m-1)\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\cdot\frac{\mathbb{P}(S_{n-m}=n-m)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T})\\ &+ n(n-k-m+1)\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T})\cdot\\ &\Bigl(\frac{\mathbb{P}(S_{n-k-m}=n-k-m+1)}{\mathbb{P}(S_n=n-1)} - \frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\frac{\mathbb{P}(S_{n-m}=n-m)}{\mathbb{P}(S_n=n-1)}\Bigr).\end{aligned} \tag{6.1}$$

If $m \leqslant k \leqslant n \leqslant k+m$, we have instead

$$\begin{aligned}\operatorname{Cov}(F_k(\mathcal{T}_n), F_m(\mathcal{T}_n)) &= n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\big(f_k(\mathcal{T})F_m(\mathcal{T})\big)\\ &- n^2\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\cdot\frac{\mathbb{P}(S_{n-m}=n-m)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T}).\end{aligned} \tag{6.2}$$

Note that by (5.6) and (4.12),

$$\mathbb{E}\, f_k(\mathcal{T}) = \mathbb{P}(|\mathcal{T}|=k)\,\mathbb{E}\, f_k(\mathcal{T}_k) = \frac{\mathbb{P}(S_k=k-1)}{k}\,\mathbb{E}\, f_k(\mathcal{T}_k) \tag{6.3}$$

and similarly (again because $f_k(\mathcal{T})=0$ unless $|\mathcal{T}|=k$)

$$\mathbb{E}\big(f_k(\mathcal{T})F_m(\mathcal{T})\big) = \mathbb{P}(|\mathcal{T}|=k)\,\mathbb{E}\big(f_k(\mathcal{T}_k)F_m(\mathcal{T}_k)\big). \tag{6.4}$$

Proof. Note first that for $n=k+m-1$ and $n=k+m$, the formulas (6.1) and (6.2) agree, in the latter case because $\mathbb{P}(S_{n-k-m}=n-k-m+1) = P(S_0=1)=0$. Hence it suffices to prove (6.1) for $n \geqslant k+m$ and (6.2) for $k \leqslant n \leqslant k+m-1$.

By (4.7) and symmetry,

$$\begin{aligned}&\mathbb{E}\big(F_k(\mathcal{T}_n)F_m(\mathcal{T}_n)\big)\\ &\quad= n\sum_{j=0}^{n-1}\mathbb{E}\big(f_k(\xi_1,\dots,\xi_k)f_m(\xi_{j+1},\dots,\xi_{j+m \bmod n}) \mid S_n=n-1\big)\end{aligned} \tag{6.5}$$

Consider first the terms with $0 \leqslant j \leqslant k-m$; these are the terms with $\{j+1,\dots,j+m\} \subseteq \{1,\dots,k\}$, and we see from (4.10) that if $(\xi_1,\dots,\xi_k)$ is the degree sequence of a tree, then

$$\sum_{j=0}^{k-m} f_m(\xi_{j+1},\dots,\xi_{j+m}) = F_m(\xi_1,\dots,\xi_k). \tag{6.6}$$

Hence, if we define $g(T) := f_k(T)F_m(T)$, $T \in \mathfrak{T}$, and use (5.1) and Lemma 5.1 (or its proof), noting that $g_k = g$,

$$\begin{aligned}
\sum_{j=0}^{k-m} &\mathbb{E}\bigl(f_k(\xi_1,\dots,\xi_k)f_m(\xi_{j+1},\dots,\xi_{j+m}) \mid S_n = n-1\bigr)\\
&= \mathbb{E}\bigl(f_k(\xi_1,\dots,\xi_k)F_m(\xi_1,\dots,\xi_k) \mid S_n = n-1\bigr)\\
&= \mathbb{E}\bigl(g_k(\xi_1,\dots,\xi_k) \mid S_n = n-1\bigr) = \frac{1}{n}\,\mathbb{E}\,G_k(\mathcal{T}_n)\\
&= \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)}\,\mathbb{E}\,g_k(\mathcal{T}).
\end{aligned} \tag{6.7}$$

This yields the first term on the right-hand side of (6.1) and (6.2).

Next, two subtrees of a tree are either disjoint or one is a subtree of the other. Hence, if $k-m < j < k$ so the index sets $\{1,\dots,k\}$ and $\{j+1,\dots,j+m\}$ overlap partly, $(\xi_1,\dots,\xi_k)$ and $(\xi_{j+1},\dots,\xi_{j+m})$ cannot both be degree sequences of trees (this can also be seen algebraically from (1.27)), and thus

$$f_k(\xi_1,\dots,\xi_k)f_m(\xi_{j+1},\dots,\xi_{j+m}) = 0.$$

Hence these terms in the sum in (6.5) vanish. The same holds if $n-m < j < n$, with indices taken modulo n, when the index sets again overlap partly (on the other side).

Finally, if $k \leqslant j \leqslant n-m$ (and thus $n \geqslant k+m$), the index sets $\{1,\dots,k\}$ and $\{j+1,\dots,j+m\}$ are disjoint. By symmetry, the expectation in (6.5) is the same for all j in this range, so we may assume $j = k$, noting that this term appears $n-k-m+1$ times. Arguing as in (5.3), and using (5.4),

$$\begin{aligned}
\mathbb{E}&\bigl(f_k(\xi_1,\dots,\xi_k)f_m(\xi_{k+1},\dots,\xi_{k+m}) \mid S_n = n-1\bigr)\\
&= \frac{\mathbb{E}\bigl(f_k(\xi_1,\dots,\xi_k)f_m(\xi_{k+1},\dots,\xi_{k+m})\mathbf{1}\{S_n - S_{k+m} = n-k-m+1\}\bigr)}{\mathbb{P}(S_n = n-1)}\\
&= \frac{\mathbb{E}\,f_k(\xi_1,\dots,\xi_k)\,\mathbb{E}\,f_m(\xi_{k+1},\dots,\xi_{k+m})\,\mathbb{P}(S_n - S_{k+m} = n-k-m+1)}{\mathbb{P}(S_n = n-1)}\\
&= \frac{\mathbb{P}(S_{n-k-m} = n-k-m+1)}{\mathbb{P}(S_n = n-1)}\mathbb{E}\,f_k(\mathcal{T})\,\mathbb{E}\,f_m(\mathcal{T}).
\end{aligned} \tag{6.8}$$

The results (6.1) and (6.2) now follow from (6.5)–(6.8), subtracting the product $\mathbb{E}\,F_k(\mathcal{T}_n)\,\mathbb{E}\,F_m(\mathcal{T}_n)$ which is given by two applications of (5.2). (Note that in (6.1), this is split into two terms.) ∎

We next estimate one of the factors in (6.1), where there typically is a lot of cancellation.

Lemma 6.2. (i) *As $n \to \infty$, uniformly for all $k \geqslant 0$ and $m \geqslant 0$ with $k + m \leqslant n/2$,*

$$\frac{\mathbb{P}(S_{n-k-m} = n-k-m+1)}{\mathbb{P}(S_n = n-1)} - \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)}\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} = -\frac{1}{n\sigma^2} + o\Bigl(\frac{1}{n}\Bigr) + O\Bigl(\frac{k+m}{n^{3/2}}\Bigr) + O\Bigl(\frac{km}{n^2}\Bigr). \tag{6.9}$$

(ii) *For all $n \geqslant 1$, $k \geqslant 0$ and $m \geqslant 0$ with $n/2 \leqslant k + m \leqslant n$,*

$$\frac{\mathbb{P}(S_{n-k-m} = n-k-m+1)}{\mathbb{P}(S_n = n-1)} - \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)}\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} = O\biggl(\frac{\min(k,m)\, n^{1/2}}{(n-k-m+1)^{3/2}}\biggr) + O\biggl(\frac{n^{1/2}}{n-k-m+1}\biggr). \tag{6.10}$$

Proof. (i): By multiplying (6.9) by $\mathbb{P}(S_n = n-1)^2$ and using (4.3), we see that (6.9) is equivalent to (assuming $h = 1$ for simplicity)

$$\mathbb{P}(S_{n-k-m} = n-k-m+1)\mathbb{P}(S_n = n-1) - \mathbb{P}(S_{n-k} = n-k)\,\mathbb{P}(S_{n-m} = n-m) = -\frac{1}{2\pi n^2\sigma^4} + o\Bigl(\frac{1}{n^2}\Bigr) + O\Bigl(\frac{k+m}{n^{5/2}}\Bigr) + O\Bigl(\frac{km}{n^3}\Bigr). \tag{6.11}$$

To prove this, we first obtain by Fourier inversion, recalling $\tilde\varphi(t) := e^{-\mathrm{i}t}\varphi(t)$,

$$\begin{aligned}
&\mathbb{P}(S_{n-k-m} = n-k-m+1)\mathbb{P}(S_n = n-1) - \mathbb{P}(S_{n-k} = n-k)\,\mathbb{P}(S_{n-m} = n-m)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-k-m}(t)e^{-\mathrm{i}t}\,\mathrm{d}t \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n}(u)e^{\mathrm{i}u}\,\mathrm{d}u - \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-k}(t)\,\mathrm{d}t \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-m}(u)\,\mathrm{d}u\\
&= \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-k-m}(t)e^{-\mathrm{i}t}\tilde\varphi^{n-k-m}(u)e^{-\mathrm{i}u}\times\\
&\qquad\Bigl(\tilde\varphi^{k}(t)e^{\mathrm{i}t} - \tilde\varphi^{k}(u)e^{\mathrm{i}u}\Bigr)\Bigl(\tilde\varphi^{m}(t)e^{\mathrm{i}t} - \tilde\varphi^{m}(u)e^{\mathrm{i}u}\Bigr)\,\mathrm{d}t\,\mathrm{d}u.
\end{aligned} \tag{6.12}$$

For all $k \geqslant 0$, we have by (5.12) $|\tilde\varphi^k(t)e^{\mathrm{i}t} - 1| = O(|t| + kt^2)$ and thus

$$|\tilde\varphi^k(t)e^{\mathrm{i}t} - \tilde\varphi^k(u)e^{\mathrm{i}u}| = O(|t| + kt^2 + |u| + ku^2); \tag{6.13}$$

similarly,

$$\Bigl|\bigl(\tilde\varphi^m(t)e^{\mathrm{i}t} - \tilde\varphi^m(u)e^{\mathrm{i}u}\bigr) - \tilde\varphi^m(t)\tilde\varphi^m(u)\bigl(e^{\mathrm{i}t} - e^{\mathrm{i}u}\bigr)\Bigr| = O(mt^2 + mu^2). \tag{6.14}$$

Furthermore, if $k + m \leqslant n/2$ and $|t| \leqslant \pi$, then by (5.10), or (5.13),

$$\bigl|\tilde\varphi^{n-k-m}(t)\bigr| \leqslant \exp(-c_5(n-k-m)t^2) \leqslant \exp(-c_6 nt^2). \tag{6.15}$$

Denote the left-hand side of (6.12) by $\Delta_{k,m}$. If we replace the factor $\tilde\varphi^m(t)e^{\mathrm{i}t}-\tilde\varphi^m(u)e^{\mathrm{i}u}$ in the right-hand side by $\tilde\varphi^m(t)\tilde\varphi^m(u)\bigl(e^{\mathrm{i}t}-e^{\mathrm{i}u}\bigr)$, we obtain after cancellation $\Delta_{k,0}$. Using (6.13)–(6.15) to estimate the resulting error we obtain

$$\begin{aligned}
&\bigl|\Delta_{k,m}-\Delta_{k.0}\bigr|\\
&\quad\ll\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{-c_6nt^2-c_6nu^2}\bigl(|t|+|u|+kt^2+ku^2\bigr)\bigl(mt^2+mu^2\bigr)\,\mathrm{d}t\,\mathrm{d}u\\
&\quad\ll mn^{-5/2}+kmn^{-3}. \qquad (6.16)
\end{aligned}$$

This is covered by the error terms in (6.11), and thus it suffices to prove (6.11) for $m=0$. By symmetry, we also obtain $|\Delta_{k,0}-\Delta_{0,0}|\ll kn^{-5/2}$, so it suffices to prove (6.11) in the case $k=m=0$.

