# CENTRAL LIMIT THEOREMS FOR ADDITIVE FUNCTIONALS AND FRINGE TREES IN TRIES

SVANTE JANSON

Abstract. We give general theorems on asymptotic normality for additive functionals of random tries generated by a sequence of independent strings. These theorems are applied to show asymptotic normality of the distribution of random fringe trees in a random trie. Formulas for asymptotic mean and variance are given. In particular, the proportion of fringe trees of size $k$ (defined as number of keys) is asymptotically, ignoring oscillations, $c/(k(k-1))$ for $k \geq 2$, where $c = 1/(1+H)$ with $H$ the entropy of the letters. Another application gives asymptotic normality of the number of $k$-protected nodes in a random trie. For symmetric tries, it is shown that the asymptotic proportion of $k$-protected nodes (ignoring oscillations) decreases geometrically as $k \to \infty$.

## 1. Introduction

We consider random tries constructed from a number of random (infinite) strings with letters in a fixed finite alphabet $\mathcal{A}$. (The most important case is $\mathcal{A} = \{0, 1\}$, and the reader may for simplicity assume this without essential loss.) See Section 2 for the definition of tries and other definitions of terms used here in the introduction.

We assume throughout the paper that the strings are i.i.d., and moreover, that the individual letters in the strings are i.i.d. The number of strings will be either fixed, or a Poisson variable; we refer to these as the *fixed $n$ model* (where $n$ is the number of strings) and the *Poisson model*.

As has been well-known since at least [15; 30], for some sets of letter probabilities (in particular, for the symmetric case with equal probabilities), there are typically (numerically small) oscillations in the asymptotics of both mean and variance for functionals of random tries; nevertheless asymptotic normality holds with suitable normalizations. The cases where oscillations occur are well understood, either from the location of poles of Mellin transforms, see e.g. [8; 16], or from (the arithmetic case of) renewal theory, see [26; 18].

One of our main results is a central limit theorem (i.e., asymptotic normality) of this type, including possible oscillations, for additive functionals of tries under rather weak conditions, for both the fixed $n$ model and the

*Date*: 5 March 2020; revised 6 March 2020.
2000 *Mathematics Subject Classification.* 60C05, 05C05, 68P05.
Partly supported by the Knut and Alice Wallenberg Foundation.

Poisson model (Theorem 3.9). This theorem assumes that the toll function is bounded (together with another technical condition). We give, as a corollary, a law of large numbers (Theorem 3.12).

In Section 4, several applications of these theorems are given. In particular, we study random fringe trees of tries, and show central limit theorems for the distribution of them. We study also the number of $k$-protected nodes in tries, $k \geq 2$, and prove a central limit theorem. We show also that for symmetric tries, ignoring oscillations, the expected number of $k$-protected nodes decreases geometrically as $k \to \infty$. We give also a couple of other applications.

Our method of proof consists of the following three separate parts:

(1) To prove asymptotic normality for the Poisson model, we use the independence of different branches in the trie and the classical central limit theorem for sums of independent random variables. The proof requires several estimates, including a moment estimate that is proved by induction using a less common version of Rosenthal's inequality (Lemma 6.4).
(2) To depoissonize, i.e., transfer results to the fixed $n$ model, we use here a novel approach, using a conditional limit theorem by Nerman [27]. The main condition for this theorem is that the functionals we consider are increasing, or at least the difference of two increasing functionals.
(3) To find asymptotic means and variances, we use results from [18] based on renewal theory, see also [26].

Note that there are several earlier papers on asymptotic normality for tries, where all three steps have been proved by detailed analyses of generating functions. That is a wonderful method, but the method used here avoids the necessity to estimate the generating functions in the complex plane; this may be useful or convenient in some applications. Furthermore, our method is easily adapted to more general sources of random strings, see Remark 1.1. The reader is encouraged to compare, and perhaps combine, the methods for future work.

We state the results of steps (1) and (2) above as general central limit theorems, in several versions (Theorems 5.3–5.8, with proofs in Section 6), where the toll function may be unbounded but we assume some technical conditions on moments of the additive functional and its toll functional. Then, as step (3), we prove separately (in Section 7) Theorem 3.1 on mean and variance of additive functionals. This is based on a theorem from [18], which for convenience is stated, and somewhat extended, in Appendix A. Finally, Theorem 3.9 follows by combining Theorem 3.1 and the general central limit theorems. (This proof is in Section 8.)

One reason for this organization is that the central limit theorems and the moment asymptotics are proved by quite different methods, and we find it instructive to present them separately, and not only their combination Theorem 3.9. This also enables us to present somewhat more general results, as said above.