In that case, (6.12) yields

$$\begin{aligned}
\Delta_{0,0}&=\frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\tilde\varphi^n(t)e^{-\mathrm{i}t}\tilde\varphi^n(u)e^{-\mathrm{i}u}\bigl(e^{\mathrm{i}t}-e^{\mathrm{i}u}\bigr)^2\,\mathrm{d}t\,\mathrm{d}u\\
&=\frac{1}{8\pi^2n^2}\int_{-\pi\sqrt n}^{\pi\sqrt n}\int_{-\pi\sqrt n}^{\pi\sqrt n}\tilde\varphi^n\Bigl(\frac{x}{\sqrt n}\Bigr)e^{-\mathrm{i}x/\sqrt n}\tilde\varphi^n\Bigl(\frac{y}{\sqrt n}\Bigr)e^{-\mathrm{i}y/\sqrt n}\times\\
&\qquad\qquad\Bigl(\sqrt n\bigl(e^{\mathrm{i}x/\sqrt n}-e^{\mathrm{i}y/\sqrt n}\bigr)\Bigr)^2\,\mathrm{d}x\,\mathrm{d}y, \qquad (6.17)
\end{aligned}$$

and it follows by dominated convergence, using (5.9) and (5.10), that as $n\to\infty$,

$$n^2\Delta_{0,0}\to-\frac{1}{8\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\sigma^2x^2/2-\sigma^2y^2/2}(x-y)^2\,\mathrm{d}x\,\mathrm{d}y=-\frac{4\pi}{8\pi^2\sigma^4}. \qquad (6.18)$$

This shows (6.11) in the special case $k=m=0$, and thus by the estimate (6.16) for all k and m with $k+m\leqslant n/2$, which completes the proof of (6.11) and (6.9).

(ii): To prove (6.10), we first observe that we may assume $m\leqslant k$ by symmetry. In this case $n\geqslant k\geqslant n/4$. We again multiply by $\mathbb{P}(S_n=n-1)^2$ and use (6.12). We now use the estimate, by (5.10), for $|t|,|u|\leqslant\pi$,

$$|\tilde\varphi^k(t)e^{\mathrm{i}t}-\tilde\varphi^k(u)e^{\mathrm{i}u}|\leqslant|\tilde\varphi^k(t)|+|\tilde\varphi^k(u)|\leqslant\exp(-c_5kt^2)+\exp(-c_5ku^2). \qquad (6.19)$$

Using this and (6.13) (with k replaced by m) in (6.12) we obtain, using $k\gg n$ and symmetry,

$$\begin{aligned}
\Delta_{k,m}&\ll\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{-c_5(n-k-m)t^2-c_5(n-k-m)u^2}\Bigl(e^{-c_5kt^2}+e^{-c_5ku^2}\Bigr)\times\\
&\qquad\qquad\bigl(|t|+mt^2+|u|+mu^2)\bigr)\,\mathrm{d}t\,\mathrm{d}u\\
&\leqslant 2e^{c_3\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-c_5kt^2-c_5(n-k-m+1)u^2}\bigl(|t|+mt^2+|u|+mu^2)\bigr)\,\mathrm{d}t\,\mathrm{d}u\\
&\ll\frac{1}{n^{1/2}(n-k-m+1)}+\frac{m}{n^{1/2}(n-k-m+1)^{3/2}}.
\end{aligned}$$

The result (6.10) follows by dividing by $\mathbb{P}(S_n = n-1)^2 \gg n^{-1}$. $\square$

In particular, we obtain a simple asymptotic result for fixed k and m.

Lemma 6.3. *For any fixed k and m with $k \geqslant m$, as $n \to \infty$,*

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

Proof. This now follows from Lemma 6.1. After division by n, the two first terms on the right-hand side of (6.1) converge to $\mathbb{E}\bigl(f_k(\mathcal{T})F_m(\mathcal{T})\bigr)$ and $-(k+m-1)\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T})$ since the probability ratios converge to 1 by Lemma 5.2(i). The third term divided by n is by Lemma 6.2(i)

$$\bigl(n+O(1)\bigr)\bigl(-\sigma^{-2}n^{-1}+o(n^{-1})\bigr)\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T}) \to -\sigma^{-2}\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T}).$$

$\square$

This yields immediately variance asymptotics for a functional f with finite support.

Corollary 6.4. *Suppose that f has finite support. Then, as $n \to \infty$,*

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

Proof. By (4.11) and Lemma 6.3, the limit exists and equals

$$2\sum_k\sum_{m<k}\mathbb{E}\bigl(f_k(\mathcal{T})F_m(\mathcal{T})\bigr) + \sum_k \mathbb{E}\bigl(f_k(\mathcal{T})F_k(\mathcal{T})\bigr) - \sum_{k,m}(k+m-1+\sigma^{-2})\,\mathbb{E}\, f_k(\mathcal{T})\,\mathbb{E}\, f_m(\mathcal{T}),$$

where all sums are finite since $f_k = 0$ for large k. Since $f_k(T) = 0$ unless $|T| = k$, we have $f_k(T)F_k(T) = f_k(T)^2$ and $f_k(T)F_m(T) = 0$ for $m > k$, $T \in \mathfrak{T}$. Using this, (4.11) and the similar relations $\sum_k f_k(T)^2 = f(T)^2$ and $\sum_k k f_k(T) = |T|f(T)$, it follows that the limit can be written as

$$\begin{aligned}
&2\sum_k\sum_m \mathbb{E}\bigl(f_k(\mathcal{T})F_m(\mathcal{T})\bigr) - \sum_k \mathbb{E}\bigl(f_k(\mathcal{T})^2\bigr) - 2\sum_k k\,\mathbb{E}\, f_k(\mathcal{T})\sum_m \mathbb{E}\, f_m(\mathcal{T}) \\
&\qquad + \bigl(1-\sigma^{-2}\bigr)\Bigl(\sum_k \mathbb{E}\, f_k(\mathcal{T})\Bigr)^2 \\
&= 2\,\mathbb{E}\bigl(f(\mathcal{T})F(\mathcal{T})\bigr) - \mathbb{E}\bigl(f(\mathcal{T})^2\bigr) - 2\,\mathbb{E}\bigl(|\mathcal{T}|f(\mathcal{T})\bigr)\,\mathbb{E}\, f(\mathcal{T}) + \bigl(1-\sigma^{-2}\bigr)\bigl(\mathbb{E}\, f(\mathcal{T})\bigr)^2.
\end{aligned}$$

$\square$

Remark 6.5. The limit in Corollary 6.4 equals γ^2 in (1.17) for every f with finite support, and more generally for every f such that the expression in (1.17) is finite and $\mathbb{E}\big(|\mathcal{T}||f(\mathcal{T})|\big) < \infty$. Conversely, if $\mathbb{E} f(\mathcal{T}) \neq 0$, the condition $\mathbb{E}\big(|\mathcal{T}||f(\mathcal{T})|\big) < \infty$ is necessary for the expression in Corollary 6.4 to be finite; note that this condition by (4.13) is equivalent to $\sum_{n=1}^{\infty} \mathbb{E}\,|f(\mathcal{T}_n)|/\sqrt{n} < \infty$ and thus imposes a stronger decay of $\mathbb{E} f(\mathcal{T}_n)$ than (1.15).

In order to extend this to more general functionals f, we prove a general upper bound for the variance. We first give another lemma estimating a combination of probability ratios where there typically is a lot of cancellation.

Lemma 6.6. (i) *If $m \leqslant k/2$ and $k \leqslant n$, then*

$$k\frac{\mathbb{P}(S_{k-m} = k-m)}{\mathbb{P}(S_k = k-1)} - \min(k+m-1,n)\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} = O\big(m\big) + O\big(k^{1/2}\big). \tag{6.20}$$

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

$$k\frac{\mathbb{P}(S_{k-m} = k-m)}{\mathbb{P}(S_k = k-1)} - \min(k+m-1,n)\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} = O\left(\frac{k^{3/2}}{(k-m+1)^{1/2}}\right). \tag{6.21}$$

(iii) *If furthermore $\mathbb{E}\,\xi^{2+\delta} < \infty$ with $0 < \delta \leqslant 1$, then the estimate in* (i) *is improved to $O(m) + O\big(k^{(1-\delta)/2}\big)$.*

Proof. (i): By Lemma 5.2(i) (twice),

$$\begin{aligned} &k\frac{\mathbb{P}(S_{k-m} = k-m)}{\mathbb{P}(S_k = k-1)} - (k+m-1)\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} \\ &\quad = k\Big(1 + O\Big(\frac{m}{k}\Big) + O\big(k^{-1/2}\big)\Big) - (k+m-1)\Big(1 + O\Big(\frac{m}{n}\Big) + O\big(n^{-1/2}\big)\Big) \\ &\quad = O(m) + O\big(k^{1/2}\big), \end{aligned} \tag{6.22}$$

which shows (6.20) if also $k+m-1 \leqslant n$.

If $k+m-1 > n$, we have $0 < k+m-1-n < m$ and thus, by Lemma 5.2(i) again (or by (4.2)–(4.3)),

$$(k+m-1-n)\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} = O(k+m-1-n) = O(m), \tag{6.23}$$

and (6.20) follows by adding (6.22) and (6.23).

(ii): By Lemma 5.2(ii),

$$\left|k\frac{\mathbb{P}(S_{k-m} = k-m)}{\mathbb{P}(S_k = k-1)} - \min(k+m-1,n)\frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)}\right|$$

$$\ll k\frac{k^{1/2}}{(k-m+1)^{1/2}}+(k+m)\frac{n^{1/2}}{(n-m+1)^{1/2}},$$

yielding the result since $n/(n-m+1)\leqslant k/(k-m+1)$.

(iii): We use the improved estimate in Lemma 5.4 instead of Lemma 5.2(i) in (6.22) and the result follows as above. ☐

Theorem 6.7. *For any functional $f:\mathfrak{T}\to\mathbb{R}$,*

$$\operatorname{Var}\bigl(F(\mathcal{T}_n)\bigr)^{1/2}\leqslant C_1n^{1/2}\biggl(\sup_k\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2}+\sum_{k=1}^{\infty}\frac{\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2}}{k}\biggr),\tag{6.24}$$

with C_1 independent of f.

Proof. Let $\mu_k:=\mathbb{E}\,f(\mathcal{T}_k)=\mathbb{E}\,f_k(\mathcal{T}_k)$. By the decomposition $f(T)=f'(T)+f''(T)$ where $f'(T):=f(T)-\mu_{|T|}$ and $f''(T):=\mu_{|T|}$, there is a corresponding decomposition $F(\mathcal{T}_n)=F'(\mathcal{T}_n)+F''(\mathcal{T}_n)$. Minkowski's inequality $\operatorname{Var}(X+Y)^{1/2}\leqslant\operatorname{Var}(X)^{1/2}+\operatorname{Var}(Y)^{1/2}$ (for any random variables X and Y) shows that it suffices to show the estimate for $F'(\mathcal{T}_n)$ and $F''(\mathcal{T}_n)$ separately. In other words, it suffices to show (6.24) in the two special cases where either $\mathbb{E}\,f(\mathcal{T}_k)=0$ for every k (so $f=f'$), or $f(T)=\mu_{|T|}$ depends on $|T|$ only (so $f=f''$). Recall the notation $\pi_n:=\mathbb{P}(|\mathcal{T}|=n)$ from (4.14).