**Remark 1.1.** The method of proof of normality (steps (1) and (2) above) applies, under suitable conditions, also to random strings where the letters are not independent, for example strings from a Markov source, or the bit expansions of random numbers with a non-uniform distribution on $(0, 1)$. (We still assume that different strings are i.i.d.) This will be studied elsewhere. □

## 2. Preliminaries

2.1. **Some general notation.** We use $\overset{\mathrm{p}}{\longrightarrow}$ and $\overset{\mathrm{d}}{\longrightarrow}$ to denote convergence in probability and distribution, respectively, of random variables. $\overset{\mathrm{d}}{=}$ denotes equality in distribution.

$(X \mid \mathcal{E})$ denotes the random variable $X$ conditioned on the event $\mathcal{E}$.

For a random variable $X$ and $r > 0$, $\|X\|_r := (\mathbb{E}\,|X|^r)^{1/r}$, the $L^r$ norm.

$C$ denotes various unimportant constants, possibly different at different occurences. We sometimes for clarity write $C_1, C_2, \dots$, and we use $C_\lambda$ for a "constant" that depends on $\lambda$.

We use standard $o$ and $O$ notation, for sequences and functions of a real variable; note that $O$ is used both in a global and an asymptotic sense: for example, $f(x) = O(g(x))$ for $x \in S$ means that $|f(x)| \le Cg(x)$ for all $x \in S$ (equivalently, if $g(x) > 0$ in $S$, $f(x)/g(x)$ is bounded in $S$), while $f(x) = O(g(x))$ as $x \to \infty$ means that $|f(x)| \le Cg(x)$ for large $x$. For positive functions or sequences we also use the notations $\Omega$ and $\Theta$: $f(x) = \Omega(g(x))$ as $x \to \infty$ means that $f(x) \ge cg(x)$ for some $c > 0$ and large $x$, or, equivalently, $g(x) = O(f(x))$ as $x \to \infty$; $f(x) = \Theta(g(x))$ means $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$, and similarly for sequences.

For $x \in \mathbb{R}^d$, $|x|$ denotes the usual Euclidean norm. (Any other norm would do as well.)

For $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the largest integer $\le x$.

log denotes the natural logarithm.

2.2. **Strings.** We consider strings with letters in a finite alphabet $\mathcal{A}$. ($\mathcal{A}$ is fixed throughout the paper.) Let $\mathcal{A}^* := \bigcup_{n=0}^{\infty} \mathcal{A}^n$, the set of finite strings from $\mathcal{A}$. The empty string is denoted by $\boldsymbol{\epsilon}$.

We write $\boldsymbol{\alpha} \preceq \boldsymbol{\beta}$ if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are two strings and $\boldsymbol{\alpha}$ is a prefix of $\boldsymbol{\beta}$.

The tries will be constructed from $n$ random infinite strings $\Xi^{(1)}, \Xi^{(2)}, \dots, \Xi^{(n)}$, where $\Xi^{(k)} = \xi_1^{(k)} \xi_2^{(k)} \cdots$ with letters $\xi_i^{(k)} \in \mathcal{A}$. (We may drop the superscript and write $\Xi = \xi_1 \xi_2 \cdots$ for a generic string in the sequence.) We suppose that the strings $\Xi^{(k)}$ are independent, and furthermore that the individual letters $\xi_i^{(k)}$ are i.i.d. We thus assume throughout the paper that we are given a probability distribution $\mathbf{p} = (p_\alpha)_{\alpha \in \mathcal{A}}$, and that

$$\mathbb{P}(\xi_i^{(k)} = \alpha) = p_\alpha, \qquad \alpha \in \mathcal{A}. \tag{2.1}$$

To avoid trivialities, we assume that each $p_\alpha > 0$ (otherwise we may reduce $\mathcal{A}$), and that $|\mathcal{A}| > 1$, and thus each $p_\alpha < 1$. We let $p_{\min} := \min_\alpha p_\alpha$ and $p_{\max} := \max_\alpha p_\alpha$, and note that $0 < p_{\min} \le p_{\max} < 1$.

The *entropy* $H$ is defined by

$$H := -\sum_{\alpha\in\mathcal{A}} p_\alpha \log p_\alpha > 0. \tag{2.2}$$

Given a finite string $\alpha_1 \cdots \alpha_m \in \mathcal{A}^*$, let $P(\alpha_1 \cdots \alpha_m)$ be the probability that the random string $\Xi$ has prefix $\alpha_1 \cdots \alpha_m$, i.e., that $\xi_i = \alpha_i$ for $i \le m$. In particular, for a single letter, $P(\alpha) = p_\alpha$, and in general

$$P(\alpha_1 \cdots \alpha_m) = \prod_{i=1}^m p_{\alpha_i}. \tag{2.3}$$

For later use we define, for complex $s \in \mathbb{C}$,

$$\rho(s) := \sum_{\alpha\in\mathcal{A}} p_\alpha^s \tag{2.4}$$

and note that (2.3) implies that, for any $m \ge 0$,

$$\sum_{|\boldsymbol{\alpha}|=m} P(\boldsymbol{\alpha})^s = \sum_{\alpha_1,\dots,\alpha_m\in\mathcal{A}^*} \prod_{i=1}^m p_{\alpha_i}^s = \rho(s)^m. \tag{2.5}$$

For any real $r > 1$, we have

$$\rho(r) = \sum_{\alpha\in\mathcal{A}} p_\alpha^r < \sum_{\alpha\in\mathcal{A}} p_\alpha = 1, \tag{2.6}$$

and thus by (2.5)

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} P(\boldsymbol{\alpha})^r = \sum_{m=0}^\infty \sum_{|\boldsymbol{\alpha}|=m} P(\boldsymbol{\alpha})^r = \sum_{m=0}^\infty \rho(r)^m < \infty. \tag{2.7}$$

Furthermore, we note that

$$\frac{\mathrm{d}}{\mathrm{d}s}\rho(s)\big|_{s=1} = \sum_{\alpha\in\mathcal{A}} p_\alpha \log p_\alpha = -H. \tag{2.8}$$

2.3. **Trees.** A *leaf* in a rooted tree is a node without children; leaves are also called *external nodes*, while the remaining nodes are called *internal nodes*.

Let $T_\infty$ be the infinite $|\mathcal{A}|$-ary tree where the nodes are the finite strings $\boldsymbol{\alpha} \in \mathcal{A}^*$; the root is the empty string $\boldsymbol{\epsilon}$, and the children of a node $\boldsymbol{\alpha}$ are the nodes $\boldsymbol{\alpha}\gamma$ with $\gamma \in \mathcal{A}$. Hence $\boldsymbol{\alpha}$ is a (strict) ancestor of $\boldsymbol{\beta}$ if and only if $\boldsymbol{\alpha} \prec \boldsymbol{\beta}$ (i.e., $\boldsymbol{\alpha}$ is a strict prefix of $\boldsymbol{\beta}$).

A finite $|\mathcal{A}|$-ary tree is a finite subtree of $T_\infty$ containing its root $\boldsymbol{\epsilon}$; for convenience we regard also the empty tree $\emptyset$ with no nodes as a finite $|\mathcal{A}|$-ary tree. Let $\overline{\mathfrak{T}}$ be the countable set of all finite $|\mathcal{A}|$-ary trees, and let $\overline{\mathfrak{T}}_+ := \overline{\mathfrak{T}} \setminus \{\emptyset\}$, the subset of nonempty trees.

We may identify trees in $\overline{\mathfrak{T}}$ with their sets of nodes, and we write $|T|$ for the number of nodes in $T$; we denote the numbers of internal and external nodes (= leaves) by $|T|_{\mathsf{i}}$ and $|T|_{\mathsf{e}}$, respectively; thus $|T| = |T|_{\mathsf{i}} + |T|_{\mathsf{e}}$. Let $\overline{\mathfrak{T}}_n := \{T \in \overline{\mathfrak{T}} : |T|_{\mathsf{e}} = n\}$, the set of finite $|\mathcal{A}|$-ary trees with exactly $n$ leaves. (Note that we in the present paper thus count the size by the number of leaves; this is natural in the context of tries.)

Let $\bullet \in \overline{\mathfrak{T}}$ denote the tree consisting of only the root $\boldsymbol{\epsilon}$. Thus $|\bullet| = |\bullet|_{\mathsf{e}} = 1$ and $\bullet \in \overline{\mathfrak{T}}_1$.

2.4. **Tries.** A *trie* (for a given alphabet $\mathcal{A}$) is an $|\mathcal{A}|$-ary tree that is constructed in the following way from a set of $n \geq 0$ distinct strings in $\mathcal{A}^\infty$, see e.g. [22, Section 6.3], [31, Section 1.1] and [6, Section 7.1]. If $n = 0$, the trie is defined to be the empty tree $\emptyset$. Otherwise, we begin with a root, and put every string in the root. If $n = 1$, then we stop there, so the trie has just one node. Otherwise, i.e., if $n \geq 2$, we pass all strings to new nodes; for each letter $\alpha \in \mathcal{A}$, we pass all strings beginning with $\alpha$, if any, to a new node labelled $\alpha$. We continue recursively, the next time partitioning the strings according to the second letter, and so on, always looking at the first letter not yet inspected; hence, the strings passed to a node $\boldsymbol{\alpha} \in \mathcal{A}^*$, if any, are the strings with prefix $\boldsymbol{\alpha}$, and if there are at least two such strings, then they are all passed further to children of $\boldsymbol{\alpha}$. At the end there is a tree with $n$ leaves, each containing one string.