Case 1: $\mathbb{E}\,f(\mathcal{T}_k)=0$ *for every* k. In this case, since $f(\mathcal{T}_k)=f_k(\mathcal{T}_k)$, also $\mathbb{E}\,f_k(\mathcal{T}_k)=0$ and by (5.6) and (5.2) (and the trivial $F_k(\mathcal{T}_n)=0$ for $k>n$),

$$\mathbb{E}\,f_k(\mathcal{T})=0,\quad \mathbb{E}\,F_k(\mathcal{T}_n)=0,\quad \mathbb{E}\,F_k(\mathcal{T})=\sum_{n=1}^{\infty}\pi_n\,\mathbb{E}\,F_k(\mathcal{T}_n)=0,\tag{6.25}$$

for all $k\geqslant1$, $n\geqslant1$. Hence, (6.1) and (6.2) yield the same result and Lemma 6.1 reduces to, for $m\leqslant k\leqslant n$,

$$\begin{aligned}\operatorname{Cov}(F_k(\mathcal{T}_n),F_m(\mathcal{T}_n))&=n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\,\mathbb{E}\bigl(f_k(\mathcal{T})F_m(\mathcal{T})\bigr)\\&=n\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\pi_k\,\mathbb{E}\bigl(f_k(\mathcal{T}_k)F_m(\mathcal{T}_k)\bigr).\end{aligned}\tag{6.26}$$

Consequently, using again $F_m(\mathcal{T}_k)=0$ for $m>k$ and $F_k(\mathcal{T}_k)=f_k(\mathcal{T}_k)$,

$$\begin{aligned}\frac1n\operatorname{Var}F(\mathcal{T}_n)&=\frac1n\sum_{k=1}^{n}\sum_{m=1}^{n}\operatorname{Cov}(F_k(\mathcal{T}_n),F_m(\mathcal{T}_n))\\&=\frac1n\sum_{k=1}^{n}\sum_{m=1}^{k}(2-\delta_{km})\operatorname{Cov}(F_k(\mathcal{T}_n),F_m(\mathcal{T}_n))\\&=\sum_{k=1}^{n}\sum_{m=1}^{k}(2-\delta_{km})\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\pi_k\,\mathbb{E}\bigl(f_k(\mathcal{T}_k)F_m(\mathcal{T}_k)\bigr)\\&=\sum_{k=1}^{n}\frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\pi_k\,\mathbb{E}\Bigl(f_k(\mathcal{T}_k)\bigl(2F(\mathcal{T}_k)-f_k(\mathcal{T}_k)\bigr)\Bigr)\end{aligned}\tag{6.27}$$

$$\leqslant 2\sum_{k=1}^{n} \frac{\mathbb{P}(S_{n-k}=n-k)}{\mathbb{P}(S_n=n-1)}\pi_k\,\mathbb{E}\big(f_k(\mathcal{T}_k)F(\mathcal{T}_k)\big). \tag{6.28}$$

By (4.2)–(4.3), (4.13), (6.25) and the Cauchy–Schwarz inequality, this yields

$$\frac{1}{n}\operatorname{Var} F(\mathcal{T}_n) \ll \sum_{k=1}^{n} \frac{n^{1/2}}{(n-k+1)^{1/2}k^{3/2}}\big(\operatorname{Var} f_k(\mathcal{T}_k)\big)^{1/2}\big(\operatorname{Var} F(\mathcal{T}_k)\big)^{1/2}.$$

Let us write $\operatorname{Var} f_k(\mathcal{T}_k)=\alpha_k^2$ and $\operatorname{Var} F(\mathcal{T}_k)=k\beta_k^2$, and let further

$$B := \sup_k \alpha_k + \sum_{k=1}^{\infty}\frac{\alpha_k}{k} \tag{6.29}$$

and $\beta_n^* := \sup_{k\leqslant n}\beta_k$. Then we have shown

$$\begin{aligned}
\beta_n^2 &\ll \sum_{k=1}^{n} \frac{n^{1/2}}{(n-k+1)^{1/2}k^{3/2}}\alpha_k k^{1/2}\beta_k \\
&\ll \sum_{k=1}^{n/2}\frac{\alpha_k}{k}\beta_k + n^{-1/2}\sum_{k=n/2}^{n}\frac{\alpha_k}{(n-k+1)^{1/2}}\beta_k \\
&\ll \beta_n^*\sum_{k=1}^{\infty}\frac{\alpha_k}{k} + n^{-1/2}\beta_n^*\sup_k\alpha_k\sum_{k=n/2}^{n}\frac{1}{(n-k+1)^{1/2}} \\
&\ll B\beta_n^*.
\end{aligned} \tag{6.30}$$

In other words, $\beta_n^2 \leqslant C_1 B\beta_n^*$ for some C_1. The sequence β_n^* is increasing, and thus we obtain

$$(\beta_n^*)^2 = \sup_{1\leqslant m\leqslant n}\beta_m^2 \leqslant C_1 B\beta_n^*. \tag{6.31}$$

Consequently, recalling that β_n and β_n^* are finite by Remark 3.3,

$$\beta_n \leqslant \beta_n^* \leqslant C_1 B, \tag{6.32}$$

i.e., $\operatorname{Var} F(\mathcal{T}_n)=n\beta_n^2\leqslant nC_1^2B^2$, which, recalling (6.29), completes the proof of Case 1.

Case 2: $f(T)=\mu_{|T|}$. In this case, $f(\mathcal{T}_k)=f_k(\mathcal{T}_k)=\mu_k$, and (6.4) and (5.2) yield

$$\begin{aligned}
\mathbb{E}\big(f_k(\mathcal{T})F_m(\mathcal{T})\big) &= \mathbb{P}(|\mathcal{T}|=k)\,\mathbb{E}\big(f_k(\mathcal{T}_k)F_m(\mathcal{T}_k)\big) = \mathbb{P}(|\mathcal{T}|=k)\mu_k\,\mathbb{E}\big(F_m(\mathcal{T}_k)\big) \\
&= \mathbb{E}\, f_k(\mathcal{T})k\frac{\mathbb{P}(S_{k-m}=k-m)}{\mathbb{P}(S_k=k-1)}\,\mathbb{E}\, f_m(\mathcal{T}).
\end{aligned}$$

Thus Lemma 6.1 can be written, for $n\geqslant k\geqslant m$ (and assuming $\mathbb{P}(S_k=k-1)>0$; otherwise $\mathbb{P}(|\mathcal{T}|=k)=0$ and $F_k(\mathcal{T}_n)=0$ a.s., so we may ignore this case),

$$\frac{1}{n}\operatorname{Cov}(F_k(\mathcal{T}_n),F_m(\mathcal{T}_n))$$

$$= \mathbb{E} f_k(\mathcal{T}) \mathbb{E} f_m(\mathcal{T}) \Biggl(\frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} \times \\ \Biggl(k \frac{\mathbb{P}(S_{k-m} = k-m)}{\mathbb{P}(S_k = k-1)} - \min(k+m-1, n) \frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} \Biggr) \\ + (n-k-m+1)_+ \times \\ \Biggl(\frac{\mathbb{P}(S_{n-k-m} = n-k-m+1)}{\mathbb{P}(S_n = n-1)} - \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} \frac{\mathbb{P}(S_{n-m} = n-m)}{\mathbb{P}(S_n = n-1)} \Biggr) \Biggr) \\ =: (A_1 + A_2) \mathbb{E} f_k(\mathcal{T}) \mathbb{E} f_m(\mathcal{T}). \tag{6.33}$$

We take absolute values and sum over all $m \leqslant k \leqslant n$ (the terms with $k > m$ are covered by symmetry). Cancellations inside A_1 and A_2 will be important, but we treat the two terms A_1 and A_2 separately.

For convenience, we write

$$x_k := |\mu_k|/\sqrt{k}. \tag{6.34}$$

Thus, by (6.3) and (4.13),

$$\mathbb{E} f_k(\mathcal{T}) = \mathbb{P}(|\mathcal{T}| = k)\mu_k = O\bigl(|\mu_k|/k^{3/2}\bigr) = O(x_k/k). \tag{6.35}$$

Consequently, by (4.11), (6.33) and symmetry, we have

$$\frac{1}{n} \operatorname{Var} F(\mathcal{T}_n) \ll \sum_{\substack{m \leqslant k \\ k \leqslant n}} (|A_1| + |A_2|) \frac{x_k x_m}{km}. \tag{6.36}$$

To estimate this sum, we consider several cases. We define

$$B_1 := \sum_{k=1}^{\infty} \frac{x_k}{\sqrt{k}} = \sum_{k=1}^{\infty} \frac{|\mu_k|}{k}, \tag{6.37}$$

$$B_2 := \sum_{k=1}^{\infty} \frac{x_k}{k} = \sum_{k=1}^{\infty} \frac{|\mu_k|}{k^{3/2}}, \tag{6.38}$$

$$B_3 := \sup_{n \geqslant 1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} x_k = \sup_{n \geqslant 1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{|\mu_k|}{\sqrt{k}}, \tag{6.39}$$

$$B_4 := \sup_{n} \sum_{n/2}^{n} \frac{x_k}{\sqrt{n-k+1}} \ll \sup_{n} \frac{1}{\sqrt{n}} \sum_{n/2}^{n} \frac{|\mu_k|}{\sqrt{n-k+1}}. \tag{6.40}$$

Case 2(i): *A_1 for $m \leqslant k/2$, $k \leqslant n/2$.* By Lemma 5.2(i) (or (4.2)–(4.3)) and Lemma 6.6(i),

$$|A_1| \ll m + k^{1/2}. \tag{6.41}$$

Hence, the contribution to (6.36) from these terms is

$$\sum_{\substack{m\leqslant k/2\\ k\leqslant n/2}} |A_1|\frac{x_kx_m}{km} \ll \sum_{k\leqslant n}\frac{x_k}{k}\sum_{m\leqslant k}x_m + \sum_{k\leqslant n}\frac{x_k}{\sqrt{k}}\sum_m \frac{x_m}{m}$$
$$\leqslant \sum_{k\leqslant n}\frac{x_k}{\sqrt{k}}(B_3+B_2) \leqslant B_1(B_3+B_2). \tag{6.42}$$

Case 2(ii): *A_1 for $m\leqslant k/2$, $n/2<k\leqslant n$.* By Lemmas 5.2(ii) and 6.6(i),

$$|A_1| \ll \frac{n^{1/2}}{(n-k+1)^{1/2}}\bigl(m+k^{1/2}\bigr) \leqslant \frac{n^{1/2}m}{(n-k+1)^{1/2}} + \frac{n}{(n-k+1)^{1/2}}. \tag{6.43}$$

Hence, the contribution to (6.36) from these terms is

$$\sum_{\substack{m\leqslant k/2\\ n/2<k\leqslant n}} |A_1|\frac{x_kx_m}{km} \ll \sum_{n/2<k\leqslant n}\frac{x_k}{(n-k+1)^{1/2}}\left(\sum_{m\leqslant n}\frac{x_m}{n^{1/2}} + \sum_m\frac{x_m}{m}\right)$$
$$\leqslant B_4\bigl(B_3+B_2\bigr). \tag{6.44}$$

Case 2(iii): *A_1 for $k/2<m\leqslant k$, $k\leqslant n/2$.* By Lemma 5.2(i) (or (4.2)–(4.3)) and Lemma 6.6(ii),

$$|A_1| \ll \frac{k^{3/2}}{(k-m+1)^{1/2}} \ll \frac{k^{1/2}m}{(k-m+1)^{1/2}}. \tag{6.45}$$

Hence, the contribution to (6.36) from these terms is

$$\sum_{\substack{k/2<m\leqslant k\\ k\leqslant n/2}} |A_1|\frac{x_kx_m}{km} \ll \sum_{\substack{k/2<m\leqslant k\\ k\leqslant n/2}} \frac{x_kx_m}{\sqrt{k}\sqrt{k-m+1}} \leqslant B_4\sum_{k\leqslant n/2}\frac{x_k}{\sqrt{k}} \leqslant B_4B_1. \tag{6.46}$$

Case 2(iv): *A_1 for $k/2<m\leqslant k$, $n/2<k\leqslant n$.* By Lemmas 5.2(ii) and 6.6(ii), noting $k\leqslant n\ll m$,

$$|A_1| \ll \frac{n^{1/2}}{(n-k+1)^{1/2}}\frac{k^{3/2}}{(k-m+1)^{1/2}} \ll \frac{km}{\sqrt{n-k+1}\sqrt{k-m+1}}. \tag{6.47}$$

Hence, the contribution to (6.36) from these terms is

$$\sum_{\substack{k/2<m\leqslant k\\ n/2<k\leqslant n}} |A_1|\frac{x_kx_m}{km} \ll \sum_{n/2<k\leqslant n}\frac{x_k}{\sqrt{n-k+1}}\sum_{k/2<m\leqslant k}\frac{x_m}{\sqrt{k-m+1}} \leqslant B_4^2. \tag{6.48}$$

Case 2(v): *A_2 for $m\leqslant k$ and $m+k\leqslant n/2$.* In this case, Lemma 6.2(i) yields

$$|A_2| \ll (n-k-m+1)\Bigl(\frac{1}{n}+\frac{k+m}{n^{3/2}}+\frac{km}{n^2}\Bigr) \leqslant 1+\frac{2k}{n^{1/2}}+\frac{km}{n} \tag{6.49}$$

and the contribution to (6.36) from these terms is dominated by

$$\sum_{k+m\leqslant n/2}|A_2|\frac{x_kx_m}{km} \ll \Bigl(\sum_{k\leqslant n}\frac{x_k}{k}\Bigr)^2+\frac{1}{n^{1/2}}\sum_{k\leqslant n}x_k\sum_{m\leqslant n}\frac{x_m}{m}+\frac{1}{n}\Bigl(\sum_{k\leqslant n}x_k\Bigr)^2 \leqslant B_2^2+B_3B_2+B_3^2. \tag{6.50}$$

Case 2(vi): *A_2 for $m\leqslant k\leqslant n$ and $m+k>n/2$.* Note that A_2 vanishes unless $n\geqslant k+m$. In this case, Lemma 6.2(ii) yields

$$|A_2| \ll \frac{m\,n^{1/2}}{(n-k-m+1)^{1/2}}+n^{1/2}. \tag{6.51}$$

Since $k+m>n/2$ and $k\geqslant m$ imply $k>n/4$, the contribution from these terms to (6.36) is at most

$$\sum_{k+m>n/2}|A_2|\frac{x_kx_m}{km} \ll \sum_{k=n/4}^{n}\frac{x_k}{\sqrt n}\sum_{m=1}^{n-k}\frac{x_m}{(n-k-m+1)^{1/2}}+\sum_{k=n/4}^{n}\frac{x_k}{\sqrt n}\sum_{m=1}^{n-k}\frac{x_m}{m} \ll B_3(B_3+B_4)+B_3B_2. \tag{6.52}$$

Conclusion: Consequently, (6.36) together with (6.42), (6.44), (6.46), (6.48), (6.50) and (6.52) show that

$$\frac{1}{n}\operatorname{Var}F(\mathcal{T}_n) \ll \bigl(B_1+B_2+B_3+B_4\bigr)^2. \tag{6.53}$$

Furthermore, trivially $B_2\leqslant B_1$ and $B_3,B_4\ll\sup_k|\mu_k|$, and $|\mu_k|\leqslant\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2}$. Hence, (6.53) proves (6.24) in Case 2, which completes the proof. □

Remark 6.8. The proof actually yields, noting that $B_3\ll B_4$, the slightly stronger

$$\operatorname{Var}\bigl(F(\mathcal{T}_n)\bigr)^{1/2} \leqslant C_2n^{1/2}\Bigl(\sup_k\frac{1}{\sqrt k}\sum_{j=k/2}^{k}\frac{\sqrt{\mathbb{E}\,f(\mathcal{T}_j)^2}}{\sqrt{k-j+1}}+\sum_{k=1}^{\infty}\frac{\sqrt{\mathbb{E}\,f_k(\mathcal{T}_k)^2}}{k}\Bigr).$$

Remark 6.9. Example 6.13 below shows that the term $\sum_k(\mathbb{E}\,f_k(\mathcal{T}_k)^2)^{1/2}/k$ in (6.24) cannot be improved in general. In the case $f(T)=\mu_{|T|}$, there is, however, a minor improvment in the following version of Theorem 6.7, provided we have more than a second moment of ξ. This theorem implies, as mentioned in Remark 1.10, a corresponding minor improvement of the condition (1.15) in Theorem 1.5; we omit the details. (For example, it allows $f(T)=1/\log|T|$.) We do not know whether $\operatorname{Var}F(\mathcal{T}_n)=O(n)$ for every bounded $f(T)$ that depends on $|T|$ only.

Theorem 6.10. *Suppose that $\mathbb{E}\,\xi^{2+\delta}<\infty$ with $\delta>0$ and let $\mu_k:=\mathbb{E}\,f(\mathcal{T}_k)=\mathbb{E}\,f_k(\mathcal{T}_k)$. Then*

$$\operatorname{Var}\bigl(F(\mathcal{T}_n)\bigr)^{1/2}$$

$$\leqslant C_3 n^{1/2}\biggl(\sup_k \bigl(\mathbb{E}\, f_k(\mathcal{T}_k)^2\bigr)^{1/2} + \sum_{k=1}^{\infty} \frac{(\operatorname{Var} f_k(\mathcal{T}_k))^{1/2}}{k} + \biggl(\sum_{k=1}^{\infty} \frac{\mu_k^2}{k}\biggr)^{1/2}\biggr).$$

Proof. We use the notation of the proof of Theorem 6.7, and define B_5 by

$$B_5^2 := \sum_{k=1}^{\infty} x_k^2 = \sum_{k=1}^{\infty} \frac{\mu_k^2}{k}. \tag{6.54}$$

Observe that $B_2 \ll B_5$ by the Cauchy–Schwarz inequality. We modify the proof of Theorem 6.7. Case 1 is as before, and so are Case 2(ii),(iv),(v),(vi), leaving only two cases where we have to replace B_1.

Case 2 (i): Using Lemma 6.6(iii) instead of Lemma 6.6(i), we obtain

$$|A_1| \ll m + k^{(1-\delta)/2}. \tag{6.55}$$

Hence, the contribution to (6.36) from these terms is

$$\sum_{\substack{m\leqslant k/2\\ k\leqslant n/2}} |A_1| \frac{x_k x_m}{km} \ll \sum_{k,m} \frac{x_k x_m}{\max(k,m)} + \sum_k \frac{x_k}{k^{(1+\delta)/2}} \sum_m \frac{x_m}{m}. \tag{6.56}$$

The second term on the right-hand side is $\ll \sum_k x_k^2 = B_5^2$ by two applications of the Cauchy–Schwarz inequality. The same holds for the first term, which says that the infinite matrix $(1/\max\{k,m\})_{k,m=1}^{\infty}$ defines a bounded operator on ℓ^2; this follows from Hilbert's inequality $\sum_{k,m} x_k x_m/(k+m) \leqslant \pi \sum_k x_k^2$, see for example [22, Chapter IX].

Case 2 (iii): We use again (6.45), but in (6.46) we set $k = m+j$ and observe that $j < k/2 < m$ and thus, using the Cauchy–Schwarz inequality,

$$\sum_{\substack{k/2<m\leqslant k\\ k\leqslant n/2}} \frac{x_k x_m}{\sqrt{k}\sqrt{k-m+1}} \leqslant \sum_{j=0}^{\infty} \frac{1}{\sqrt{j+1}} \sum_{m=j+1}^{\infty} \frac{x_{m+j} x_m}{(m+j)^{1/4} m^{1/4}}$$

$$\leqslant \sum_{j=0}^{\infty} \frac{1}{\sqrt{j+1}} \sum_{m=j+1}^{\infty} \frac{x_m^2}{m^{1/2}} = \sum_{m=1}^{\infty} \frac{x_m^2}{m^{1/2}} \sum_{j=0}^{m-1} \frac{1}{\sqrt{j+1}} \ll \sum_{m=1}^{\infty} x_m^2 = B_5^2.$$

The different terms for Case 2 thus all are dominated by $(B_2 + B_3 + B_4 + B_5)^2 \ll (B_4 + B_5)^2$, which completes the proof. □

Remark 6.11. In (6.56) we used Hilbert's inequality. One can also see in other ways that $(1/\max\{k,m\})_{k,m=1}^{\infty}$ defines a bounded operator on ℓ^2; a much more general result is shown in [43] and [6, Theorem 3.1] (for the continuous case, which implies the discrete).

In the special case of (weakly) local functionals, we can improve Theorem 6.7. For simplicity we consider only bounded functionals.

Theorem 6.12. *Suppose that $f(T)$ is a bounded and weakly local functional on $\mathfrak{T}$ with cut-off M. Let $A_k := \sup\{|f(T)| : |T| = k\}$ and $\mu_k := \mathbb{E} f(\mathcal{T}_k)$. Then,*

$$\operatorname{Var}\bigl(F(\mathcal{T}_n)\bigr)^{1/2} \leqslant C_1 n^{1/2}\biggl(\Bigl(\sup_k A_k \sup_k k^{-1/4} A_k\Bigr)^{1/2} + \sup_k |\mu_k| + \sum_{k=1}^{\infty} \frac{|\mu_k|}{k}\biggr)$$

where the constant C_1 may depend on the cut-off M but not otherwise on f.

Proof. We modify the proof of Theorem 6.7; Case 2 is the same so we consider Case 1 only. Thus assume that $\mathbb{E} f(\mathcal{T}_k) = 0$ for each k. We have

$$F(\mathcal{T}_k) = \sum_{v\in\mathcal{T}_k} f(\mathcal{T}_{k,v}) = \sum_{d(v)<M} f(\mathcal{T}_{k,v}) + \sum_{d(v)\geqslant M} f(\mathcal{T}_{k,v}) =: S_1 + S_2. \tag{6.57}$$

Let again $w_j(T)$ be the number of nodes at depth j. Since $|f|$ is bounded by A, the sum S_1 is bounded by

$$|S_1| \leqslant A \sum_{j=0}^{M-1} w_j(\mathcal{T}_k). \tag{6.58}$$

As said in the proof of Lemma 5.8, $\mathbb{E}\, w_j(\mathcal{T}_k) = O(j)$ for each j, see [44] and [29, Theorem 1.13]. Hence (6.58) implies $\mathbb{E}|S_1| = O(A)$ and

$$\mathbb{E}|f_k(\mathcal{T}_k)S_1| \leqslant \mathbb{E}|A_k S_1| = O(A_k A). \tag{6.59}$$

For S_2 we condition on $\mathcal{T}_k^{(M)}$ and on the sizes n_v of the $w_M(\mathcal{T}_k)$ subtrees T_v with $d(v) = M$. Given this, the forest $\mathcal{T}_k \setminus \mathcal{T}_k^{(M-1)}$ consists of independent copies of random trees $\mathcal{T}_{d_v}$, and S_2 is the sum of $f(\mathcal{T}_{k,v})$ over all nodes in these trees, which equals the sum of F over these trees. Since each $\mathbb{E} F(\mathcal{T}_{d_i}) = 0$ by Lemma 5.1, it follows that the conditional expectation $\mathbb{E}\bigl(S_2 \mid \mathcal{T}_k^{(M)}\bigr) = 0$. However, $f_k(\mathcal{T}_k)$ depends only on $\mathcal{T}_k^{(M)}$, and thus $\mathbb{E}\bigl(f_k(\mathcal{T}_k)S_2\bigr) = 0$. Consequently, (6.59) yields

$$\mathbb{E}\bigl(f_k(\mathcal{T}_k)F(\mathcal{T}_k)\bigr) = O(A_k A). \tag{6.60}$$

By (6.28), (6.60) and estimates using (4.2)–(4.3) and (4.13) as before,

$$\frac{1}{n}\operatorname{Var} F(\mathcal{T}_n) \ll \sum_{k=1}^{n/2} k^{-3/2} A_k A + \sup_{k>n/2} k^{-1/2} A_k A,$$

and the result follows. (The exponent $-1/4$ can be replaced by any exponent $> -1/2$.) ∎

Example 6.13. We show by an example, where we make correlations between different $F_k(\mathcal{T}_n)$ large, that the condition (1.15) in general cannot be relaxed. We consider any offspring distribution such that $p_0, p_1, p_2 > 0$.