Given a set of infinite strings, let $\nu_{\boldsymbol{\alpha}}$ be the number of these strings that have $\boldsymbol{\alpha}$ as a prefix, for $\boldsymbol{\alpha} \in \mathcal{A}^*$, and note that the trie $T$ just constructed can be defined as the subtree of $T_\infty$ consisting of all nodes $\boldsymbol{\alpha}$ such that one of the following holds:

- $\nu_{\boldsymbol{\alpha}} \geq 2$ (then $\boldsymbol{\alpha}$ is an internal node in $T$),
- $\nu_{\boldsymbol{\alpha}} = 1$ and either $\boldsymbol{\alpha} = \boldsymbol{\epsilon}$ or the parent of $\boldsymbol{\alpha}$ is an internal node (then $\boldsymbol{\alpha}$ is an external node in $T$).

We are mainly interested in random tries, see below, but we say also that a deterministic $|\mathcal{A}|$-ary tree is a trie if it can be generated in this way from some set of strings. (It is easily seen that a finite $|\mathcal{A}|$-ary tree is a trie if and only if there is no leaf with a parent that has only one child.) Denote the set of all tries by $\mathfrak{T} \subset \overline{\mathfrak{T}}$. Let $\mathfrak{T}_n := \overline{\mathfrak{T}}_n \cap \mathfrak{T}$, the set of tries with $n$ leaves, and $\mathfrak{T}_+ := \bigcup_1^\infty \mathfrak{T}_n = \mathfrak{T} \setminus \{\emptyset\}$.

Note that adding a new string to the ones generating a trie $T$ means either adding a new leaf to an internal node of $T$, or converting a leaf to a path of $k \geq 1$ additional internal nodes, and adding two new leaves to the last node in this path. We call this *adding a new string to $T$*.

A *functional* of tries is a function $\varphi : \mathfrak{T} \to \mathbb{R}$ such that (to avoid uninteresting complications) $\varphi(\emptyset) = 0$.

We say that a functional $\Phi$ of tries is *increasing* if $\Phi(T_1) \leq \Phi(T_2)$ whenever $T_1$ is a subtree of $T_2$. It is easily seen that it suffices to consider the case when $T_2$ is obtained from $T_1$ by adding a new string.

2.5. **Random tries.** Let $\mathcal{T}_n$ denote the random trie generated by the $n$ i.i.d. random infinite strings $\Xi^{(1)}, \Xi^{(2)}, \dots, \Xi^{(n)}$ (see Section 2.2). Note that $\mathcal{T}_n$ has $n$ leaves, so $\mathcal{T}_n \in \mathfrak{T}_n$.

In the trivial case $n = 1$, we see that $\mathcal{T}_1 = \bullet$ is non-random. (This is the only trie in $\mathfrak{T}_1$.)

We consider also the Poisson version. In general, for any random variable $N \in \mathbb{N}_0$, independent of the strings $\Xi^{(k)}$, $k \geq 1$, we may consider the tree $T_N$ constructed from the $N$ strings $\Xi^{(1)}, \dots, \Xi^{(N)}$. We will only consider the case $N = N_\lambda \sim \mathrm{Po}(\lambda)$ for some $\lambda > 0$, and we then use the notation

$$\widetilde{\mathcal{T}}_\lambda := T_{N_\lambda}. \tag{2.9}$$

In the Poisson case, we use the notation $N_{\lambda,\boldsymbol{\alpha}}$ for the (random) number $\nu_{\boldsymbol{\alpha}}$ of strings with prefix $\boldsymbol{\alpha}$, i.e.,

$$N_{\lambda,\boldsymbol{\alpha}} := \bigl|\{k \leq N_\lambda : \Xi^{(k)} \succeq \boldsymbol{\alpha}\}\bigr|. \tag{2.10}$$

By standard properties of the Poisson distribution, for any $\boldsymbol{\alpha} \in \mathcal{A}^*$,

$$N_{\lambda,\boldsymbol{\alpha}} \sim \mathrm{Po}\bigl(\lambda P(\boldsymbol{\alpha})\bigr). \tag{2.11}$$

Furthermore, for any finite strings $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_\ell$ such that none of them is a prefix of another, the random variables $N_{\lambda,\boldsymbol{\alpha}_1}, \dots, N_{\lambda,\boldsymbol{\alpha}_\ell}$ are independent.