Suppose that $(\alpha_k)_3^\infty$ is a given sequence of positive numbers. Define $f_1 = f_2 := 0$ and let $g_3(T) := \#\{\text{leaves in } T\} - c_0$ when $|T| = 3$, where the constant c_0 is chosen such that $\mathbb{E}\, g_3(\mathcal{T}_3) = 0$. (Note that $\mathcal{T}_3$ has one or two leaves, both with positive probability, so $g_3(\mathcal{T}_3) \neq 0$.) Let $f_3(T) = s_3 g_3(T)$ for a constant $s_3 > 0$ such that $\operatorname{Var} f_3(\mathcal{T}_3) = \alpha_3^2$. Continue recursively. If we have chosen $f_1, \dots, f_{k-1}$, let for a tree T with $|T| = k$,

$$g_k(T) := \sum_{j=1}^{k-1} F_j(T) = \sum_{v \in T}{}' f_{|T_v|}(T_v), \tag{6.61}$$

where $\sum'$ denotes summation over all nodes except the root. Define

$$f_k(T) = s_k g_k(T), \qquad T \in \mathfrak{T}_k, \tag{6.62}$$

for a constant $s_k > 0$ such that $\operatorname{Var} f_k(\mathcal{T}_k) = \alpha_k^2$. Note that, by induction, and Lemma 5.1, $\mathbb{E} f_k(\mathcal{T}_k) = \mathbb{E} g_k(\mathcal{T}_k) = 0$ for every k.

Consider $f = \sum_k f_k$ and the corresponding $F = \sum_k F_k$. By construction, for a tree T with $|T| = k > 3$,

$$F(T) = f_k(T) + g_k(T) = (1 + s_k) g_k(T). \tag{6.63}$$

If we let $b_k^2 := \operatorname{Var} g_k(\mathcal{T}_k)$, we have $\alpha_k^2 = s_k^2 b_k^2$ so $\alpha_k = s_k b_k$. Thus, for $k > 3$,

$$\operatorname{Var} F(\mathcal{T}_k) = (1 + s_k)^2 \operatorname{Var} g_k(\mathcal{T}_k) = (1 + s_k)^2 b_k^2 = (\alpha_k + b_k)^2 \tag{6.64}$$

and

$$\begin{aligned} \mathbb{E}\bigl(f_k(\mathcal{T}_k)\bigl(2F(\mathcal{T}_k) - f_k(\mathcal{T}_k)\bigr)\bigr) &= \mathbb{E}\bigl(s_k g_k(\mathcal{T}_k)(2 + s_k) g_k(\mathcal{T}_k)\bigr) \\ &= s_k(2 + s_k) b_k^2 = \alpha_k(2b_k + \alpha_k). \end{aligned} \tag{6.65}$$

For $k = 3$, $F(\mathcal{T}_3) = f_3(\mathcal{T}_3)$, and (6.64)–(6.65) hold if we redefine $b_3 := 0$.

By (6.61), (6.27) (with f_n temporarily redefined as 0) and (6.65),

$$\begin{aligned} b_n^2 = \operatorname{Var} g_n(\mathcal{T}_n) &= n \sum_{k=3}^{n-1} \frac{\mathbb{P}(S_{n-k} = n - k)}{\mathbb{P}(S_n = n - 1)} \pi_k \, \mathbb{E}\bigl(f_k(\mathcal{T}_k)\bigl(2F(\mathcal{T}_k) - f_k(\mathcal{T}_k)\bigr)\bigr) \\ &= n \sum_{k=3}^{n-1} \frac{\mathbb{P}(S_{n-k} = n - k)}{\mathbb{P}(S_n = n - 1)} \pi_k \alpha_k (2b_k + \alpha_k). \end{aligned} \tag{6.66}$$

In particular, for $n > 3$, using (4.2) (or Lemma 5.2),

$$b_n^2 \geqslant n \frac{\mathbb{P}(S_{n-3} = n - 3)}{\mathbb{P}(S_n = n - 1)} \pi_3 \alpha_3^2 \geqslant c_1 n. \tag{6.67}$$

If $n - k \geqslant 1$, (4.2) implies $\mathbb{P}(S_{n-k} = n - k) \geqslant c_2 (n - k)^{-1/2} \geqslant c_2 n^{-1/2}$ and thus, using also (4.3), $\mathbb{P}(S_{n-k} = n - k) / \mathbb{P}(S_{n-1} = n) \geqslant c_3$. Using this, (6.67) and (4.13) in (6.66) yields, noting (6.64),

$$\operatorname{Var} F(\mathcal{T}_n) \geqslant b_n^2 \geqslant c_4 n \sum_{k=4}^{n-1} k^{-3/2} \alpha_k k^{1/2}. \tag{6.68}$$

It follows that if $\sum_{k=3}^{\infty} \alpha_k / k = \infty$, then $\operatorname{Var} F(\mathcal{T}_n)/n \to \infty$.

Consequently, the condition (1.15) is in general necessary for $\operatorname{Var} F(\mathcal{T}_n) = O(n)$, even if we assume $\mathbb{E} f(\mathcal{T}_n) = 0$. In particular, taking $\alpha_k = 1/\log k$,

$k \geqslant 3$, we find an example where $\mathbb{E} f(\mathcal{T}_k)^2 \to 0$ as $k \to 0$ but $\operatorname{Var} F(\mathcal{T}_n)/n \to \infty$.

Example 6.14. We modify example Example 6.13 to have $f_k(T)$ uniformly bounded by defining recursively, instead of (6.62),

$$f_k(T) = s_k\big(\operatorname{sign}(g_k(T)) - \mathbb{E}\operatorname{sign}(g_k(\mathcal{T}_k))\big), \qquad T \in \mathfrak{T}_k, \tag{6.69}$$

for a given bounded sequence $(s_k)_3^\infty$ of positive numbers. We still have $\mathbb{E} f_k(\mathcal{T}_k) = \mathbb{E} g_k(\mathcal{T}_k) = 0$. Furthermore,

$$\mathbb{E}\big(f_k(\mathcal{T}_k) g_k(\mathcal{T}_k)\big) = s_k \,\mathbb{E}\,|g_k(\mathcal{T}_k)|. \tag{6.70}$$

Since $F(\mathcal{T}_n) = f_n(\mathcal{T}_n) + g_n(\mathcal{T}_n)$, it follows, similarly to (6.66), that

$$\begin{aligned}\operatorname{Var} F(\mathcal{T}_n) &\geqslant \operatorname{Var} g_n(\mathcal{T}_n)\\ &= n\sum_{k=3}^{n-1} \frac{\mathbb{P}(S_{n-k} = n-k)}{\mathbb{P}(S_n = n-1)} \pi_k \,\mathbb{E}\big(f_k(\mathcal{T}_k)\big(2F(\mathcal{T}_k) - f_k(\mathcal{T}_k)\big)\big)\\ &\geqslant c_5 n \sum_{k=3}^{n-1} \pi_k s_k \,\mathbb{E}\,|g_k(\mathcal{T}_k)| \geqslant c_6 n \sum_{k=3}^{n-1} s_k \,\mathbb{E}\,|g_k(\mathcal{T}_k)|/k^{3/2}. \end{aligned} \tag{6.71}$$

In particular $\operatorname{Var} g_n(\mathcal{T}_n) \geqslant c_7 n$. It seems likely that also

$$\mathbb{E}\,|g_n(\mathcal{T}_n)| \geqslant c_8 n^{1/2}; \tag{6.72}$$

if this is the case with, for example, $s_k = 1/\log k$, then (6.71) shows that $\operatorname{Var} F(\mathcal{T}_n)/n \to \infty$ although f is bounded, (1.14) holds and $\mathbb{E} f(\mathcal{T}_n) = 0$ for all n.

Unfortunately, we have not been able to show (6.72), but we note that if (1.18) holds (with $\mu = 0$ as in our case), then $\liminf_{n\to\infty} \mathbb{E}\,|F(\mathcal{T}_n)|/\sqrt{n} \geqslant \sqrt{2/\pi}\gamma$ by Fatou's lemma, and since $F(\mathcal{T}_n) - g_n(\mathcal{T}_n) = f_n(\mathcal{T}_n) = O(1)$, it follows that (6.72) holds. Hence we can at least conclude that, for $s_k = 1/\log k$, (1.16) and (1.18) cannot both hold (for any finite γ^2).

7. Asymptotic normality

In this section we consider only functionals f with finite support. Recall that this implies that f is bounded.

Lemma 7.1. *Suppose that f has finite support. Then, with notations as in Theorem 1.5,*

$$\frac{F(\mathcal{T}_n) - n\mu}{\sqrt{n}} \xrightarrow{\text{d}} N(0,\gamma^2). \tag{7.1}$$

Proof. We use the representation (4.7). Since f has finite support, there exists m such that $f_k = 0$ for $k > m$; this means that it suffices to sum over $k \leqslant m$ in (4.7). We define

$$g(x_1,\dots,x_m) := \sum_{k=1}^m f(x_1,\dots,x_k) = \sum_{k=1}^m f_k(x_1,\dots,x_k). \tag{7.2}$$

Then (4.7) can be written (assuming $n \geqslant m$)

$$F(\mathcal{T}_n) \overset{\mathrm{d}}{=} \biggl(\sum_{i=1}^{n} g(\xi_i, \dots, \xi_{i+m-1 \bmod n}) \biggm| S_n = n-1\biggr). \tag{7.3}$$

We now use a method by Le Cam [39] and Holst [25], see also Kudlaev [38]. (See in particular [25, Theorem 5.1]; the conditions are somewhat different but we use essentially the same proof.)

Note first that by (7.2) and (5.4),

$$\mathbb{E}\, g(\xi_1, \dots, \xi_m) = \sum_{k=1}^{m} \mathbb{E}\, f_k(\mathcal{T}) = \mathbb{E}\, f(\mathcal{T}) = \mu. \tag{7.4}$$

Furthermore, g is bounded, because f is.

Fix α with $0 < \alpha < 1$ and a sequence $n' = n'(n)$ with $n'/n \to \alpha$, for example $n' = \lfloor \alpha n \rfloor$. Define the centred sum

$$Y_n := \sum_{i=1}^{n} \bigl(g(\xi_i, \dots, \xi_{i+m-1}) - \mu\bigr). \tag{7.5}$$

Then, by the standard central limit for m-dependent variables [23], [14], applied to the random vectors $\bigl(g(\xi_i, \dots, \xi_{i+m-1}) - \mu, \xi_i\bigr)$,

$$\left(\frac{Y_{n'}}{\sqrt{n}}, \frac{S_{n'} - n'}{\sqrt{n}}\right) \overset{\mathrm{d}}{\longrightarrow} N\left(0, \alpha \begin{pmatrix} \beta^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}\right) \tag{7.6}$$

where

$$\beta^2 = \operatorname{Var}\bigl(g(\xi_1, \dots, \xi_m)\bigr) + 2 \sum_{i=2}^{m} \operatorname{Cov}\bigl(g(\xi_1, \dots, \xi_m), g(\xi_i, \dots, \xi_{i+m-1})\bigr),$$

$$\rho = \sum_{i=1}^{m} \operatorname{Cov}\bigl(g(\xi_1, \dots, \xi_m), \xi_i\bigr) = \operatorname{Cov}\bigl(g(\xi_1, \dots, \xi_m), S_m\bigr).$$

We calculate β^2 by expanding g using (7.2) and arguing as in (6.5)—(6.8) in the proof of Lemma 6.1 (where we condition on $S_n = n-1$, making the present calculation simpler). This yields, omitting the details, cf. also the proof of Corollary 6.4, and using (1.17),

$$\begin{aligned} \beta^2 &= \sum_{\ell \leqslant k}^{m} (2 - \delta_{k,\ell})\, \mathbb{E}\bigl(f_k(\mathcal{T}) F_\ell(\mathcal{T})\bigr) - \sum_{k,\ell=1}^{m} (k + \ell - 1)\, \mathbb{E}\, f_k(\mathcal{T})\, \mathbb{E}\, f_\ell(\mathcal{T}) \\ &= \mathbb{E}\bigl(f(\mathcal{T})(2F(\mathcal{T}) - f(\mathcal{T}))\bigr) - 2\,\mathbb{E}\bigl(|\mathcal{T}| f(\mathcal{T})\bigr)\mu + \mu^2 \\ &= \gamma^2 + \mu^2/\sigma^2. \end{aligned} \tag{7.7}$$

Furthermore, since $f_k(\xi_1, \dots, \xi_k) \neq 0$ only when $(\xi_1, \dots, \xi_k)$ is the degree sequence of a tree, and thus $S_k = k-1$, while $\mathbb{E}\, S_k = k$, and using (5.4) again,

$$\begin{aligned}\rho &= \sum_{k=1}^{m} \operatorname{Cov}\bigl(f_k(\xi_1,\dots,\xi_k), S_m\bigr) = \sum_{k=1}^{m} \operatorname{Cov}\bigl(f_k(\xi_1,\dots,\xi_k), S_k\bigr) \\ &= \sum_{k=1}^{m} \mathbb{E}\bigl(f_k(\xi_1,\dots,\xi_k)(S_k-k)\bigr) = \sum_{k=1}^{m} \mathbb{E}\bigl(-f_k(\xi_1,\dots,\xi_k)\bigr) \\ &= -\sum_{k=1}^{m} \mathbb{E}\, f_k(\mathcal{T}) = -\mu. \end{aligned} \tag{7.8}$$

We define for convenience

$$\tilde{Y}_n := Y_n + \frac{\mu}{\sigma^2}(S_n - n). \tag{7.9}$$

Then (7.6) yields, using (7.7)–(7.8),

$$\left(\frac{\tilde{Y}_{n'}}{\sqrt{n}}, \frac{S_{n'}-n'}{\sqrt{n}}\right) \xrightarrow{\mathrm{d}} N\left(0, \alpha\begin{pmatrix}\beta^2-\mu^2/\sigma^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right) = N\left(0, \alpha\begin{pmatrix}\gamma^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right). \tag{7.10}$$

In other words, $\tilde{Y}_{n'}/\sqrt{n}$ and $(S_{n'}-n')/\sqrt{n}$ are jointly asymptotically normal with independent limits $W \sim N(0, \alpha\gamma^2)$ and $Z \sim N(0, \alpha\sigma^2)$.

Next, let h be any bounded continuous function on $\mathbb{R}$. Then, using (4.2)–(4.3) and (7.10),

$$\begin{aligned}&\mathbb{E}\bigl(h\bigl(\tilde{Y}_{n'}/\sqrt{n}\bigr) \mid S_n = n-1\bigr) \\ &= \frac{\mathbb{E}\sum_j h\bigl(\tilde{Y}_{n'}/\sqrt{n}\bigr)\mathbf{1}\{S_{n'}=j\}\mathbf{1}\{S_n - S_{n'} = n-1-j\}}{\mathbb{P}(S_n = n-1)} \\ &= \frac{\sum_j \mathbb{E}\bigl(h\bigl(\tilde{Y}_{n'}/\sqrt{n}\bigr)\mathbf{1}\{S_{n'}=j\}\bigr)\,\mathbb{P}\bigl(S_{n-n'} = n-1-j\bigr)}{\mathbb{P}(S_n = n-1)} \\ &= \sum_j \mathbb{E}\bigl(h\bigl(\tilde{Y}_{n'}/\sqrt{n}\bigr)\mathbf{1}\{S_{n'}=j\}\bigr)\Bigl(\frac{n}{n-n'}\Bigr)^{1/2}\bigl(e^{-(j+1-n')^2/(2(n-n')\sigma^2)} + o(1)\bigr) \\ &= (1-\alpha)^{-1/2}\,\mathbb{E}\Bigl(h\bigl(\tilde{Y}_{n'}/\sqrt{n}\bigr)e^{-(S_{n'}+1-n')^2/(2(n-n')\sigma^2)}\Bigr) + o(1) \\ &\to (1-\alpha)^{-1/2}\,\mathbb{E}\Bigl(h(W)e^{-Z^2/(2(1-\alpha)\sigma^2)}\Bigr) \\ &= \mathbb{E}\bigl(h(W)\bigr)(1-\alpha)^{-1/2}\,\mathbb{E}\, e^{-Z^2/(2(1-\alpha)\sigma^2)} = \mathbb{E}\, h(W),\end{aligned}$$

where the final equality follows by a simple calculation, or even more simply by using the special case $h \equiv 1$. Since h is arbitrary, this proves

$$\Bigl(\tilde{Y}_{n'}/\sqrt{n} \mid S_n = n-1\Bigr) \xrightarrow{\mathrm{d}} W \sim N\bigl(0, \alpha\gamma^2\bigr). \tag{7.11}$$

Next, conditioned on $S_n = n-1$ we have by (7.5), (7.9) and symmetry (for n so large that $n', n-n' > m$)

$$\begin{aligned}\sum_{i=1}^{n}\bigl(g(\xi_i,\dots,\xi_{i+m-1 \bmod n})-\mu\bigr)-\tilde{Y}_{n'} \\ &\overset{\mathrm{d}}{=} \sum_{i=1}^{n-n'}\bigl(g(\xi_i,\dots,\xi_{i+m-1})-\mu\bigr)+\frac{\mu}{\sigma^2}\bigl(S_{n-n'}-(n-1-n')\bigr) \\ &= \tilde{Y}_{n-n'}+\mu/\sigma^2. \end{aligned} \tag{7.12}$$

We may for notational convenience pretend that the equality in distribution (4.7) is an equality, and we then have, for each $\alpha \in (0,1)$, a decomposition

$$\frac{F(\mathcal{T}_n)-n\mu}{\sqrt{n}} = X'_{n,\alpha}+X''_{n,\alpha} \tag{7.13}$$

where $X'_{n,\alpha} = \bigl(\tilde{Y}_{n'}/\sqrt{n} \mid S_n = n-1\bigr)$ and, by (7.12),

$$X''_{n,\alpha} \overset{\mathrm{d}}{=} \left(\frac{\tilde{Y}_{n-n'}}{\sqrt{n}} \mid S_n = n-1\right)+\frac{\mu}{\sigma^2\sqrt{n}} = \left(\frac{\tilde{Y}_{n-n'}}{\sqrt{n}} \mid S_n = n-1\right)+o(1). \tag{7.14}$$

By (7.11), $X'_{n,\alpha} \overset{\mathrm{d}}{\longrightarrow} W'_\alpha \sim N\bigl(0,\alpha\gamma^2\bigr)$. Furthermore, $(n-n')/n \to 1-\alpha$, and thus by (7.11) applied to $1-\alpha$ instead of α, $X''_{n,\alpha} \overset{\mathrm{d}}{\longrightarrow} W''_\alpha \sim N\bigl(0,(1-\alpha)\gamma^2\bigr)$.

Now let $\alpha \to 1$ (along a sequence, if you like). Then $W'_\alpha \overset{\mathrm{d}}{\longrightarrow} N(0,\gamma^2)$ and $W''_\alpha \overset{\mathrm{p}}{\longrightarrow} 0$, and the conclusion (7.1) follows from (7.13), see e.g. [9, Theorem 4.2] or [32, Theorem 4.28]. □

8. Final proofs

Proof of Theorem 1.5. We have already proved part (i) in Section 5.

Futhermore, we have proved part (ii) in the special case of a functional f with finite support in Corollary 6.4 and Lemma 7.1. In general, we use a truncation.

We begin by verifying that γ^2 is finite, with the expectations in (1.17) absolutely convergent.

First, $\mathbb{E}\,|f(\mathcal{T})| < \infty$ by assumption, see also Remark 1.6. Similarly, by (4.13), since $\mathbb{E}\,f(\mathcal{T}_n)^2 = O(1)$ by (1.14),

$$\mathbb{E}\,f(\mathcal{T})^2 = \sum_{n=1}^{\infty}\pi_n\,\mathbb{E}\,f(\mathcal{T}_n)^2 \ll \sum_{n=1}^{\infty}\frac{\mathbb{E}\,f(\mathcal{T}_n)^2}{n^{3/2}} < \infty. \tag{8.1}$$

Hence $\operatorname{Var} f(\mathcal{T}) < \infty$.

To show that $\mathbb{E}\bigl|f(\mathcal{T})\bigl(F(\mathcal{T})-|\mathcal{T}|\mu\bigr)\bigr| < \infty$, note that by Theorem 6.7 and (1.13),

$$\mathbb{E}\bigl(F(\mathcal{T}_n)-n\mu\bigr)^2 = \operatorname{Var} F(\mathcal{T}_n)+\bigl(\mathbb{E}\,F(\mathcal{T}_n)-n\mu\bigr)^2 = O(n). \tag{8.2}$$

Thus, using the Cauchy–Schwarz inequality, (4.13), (8.2) and (1.15),

$$\begin{aligned}\mathbb{E}\bigl|f(\mathcal{T})\bigl(F(\mathcal{T})-|\mathcal{T}|\mu\bigr)\bigr| &= \sum_{n=1}^{\infty}\pi_n\,\mathbb{E}\bigl|f(\mathcal{T}_n)\bigl(F(\mathcal{T}_n)-n\mu\bigr)\bigr| \\ &\leqslant \sum_{n=1}^{\infty}\pi_n\sqrt{\mathbb{E}\,f(\mathcal{T}_n)^2}\sqrt{\mathbb{E}\bigl(F(\mathcal{T}_n)-n\mu\bigr)^2} \\ &\ll \sum_{n=1}^{\infty} n^{-3/2}\sqrt{\mathbb{E}\,f(\mathcal{T}_n)^2}\,n^{1/2} < \infty. \end{aligned} \tag{8.3}$$

Hence, γ^2 is well-defined by (1.17), and finite.