2.6. **Bucket tries.** A *bucket trie* (or *b-trie* [31]) is a generalization of tries; it is constructed from a number of strings recursively in the same way as a trie, see Section 2.4, but stopping when the number of strings in a node is at most some given number $b$, known as the *bucket size*. Thus ordinary tries is the case $b = 1$. In general, a leaf (external node) will contain from 1 to $b$ strings. (The leafs are also called *buckets*.) In the notation above for random tries, the internal nodes are $\{\boldsymbol{\alpha} \in \mathcal{A}^* : \nu_{\boldsymbol{\alpha}} \geq b + 1\}$.

Note that, for any given bucket size $b \geq 2$, we can construct the trie $T$ based on a set of strings by first constructing the bucket trie $T'$ with bucket size $b$, and then letting a small trie grow from each bucket. Moreover, for i.i.d. random strings $\Xi^{(1)}, \dots, \Xi^{(n)}$ as above, conditioned on the bucket trie, these small tries are independent, and the small trie grown from a bucket that contains $k$ strings is a copy of $\mathcal{T}_k$.

We use bucket tries as a tool in some proofs.

2.7. **Fringe trees.** Given a rooted tree $T$ and a node $v$ in $T$, let $T^v$ be the subtree of $T$ consisting of $v$ and all its descendents (with $v$ as the root of $T^v$). Such subtrees are called *fringe subtrees*, or just *fringe trees*, of $T$. For convenience, we also define $T^v := \emptyset$, the empty tree, if $v \notin T$. We consider in the present paper only trees $T \in \overline{\mathfrak{T}}$, i.e., finite $|\mathcal{A}|$-ary trees; we then also regard the fringe trees $T^v$ as elements of $\overline{\mathfrak{T}}$ in the obvious way. (Recall that we have defined trees in $\overline{\mathfrak{T}}$ as subtrees of $T_\infty$ with root $\boldsymbol{\epsilon}$, the empty string.) Thus, formally,

$$T^{\boldsymbol{\alpha}} = \{\boldsymbol{\beta} \in \mathcal{A}^* : \boldsymbol{\alpha}\boldsymbol{\beta} \in T\} \tag{2.12}$$

Note that the fringe trees of a trie are tries. Furthermore, for a trie $T$ generated as in Section 2.4 from a set of strings, and any $\alpha \in T$,

$$|T^{\boldsymbol{\alpha}}|_{\mathsf{e}} = \nu_{\boldsymbol{\alpha}}, \tag{2.13}$$

the number of generating strings with prefix $\boldsymbol{\alpha}$.

The *random fringe subtree* $T^*$ is the random rooted tree obtained by taking the subtree $T^v$ at a uniformly random node $v$ in $T$; see [1]. (We assume $T \neq \emptyset$.) Let, for $|\mathcal{A}|$-ary trees $T, T' \in \overline{\overline{\mathfrak{T}}}$,

$$n_{T'}(T) := \bigl|\{v \in T' : T^v = T'\}\bigr|, \tag{2.14}$$

i.e., the number of subtrees of $T$ that are equal to $T'$. Then the distribution of $T^*$ is given by

$$\mathbb{P}(T^* = T') = n_{T'}(T)/|T|, \qquad T' \in \overline{\overline{\mathfrak{T}}}. \tag{2.15}$$

When $T$ is a random tree, as in [1] as well as in the present paper where we consider $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_\lambda$, $n_{T'}(T)$ is a random variable for each $T' \in \overline{\overline{\mathfrak{T}}}$, and (2.15) holds for the conditional probability $\mathbb{P}\bigl(T^* = T' \mid T\bigr)$.

2.8. **Additive functionals.** Let $\varphi$ be a functional of tries, and consider the functional $\Phi$ defined for a trie $T \in \mathfrak{T}$ by the sum

$$\Phi(T) = \Phi(T; \varphi) := \sum_{v \in T} \varphi(T^v). \tag{2.16}$$

(Thus, $\Phi(\emptyset) = 0$.) Recall that we assume $\varphi(\emptyset) = 0$. Hence, (2.16) can be written as the formally infinite sum

$$\Phi(T) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^*} \varphi(T^{\boldsymbol{\alpha}}). \tag{2.17}$$

Moreover, the definition (2.16) can also be written recusively as

$$\Phi(T) = \varphi(T) + \sum_{\alpha \in \mathcal{A}} \Phi(T^\alpha), \tag{2.18}$$

where $T^\alpha$, $\alpha \in \mathcal{A}$, are the principal branches of $T$, i.e., the fringe subtrees rooted at the children of the root.

A functional $\Phi$ that can be written as (2.16)–(2.18) is often called an *additive functional* with *toll function* $\varphi$. (Any functional can be written in this form for some $\varphi$, so the important part of this terminology is the relation between $\Phi$ and $\varphi$.)

**Example 2.1.** A simple example, which will be important in the sequel, is the toll function

$$\varphi_\bullet(T) := \mathbf{1}\{|T| = 1\} = \mathbf{1}\{T = \bullet\}; \tag{2.19}$$

then (2.16) shows that the corresponding additive functional $\Phi_\bullet$ counts the number of leaves in $T$. In particular, a random trie $\mathcal{T}_n$ has always $n$ leaves, and thus $\Phi_\bullet(\mathcal{T}_n) = n$ is non-random. Similarly, by (2.9),

$$\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = N_\lambda \sim \mathrm{Po}(\lambda). \tag{2.20}$$

□

**Example 2.2.** A more general example is to take $\varphi(T) = \mathbf{1}\{T = T'\}$, the indicator that $T$ equals some given tree $T' \in \overline{\mathfrak{T}}$; then $\Phi(T) = n_{T'}(T)$ defined in (2.14). Conversely, for any $\varphi$, (2.16) can be written

$$\Phi(T) = \sum_{T' \in \mathfrak{T}_+} \varphi(T') n_{T'}(T); \tag{2.21}$$

hence any additive functional can be written as a (potentially infinite) linear combination of the subtree counts $n_{T'}(T)$, where it suffices to consider (nonempty) tries $T'$. □

2.9. **Fringe trees of tries.** For the random trie $\widetilde{\mathcal{T}}_\lambda$ and any string $\boldsymbol{\alpha} \in \mathcal{A}^*$, we have by the recursive construction of tries that the fringe tree $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}$ is a trie constructed from $N_{\lambda,\boldsymbol{\alpha}}$ strings, except in the case $N_{\lambda,\boldsymbol{\alpha}} = 1$, when it is also possible that $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}} = \emptyset$ because $\boldsymbol{\alpha} \notin \widetilde{\mathcal{T}}_\lambda$ (when $\boldsymbol{\alpha}$ has a parent that is not an internal node). We therefore define

$$\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+} := \begin{cases} \widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}, & N_{\lambda,\boldsymbol{\alpha}} \neq 1, \\ \bullet, & N_{\lambda,\boldsymbol{\alpha}} = 1. \end{cases} \tag{2.22}$$

Then, $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ is always a trie constructed from $N_{\lambda,\boldsymbol{\alpha}}$ strings, and thus, by (2.11), for any (fixed) $\boldsymbol{\alpha} \in \mathcal{A}^*$,

$$\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+} \overset{\mathrm{d}}{=} \widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}. \tag{2.23}$$

Furthermore, $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}$ and $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ differ by (2.22) only in the case $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}} = \emptyset$ and $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+} = \bullet$; hence, for any functional $\varphi$ on $\overline{\mathfrak{T}}$,

$$\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}) + O(1). \tag{2.24}$$

Moreover, if $\varphi : \mathfrak{T} \to \mathbb{R}$ is a functional such that $\varphi(\bullet) = 0$, then

$$\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}) \overset{\mathrm{d}}{=} \varphi\big(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}\big). \tag{2.25}$$

For any finite strings $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_\ell$ such that none of them is a prefix of another, the random tries $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}_1+}, \dots, \widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}_\ell+}$ are independent, since this holds for $N_{\lambda,\boldsymbol{\alpha}_1}, \dots, N_{\lambda,\boldsymbol{\alpha}_\ell}$ as pointed out above. Note that this does not hold for the fringe tries $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}_1}, \dots, \widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}_\ell}$ in general, again because of the special case $N_{\lambda,\boldsymbol{\alpha}} = 1$.

For these reasons, we will often as a technical tool use $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ instead of $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}$.

**Remark 2.3.** For any additive functional $\Phi$ with toll function $\varphi$, $\Phi(\bullet) = \varphi(\bullet)$ by (2.16), and thus it follows from (2.22) that $\Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) - \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}) - \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+})$. Hence, for any $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_\ell$ such that none is a prefix of another, by the comments just made, the random variables $\Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) - \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})$ are independent. This could be used in the proofs below as an alternative to using the modified fringe tree $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$; it seems that the choice is mainly a matter of taste, but we invite the reader to explore this further. □

2.10. **Greatest common divisor.** Given a set $S$ of real numbers, we define $\gcd(S)$ to be the largest positive real number $d$ such that $S \subseteq d\mathbb{Z}$ (equivalently: $x/d \in \mathbb{Z}$ for every $x \in S$), provided that some such $d > 0$ exists; if no such $d$ exists, we define $\gcd(S) := 0$. (We assume that $S$ contains some non-zero element; otherwise this definition would give $\infty$.) We will only use this in the case $S := \{-\log p_\alpha : \alpha \in \mathcal{A}\}$, and we then use the special notation $d_{\mathbf{p}} := d(S)$ for this $S$. We say that $\mathbf{p}$ is *periodic* if $d_{\mathbf{p}} > 0$. (This is when periodic oscillations typically occur in the results below.)