Define the truncation

$$f^{(N)}(T) := \sum_{k=1}^{N} f_k(T) = f(T)\mathbf{1}\{|T| \leqslant N\} \tag{8.4}$$

and the corresponding sum $F^{(N)}(T)$. Furthermore, let $\mu^{(N)} := \mathbb{E}\,f^{(N)}(\mathcal{T})$ and

$$(\gamma^{(N)})^2 := 2\,\mathbb{E}\Bigl(f^{(N)}(\mathcal{T})\bigl(F^{(N)}(\mathcal{T})-|\mathcal{T}|\mu^{(N)}\bigr)\Bigr) - \operatorname{Var} f^{(N)}(\mathcal{T}) - (\mu^{(N)})^2/\sigma^2. \tag{8.5}$$

Then $\mu^{(N)} \to \mu$ as $N\to\infty$ by dominated convergence, and similarly, using (8.1), $\operatorname{Var} f^{(N)}(\mathcal{T}) \to \operatorname{Var} f(\mathcal{T})$ and, using (8.3),

$$\mathbb{E}\bigl|f^{(N)}(\mathcal{T})\bigl(F^{(N)}(\mathcal{T})-|\mathcal{T}|\mu\bigr)\bigr| \to \mathbb{E}\bigl|f(\mathcal{T})\bigl(F(\mathcal{T})-|\mathcal{T}|\mu\bigr)\bigr|.$$

Finally, using (1.14) and (4.13),

$$\begin{aligned}&\mathbb{E}\bigl|f^{(N)}(\mathcal{T})|\mathcal{T}|(\mu-\mu^{(N)})\bigr| = \mathbb{E}\bigl|f^{(N)}(\mathcal{T})|\mathcal{T}|\bigr|\cdot\bigl|\mu-\mu^{(N)}\bigr| \\ &\qquad\leqslant \sum_{k=1}^{N}\pi_k k\,\mathbb{E}\,|f(\mathcal{T}_k)|\cdot\sum_{k=N+1}^{\infty}\pi_k|\,\mathbb{E}\,f(\mathcal{T}_k)| = O\bigl(N^{1/2}\bigr)\cdot o\bigl(N^{-1/2}\bigr) = o(1),\end{aligned}$$

as $N\to\infty$. Combining these estimates we see that $(\gamma^{(N)})^2 \to \gamma^2$.

Since $f^{(N)}$ has finite support, Corollary 6.4 yields $\operatorname{Var} F^{(N)}(\mathcal{T}_n)/n \to (\gamma^{(N)})^2$ as $n\to\infty$, for every fixed N. Furthermore, Theorem 6.7 applied to $f - f^{(N)} = \sum_{k=N+1}^{\infty} f_k$ shows that

$$n^{-1/2}\operatorname{Var}\bigl(F(\mathcal{T}_n)-F^{(N)}(\mathcal{T}_n)\bigr)^{1/2} \ll \sup_{k>N}\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2} + \sum_{k=N+1}^{\infty}\frac{\sqrt{\mathbb{E}\,f(\mathcal{T}_k)^2}}{k}, \tag{8.6}$$

uniformly in n and N. The right-hand side is independent of n and tends to 0 as $N\to\infty$, and it follows by Minkowski's inequality and a standard 3ε-argument (i.e., because a uniform limit of convergent sequences is convergent) that $n^{-1/2}\operatorname{Var}(F(\mathcal{T}_n))^{1/2} \to \lim_{N\to\infty}\gamma^{(N)} = \gamma$, showing (1.16).

Similarly, Lemma 7.1 applies to each $f^{(N)}$, and the uniform estimate (8.6) implies that we can let $N\to\infty$ and conclude (1.18), see e.g. [9, Theorem 4.2] or [32, Theorem 4.28] again. □

Proof of Theorem 1.13. Suppose first that f is bounded and local. By replacing $f(T)$ by $f(T)-\mathbb{E} f(\hat{\mathcal{T}})$, which does not change $F(T)-|T|\mu$, we may assume that $\mathbb{E} f(\hat{\mathcal{T}})=0$. In this case Lemma 5.9 yields $\mathbb{E} f(\mathcal{T}_n)=O\bigl(n^{-1/2}\bigr)$ and in particular $\mathbb{E} f(\mathcal{T}_n)\to 0$ and $\sum_n |\mathbb{E} f(\mathcal{T}_n)|/n<\infty$. Hence the conditions of the second part are satisfied, so it suffices to prove it.

Hence, assume now that f is bounded and weakly local, and that $\mathbb{E} f(\mathcal{T}_n)\to 0$ and $\sum_n |\mathbb{E} f(\mathcal{T}_n)|/n<\infty$. We use truncation as in the proof of Theorem 1.5 and note first that Theorem 6.12 applied to $f-f^{(N)}$ yields

$$n^{-1/2}\operatorname{Var}\bigl(F(\mathcal{T}_n)-F^{(N)}(\mathcal{T}_n)\bigr)^{1/2}\ll N^{-1/8}\sup_{k>N}A_k+\sup_{k>N}|\mu_k|+\sum_{k>N}|\mu_k|/k \tag{8.7}$$

where the right-hand side is independent of n and tends to 0 as $N\to\infty$. (Note that $f-f^{(N)}$ is weakly local with the same cut-off as f for all N.)

Similarly, if $M\geqslant N$, then Corollary 6.4 and Theorem 6.12, together with Minkowski's inequality, show, with $\gamma^{(N)}$ given by (8.5),

$$\begin{aligned}\bigl|\gamma^{(M)}-\gamma^{(N)}\bigr| &\leqslant \limsup_{n\to\infty} n^{-1/2}\operatorname{Var}\bigl(F^{(M)}(\mathcal{T}_n)-F^{(N)}(\mathcal{T}_n)\bigr)^{1/2}\\ &\ll N^{-1/8}\sup_{k>N}A_k+\sup_{k>N}|\mu_k|+\sum_{k>N}|\mu_k|/k. \end{aligned}\tag{8.8}$$

Consequently $\bigl(\gamma^{(N)}\bigr)_N$ is a Cauchy sequence so $\gamma^{(N)}\to\gamma$ for some $\gamma<\infty$.

The rest of the proof is the same as for Theorem 1.5. ◻

Proof of Corollary 1.8. For any finite set $T_1,\dots,T_m$ of distinct trees and real numbers $a_1,\dots,a_m$, apply Theorem 1.5 to $f(T):=\sum_{i=1}^m a_i\mathbf{1}\{T=T_i\}$ and note that then $F(T)=\sum_{i=1}^m a_i n_{T_i}(T)$. The assumptions (1.14)–(1.15) hold trivially since f has finite support. We have $\mu=\mathbb{E} f(\mathcal{T})=\sum_{i=1}^n a_i\,\mathbb{P}(\mathcal{T}=T_i)=\sum_{i=1}^m a_i\pi_{T_i}$ and a simple calculation shows that (1.17) yields $\gamma^2=\sum_{i,j=1}^m a_ia_j\gamma_{T_i,T_j}$. The results now follow from Theorem 1.5 (or directly from Corollary 6.4 and Lemma 7.1), using the Cramér–Wold device. ◻

Proof of Theorem 1.9. We use the notation in the proof of Lemma 7.1 (but now simply taking $n'=n$ so $\alpha=1$). We showed in (7.6) and (7.10) asymptotic normality of Y_n, S_n and $\tilde{Y}_n$; a simple (and well-known) calculation shows that also the (co)variances converge: $\operatorname{Var}(Y_n)/n\to\beta^2$, $\operatorname{Cov}(Y_n,S_n)/n\to\rho$, $\operatorname{Var}(S_n)/n\to\sigma^2$ and, recalling (7.7)–(7.8),

$$\frac{\operatorname{Var}(\tilde{Y}_n)}{n}\to\beta^2-\frac{\mu^2}{\sigma^2}=\gamma^2=0. \tag{8.9}$$

However, by (7.9) and (7.5),

$$\tilde{Y}_n:=\sum_{i=1}^n\tilde{g}(\xi_i,\dots,\xi_{i+m-1}), \tag{8.10}$$

where

$$\tilde{g}(\xi_i,\dots,\xi_{i+m-1}):=g(\xi_i,\dots,\xi_{i+m-1})-\mu+\frac{\mu}{\sigma^2}(\xi_i-1). \tag{8.11}$$

The sequence $X_i := \tilde{g}(\xi_i, \dots, \xi_{i+m-1})$ is strictly stationary and $(m-1)$-dependent, with mean $\mathbb{E}\, X_i = 0$ (by (7.4)) and finite variance. If the partial sums $\tilde{Y}_n$ satisfy (8.9), with limit $\gamma^2 = 0$, then as a consequence of a theorem by Leonov [40], in the version given by Bradley [10, Theorem 8.6], see Janson [31, Theorem 1.6] for details, there exists a function $h : \mathbb{N}^{m-1} \to \mathbb{R}$ such that

$$\tilde{g}(\xi_i, \dots, \xi_{i+m-1}) = h(\xi_{i+1}, \dots, \xi_{i+m-1}) - h(\xi_i, \dots, \xi_{i+m-2}) \quad \text{a.s.} \tag{8.12}$$

and thus by (8.10),

$$\tilde{Y}_n = h(\xi_{n+1}, \dots, \xi_{n+m-1}) - h(\xi_1, \dots, \xi_{m-1}) \quad \text{a.s.} \tag{8.13}$$

In particular, $\tilde{Y}_n$ depends a.s. only on $\xi_1, \dots, \xi_{m-1}$ and $\xi_{n+1}, \dots, \xi_{n+m-1}$, but not on n or $\xi_m, \dots, \xi_n$.

Consider now first case (i). Take $j > 0$ with $p_j > 0$ and consider the case $\xi_i = j$ for all $i < n + m$. Then no substring of $\xi_1, \dots, \xi_{n+m-1}$ is the degree sequence of a tree. Thus, recalling (7.2), $g(\xi_i, \dots, \xi_{i+m-1}) = 0$ for every i so (8.11) and (8.10) yield $\tilde{g}(\xi_i, \dots, \xi_{i+m-1}) = \mu\big((j-1)/\sigma^2 - 1\big)$ and

$$\tilde{Y}_n = n\mu\big((j-1)/\sigma^2 - 1\big). \tag{8.14}$$

Since this vanishes for any n by (8.13), $\mu\big((j-1)/\sigma^2-1\big) = 0$. By assumption, there exist at least two different such j, and thus $\mu = 0$. Hence, (8.11) simplifies to $\tilde{g}(\xi_i, \dots, \xi_{i+m-1}) = g(\xi_i, \dots, \xi_{i+m-1})$, and thus

$$\tilde{Y}_n = \sum_{i=1}^{n} g(\xi_i, \dots, \xi_{i+m-1}). \tag{8.15}$$

Next consider the case $\xi_i = 0$ for all $i < n + m$. Since (0) is the degree sequence of the tree $\bullet$ of size 1, (7.2) yields $g(0, \dots, 0) = f_1(0) = f(\bullet)$. Hence, (8.15) yields $\tilde{Y}_n = nf(\bullet)$. Since this vanishes, by (8.13) again, we must have $f(\bullet) = 0$.

Suppose, in order to obtain a contradiction, that $f(\mathcal{T})$ does not vanish a.s. We have, for some $N \geqslant 1$, some distinct trees $T_1, \dots, T_N$ and some real numbers $a_1, \dots, a_N$,

$$f(T) = \sum_{i=1}^{N} a_i n_{T_i}(T), \tag{8.16}$$

where we may assume that $a_i \neq 0$ and $\mathbb{P}(\mathcal{T} = T_i) > 0$ for all i (otherwise we eliminate the offending terms). We may also suppose that $T_1, \dots, T_N$ are ordered with $|T_1| \leqslant |T_2| \leqslant \dots$; this implies that $f(T) = 0$ for every proper subtree T of T_1. Let T_1 have degree sequence $(d_1, \dots, d_\ell)$, and consider now the case $\xi_{m+j} = d_j$, $j = 1, \dots, \ell$, and $\xi_i = 0$ for $i \leqslant m$ and $m+\ell < i < n+m$, for $n \geqslant m + \ell$. Since $f(T) = 0$ for all proper subtrees of T_1 and $f(0) = 0$, the only non-zero contribution to $\tilde{Y}_n$ is by (8.15), (7.2) and (8.16),

$$f(\xi_{m+1}, \dots, \xi_{m+\ell}) = f(d_1, \dots, d_\ell) = f(T_1) = a_1. \tag{8.17}$$

Hence $\tilde{Y}_n = a_1 \neq 0$, which contradicts (8.13). This contradiction proves $f(\mathcal{T}) = 0$ a.s., which implies $F(\mathcal{T}) = 0$ and $F(\mathcal{T}_n) = 0$ a.s. for every n, completing the proof of (i).