In particular, if $x, y \neq 0$, then $\gcd(x, y) = 0 \iff x/y \notin \mathbb{Q}$. Hence, if $\mathcal{A} = \{0, 1\}$, then

$$d_{\mathbf{p}} = 0 \iff \frac{\log p_1}{\log p_0} \notin \mathbb{Q}. \tag{2.26}$$

2.11. **Mellin transform.** If $f$ is a (measurable) function on $(0, \infty)$, its *Mellin transform* is defined by

$$f^*(s) := \int_0^\infty f(s) x^{s-1} \, \mathrm{d}x, \tag{2.27}$$

for all complex $s$ such that the integral converges absolutely. (This domain is always a vertical strip in the complex plane, which may be infinite, finite, or empty. For simplicity we consider only absolute convergence which suffices for us; for other purposes one might also consider conditionally convergent integrals (2.27).) See further e.g. [9, Appendix B.7].

2.12. **Convergence and approximation in distribution.** As said above, we use $\overset{\mathrm{d}}{\longrightarrow}$ to denote convergence in distribution of random variables; these may take values in some metric space $\mathcal{S}$, see e.g. [2]. (We will only use $\mathcal{S} = \mathbb{R}^d$ for some $d$.) Recall that by definition [2], $X_n \overset{\mathrm{d}}{\longrightarrow} Y$ if and only if $\mathbb{E} f(X_n) \to \mathbb{E} f(Y)$ as $n \to \infty$ for every bounded continuous function $f : \mathcal{S} \to \mathbb{R}$. We extend this notion as follows.

Let $(X_n)_1^\infty$ and $(Y_n)_1^\infty$ be two sequences of random variables with values in a metric space $\mathcal{S}$. We write $X_n \overset{\mathrm{d}}{\approx} Y_n$ if, for every bounded continuous function $f : \mathcal{S} \to \mathbb{R}$,

$$\mathbb{E} f(X_n) = \mathbb{E} f(Y_n) + o(1) \qquad \text{as } n \to \infty. \tag{2.28}$$

If $\mathcal{S} = \mathbb{R}$, we say that $X_n \overset{\mathrm{d}}{\approx} Y_n$ *with moments of order* $s$ (where $s \in \mathbb{N}$) if (2.28) holds and also

$$\mathbb{E} X_n^s = \mathbb{E} Y_n^s + o(1) \tag{2.29}$$

with both sides finite. More generally, if $\mathcal{S} = \mathbb{R}^d$, we say that $X_n \overset{\mathrm{d}}{\approx} Y_n$ *with moments of order* $s$ if (2.28) holds and also, for every multi-index $\mathbf{m}$ with

$|\mathbf{m}| = s$,

$$\mathbb{E}\, X_n^{\mathbf{m}} = \mathbb{E}\, Y_n^{\mathbf{m}} + o(1) \tag{2.30}$$

with both sides finite. Similarly, still for $\mathcal{S} = \mathbb{R}^d$, we say that $X_n \overset{\mathrm{d}}{\approx} Y_n$ *with absolute moment of order* $s$ (where $s \in \mathbb{R}_+$) if (2.28) holds and also

$$\mathbb{E}\, |X_n|^s = \mathbb{E}\, |Y_n|^s + o(1) \tag{2.31}$$

with both sides finite.

For applications, ordinary moments are usually more interesting, but we use absolute moments in at least one proof; we therefore give statements including both. For brevity we will write "with [absolute] moments of order $s$", meaning with absolute moments of order $s$ and, provided $s$ is an integer, also with moments of order $s$. (For the relation between these, see Appendix B.)

We use the same notation for variables $X_\lambda$ and $Y_\lambda$ depending on a continuous parameter.

**Remark 2.4.** If $Y_n = Y$ for all $n$, then $X_n \overset{\mathrm{d}}{\approx} Y_n$ is equivalent to $X_n \overset{\mathrm{d}}{\longrightarrow} Y$, by the definitions above. More generally, the same holds if we assume $Y_n \overset{\mathrm{d}}{\longrightarrow} Y$. □