Now consider (ii), with only p_0 and p_r non-zero. (Since $\mathbb{E}\,\xi = 1$, we have $p_r = 1 - p_0 = 1/r$.) This is the case of full r-ary trees, and the random tree $\mathcal{T}_n$ has $(n-1)/r$ nodes of outdegree r and $n - (n-1)/r$ leaves. Thus the choice $f(T) = \mathbf{1}\{T = \bullet\}$, when $F(T) = n_\bullet(T)$ is the number of leaves in T, yields $\operatorname{Var} F(\mathcal{T}_n) = 0$ so $\gamma^2 = 0$, see Example 2.1.

If f is any functional with finite support such that $\gamma^2 = 0$, we may replace $f(T)$ by $f(T) - f(\bullet)\mathbf{1}\{T = \bullet\}$ without changing $\operatorname{Var} F(\mathcal{T}_n)$, so we still have $\gamma^2 = 0$. Hence we may assume $f(\bullet) = 0$. If we now consider the case $\xi_i = 0$, $i < n+m$, then by (7.2), $g(\xi_i, \dots, \xi_{i+m-1}) = f(0) = 0$ and thus (8.11) yields $\tilde{g}(\xi_i, \dots, \xi_{i+m-1}) = -\mu - \mu/\sigma^2$. Hence, (8.10) yields $\tilde{Y}_n = -n\mu(1 + \sigma^{-2})$, and (8.13) implies that this vanishes, and thus $\mu = 0$. The rest of the proof is the same as for (i). ☐

Proof of (2.6). This is a minor variation of arguments in Sections 5–6, using the special simple structure of f. We omit some details.

If T has degree sequence $(d_1, \dots, d_n)$, then $n_r(T) = \sum_{i=1}^n \mathbf{1}\{d_i = r\}$. This yields, arguing as for (4.7),

$$n_r(\mathcal{T}_n) \overset{\mathrm{d}}{=} \left(\sum_{i=1}^n \mathbf{1}\{\xi_i = r\} \;\Big|\; S_n = n-1 \right), \tag{8.18}$$

jointly for all $r \geqslant 0$. This yields, arguing as in (5.3) and (6.8),

$$\mathbb{E}\, n_r(\mathcal{T}_n) = n\,\mathbb{P}\bigl(\xi_1 = r \mid S_n = n-1\bigr) = np_r \frac{\mathbb{P}(S_{n-1} = n-1-r)}{\mathbb{P}(S_n = n-1)} \tag{8.19}$$

and, for any integers $r, s \geqslant 0$,

$$\begin{aligned}\mathbb{E}\bigl(n_r(\mathcal{T}_n) n_s(\mathcal{T}_n)\bigr) &= \delta_{rs}\,\mathbb{E}\, n_r(\mathcal{T}_n) + n(n-1)\,\mathbb{P}\bigl(\xi_1 = r, \xi_2 = s \mid S_n = n-1\bigr) \\ &= \delta_{rs}\,\mathbb{E}\, n_r(\mathcal{T}_n) + n(n-1) p_r p_s \frac{\mathbb{P}(S_{n-2} = n-1-r-s)}{\mathbb{P}(S_n = n-1)}.\end{aligned}$$

Hence

$$\begin{gathered}\operatorname{Cov}\bigl(n_r(\mathcal{T}_n), n_s(\mathcal{T}_n)\bigr) = \delta_{rs}\,\mathbb{E}\, n_r(\mathcal{T}_n) - \frac{1}{n}\,\mathbb{E}\, n_r(\mathcal{T}_n)\,\mathbb{E}\, n_s(\mathcal{T}_n) + n(n-1) p_r p_s \times \\ \left(\frac{\mathbb{P}(S_{n-2} = n-1-r-s)}{\mathbb{P}(S_n = n-1)} - \frac{\mathbb{P}(S_{n-1} = n-1-r)}{\mathbb{P}(S_n = n-1)} \frac{\mathbb{P}(S_{n-1} = n-1-s)}{\mathbb{P}(S_n = n-1)} \right).\end{gathered} \tag{8.20}$$

For the mean, (8.19) and (4.2) yield

$$\mathbb{E}\, n_r(\mathcal{T}_n)/n \to p_r, \tag{8.21}$$

cf. (2.5). (The argument is simpler than in Section 5 since we consider a fixed r.)

For the covariance, we argue as in the proof of Lemma 6.2 and consider

$$\begin{aligned}
&\mathbb{P}(S_{n-2}=n-r-s-1)\mathbb{P}(S_n=n-1)\\
&\qquad\qquad\qquad - \mathbb{P}(S_{n-1}=n-r-1)\,\mathbb{P}(S_{n-1}=n-s-1)\\
&\quad=\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-2}(t)e^{\mathrm{i}(r+s-1)t}\,\mathrm{d}t\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n}(u)e^{\mathrm{i}u}\,\mathrm{d}u\\
&\qquad\qquad - \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-1}(t)e^{\mathrm{i}rt}\,\mathrm{d}t\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-1}(u)e^{\mathrm{i}su}\,\mathrm{d}u\\
&\quad=\frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\tilde\varphi^{n-2}(t)e^{\mathrm{i}t}\tilde\varphi^{n-2}(u)e^{\mathrm{i}u}\bigl(e^{\mathrm{i}(r-1)t}\tilde\varphi(u)-\tilde\varphi(t)e^{\mathrm{i}(r-1)u}\bigr)\times\\
&\qquad\qquad\qquad\bigl(e^{\mathrm{i}(s-1)t}\tilde\varphi(u)-\tilde\varphi(t)e^{\mathrm{i}(s-1)u}\bigr)\,\mathrm{d}t\,\mathrm{d}u. \qquad (8.22)
\end{aligned}$$

We have

$$e^{\mathrm{i}(r-1)t}\tilde\varphi(u)-\tilde\varphi(t)e^{\mathrm{i}(r-1)u}=\mathrm{i}(r-1)(t-u)+O\bigl(t^2+u^2\bigr),$$

and by a change of variables as in (6.17)–(6.18), the final double integral in (8.22) is $\sim -(r-1)(s-1)/(2\pi\sigma^4 n^2)$. Hence (8.20) yields, using also (4.3) and (8.21),

$$\frac{\operatorname{Cov}\bigl(n_r(\mathcal{T}_n),n_s(\mathcal{T}_n)\bigr)}{n}\to\delta_{rs}p_r-p_rp_s-\frac{(r-1)(s-1)}{\sigma^2}p_rp_s, \qquad (8.23)$$

showing (2.3) and (2.6). □

References

[1] Louigi Addario-Berry, Luc Devroye & Svante Janson, Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.* **41** (2013), no. 2, 1072–1087.

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

[3] David Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[4] David Aldous, The continuum random tree III. *Ann. Probab.* **21** (1993), no. 1, 248–289.

[5] David Aldous & Jim Pitman, Tree-valued Markov chains derived from Galton–Watson processes. *Ann. Inst. H. Poincaré Probab. Statist.* **34** (1998), no. 5, 637–686.

[6] Alexei B. Aleksandrov, Svante Janson, Vladimir V. Peller & Richard Rochberg, An interesting class of operators with unusual Schatten-von Neumann behavior. *Function Spaces, Interpolation Theory and Related Topics (Lund, 2000)*, 61–149, de Gruyter, Berlin, 2002.

[7] Krishna B. Athreya & Peter E. Ney, *Branching Processes.* Springer-Verlag, Berlin, 1972.

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

[9] Patrick Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[10] Richard C. Bradley, *Introduction to Strong Mixing Conditions. Vol. 1–3.* Kendrick Press, Heber City, UT, 2007.

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

[12] Luc Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer, Berlin, 1998, pp. 249–314.

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

[14] P. H. Diananda, The central limit theorem for m-dependent variables. *Proc. Cambridge Philos. Soc.* **51** (1955), 92–95.

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

[16] Meyer Dwass, The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6** (1969), 682–686.

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

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

[19] 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.

[20] 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.

[21] Philippe Flajolet & Robert Sedgewick, *Analytic Combinatorics.* Cambridge Univ. Press, Cambridge, UK, 2009.

[22] G. H. Hardy, J. E. Littlewood & G. Pólya, *Inequalities.* 2nd ed., Cambridge, at the University Press, 1952.

[23] Wassily Hoeffding & Herbert Robbins, The central limit theorem for dependent random variables. *Duke Math. J.* **15** (1948), 773–780.

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

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

[26] 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.

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

[28] Svante Janson, The Wiener index of simply generated random trees. *Random Struct. Alg.* **22** (2003), no. 4, 337–358.

[29] Svante Janson, Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29** (2006), no. 2, 139–179.

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

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

[32] Olav Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[33] Johannes Henricus Bernardus Kemperman, *The General One-Dimensional Random Walk with Absorbing Barriers with Applications to Sequential Analysis.* Thesis. Excelsiors Foto-Offset, The Hague, 1950.

[34] J. H. B. Kemperman, *The Passage Problem for a Stationary Markov Chain.* University of Chicago Press, Chicago, IL (1961).

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

[36] Harry Kesten, Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.* **22** (1986), no. 4, 425–487.

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

[38] 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.

[39] 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.

[40] V. P. Leonov, On the dispersion of time-dependent means of a stationary stochastic process. (Russian.) *Teor. Verojatnost. i Primenen.* **6** (1961), 93–101; English transl.: *Theory Probab. Appl.* **6** (1961), no. 1, 87–93

[41] Russell Lyons, Robin Pemantle & Yuval Peres Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** (1995), no. 3, 1125–1138.

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

[43] Vladimir G. Maz'ya & Igor E. Verbitsky, The Schrödinger operator on the energy space: boundedness and compactness criteria. *Acta Math.* **188** (2002), no. 2, 263–302.

[44] A. Meir & J.W. Moon, On the altitude of nodes in random trees. *Canad. J. Math.* **30** (1978), 997–1015.

[45] 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.

[46] *NIST Digital Library of Mathematical Functions.* http://dlmf.nist.gov/
[47] Richard Otter, The multiplicative process. *Ann. Math. Statistics* **20** (1949), 206–224.
[48] Valentin V. Petrov, *Sums of Independent Random Variables.* Springer-Verlag, Berlin, 1975.
[49] Jim Pitman, Enumerations of trees and forests related to branching processes and random walks. *Microsurveys in Discrete Probability (Princeton, NJ, 1997)*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 41, Amer. Math. Soc., Providence, RI, 1998, pp. 163–180.
[50] Anders Rygh Swensen, A note on convergence of distributions of conditional moments. *Scand. J. Statist.* **10** (1983), no. 1, 41–44.
[51] Stephan Wagner, Additive tree functionals with small toll functions and subtrees of random trees. *Discrete Math. Theor. Comput. Sci. Proc.*, **AQ** (2012), 67–80.
[52] Sandy Lew Zabell, *A limit theorem for conditional expectations with applications to probability theory and statistical mechanics.* Ph.D. thesis, Harvard University, 1975.
[53] Sandy L. Zabell, Rates of convergence for conditional expectations. *Ann. Probab.* **8** (1980), no. 5, 928–941.
[54] S. L. Zabell, A limit theorem for expectations conditional on a sum. *J. Theoret. Probab.* **6** (1993), no. 2, 267–283.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se

URL: http://www2.math.uu.se/∼svante/