**Remark 2.5.** The standard *subsequence principle* says that a sequence in a metric space converges to a limit $x$ if and only if every subsequence has a subsubsequence that converges to $x$. It is well known that this holds also for convergence in distribution, in any metric space. (Cf. [11, Section 5.7]). It holds also for $\overset{\mathrm{d}}{\approx}$ (and any metric space $\mathcal{S}$): If every subsequence $(n_k)$ has a subsubsequence along which $X_n \overset{\mathrm{d}}{\approx} Y_n$, then $X_n \overset{\mathrm{d}}{\approx} Y_n$ along the full sequence. (This follows by fixing $f : \mathcal{S} \to \mathbb{R}$: each subsequence then has a subsubsequence such that (2.28) holds, and thus (2.28) holds for the full sequence.) The same holds with a continuous parameter. □

We use the subsequence principle several times in our proofs, often omitting some details. Here follows one example, extending to $\overset{\mathrm{d}}{\approx}$ the standard result that if $X_n \overset{\mathrm{d}}{\longrightarrow} Y$, then uniform integrability of $|X_n|^s$ implies convergence of [absolute] moments of order $s$, see e.g. [11, Theorem 5.5.9].

**Lemma 2.6.** *Let $(X_n)_1^\infty$ and $(Y_n)_1^\infty$ be random vectors in $\mathbb{R}^d$ such that $X_n \overset{\mathrm{d}}{\approx} Y_n$. Let further $s > 0$, and suppose that the sequence $(|X_n|^s)$ and $(|Y_n|^s)$ are uniformly integrable. Then, $X_n \overset{\mathrm{d}}{\approx} Y_n$ with [absolute] moments of order $s$.*

We give a detailed proof in Appendix B, together with a converse and some further comments.

## 3. A central limit theorem

We begin with our main results. Proofs are given in Sections 7 and 8.

The first theorem is a preliminary result giving asymptotics for mean and variance of additive functionals in the Poisson model under rather weak conditions (implying a linear growth), including the case of bounded toll functions; it also introduces some notation that will be used in the sequel. Corresponding results for the fixed $n$ model (under stronger conditions) are included in Theorem 3.9.

Recall the definition of the entropy $H$ in (2.2), the greatest common divisor $d_{\mathbf{p}} := \gcd\{-\log p_\alpha : \alpha \in \mathcal{A}\}$ in Section 2.10, and the Mellin transform $f^*$ in (2.27).

**Theorem 3.1.** *Let $\varphi$ be a toll function and let $\Phi$ be the corresponding additive functional given by* (2.16). *Suppose that, for some $\varepsilon > 0$, as $\lambda \to \infty$,*

$$\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) = O\big(\lambda^{1-\varepsilon}\big), \tag{3.1}$$

$$\operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) = O\big(\lambda^{1-\varepsilon}\big). \tag{3.2}$$

*Let*

$$\chi := \varphi(\bullet), \tag{3.3}$$

$$f_{\mathsf{E}}(\lambda) := \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) - \chi\lambda e^{-\lambda}, \tag{3.4}$$

$$\begin{aligned} f_{\mathsf{V}}(\lambda) := {}& 2\operatorname{Cov}\big(\varphi(\widetilde{\mathcal{T}}_\lambda), \Phi(\widetilde{\mathcal{T}}_\lambda)\big) - \operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) \\ & + 2\chi\lambda e^{-\lambda}\big(\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)\big) - \chi^2\lambda e^{-\lambda}\big(1 - \lambda e^{-\lambda}\big), \end{aligned} \tag{3.5}$$

$$f_{\mathsf{C}}(\lambda) := \operatorname{Cov}\big(\varphi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big) + \chi\lambda(\lambda-1)e^{-\lambda}. \tag{3.6}$$

*Then the following hold.*

(i) *If $d_{\mathbf{p}} = 0$, then, as $\lambda \to \infty$,*

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} \to \chi + \frac{1}{H} f_{\mathsf{E}}^*(-1) = \chi + \frac{1}{H}\int_0^\infty f_{\mathsf{E}}(x) x^{-2}\,\mathrm{d}x, \tag{3.7}$$

$$\frac{\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} \to \chi^2 + \frac{1}{H} f_{\mathsf{V}}^*(-1) = \chi^2 + \frac{1}{H}\int_0^\infty f_{\mathsf{V}}(x) x^{-2}\,\mathrm{d}x, \tag{3.8}$$

$$\frac{\operatorname{Cov}\big(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big)}{\lambda} \to \chi + \frac{1}{H} f_{\mathsf{C}}^*(-1) = \chi + \frac{1}{H}\int_0^\infty f_{\mathsf{C}}(x) x^{-2}\,\mathrm{d}x. \tag{3.9}$$

(ii) *More generally, for any $d_{\mathbf{p}}$, as $\lambda \to \infty$,*

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \chi + \frac{1}{H}\psi_{\mathsf{E}}(\log\lambda) + o(1), \tag{3.10}$$

$$\frac{\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \chi^2 + \frac{1}{H}\psi_{\mathsf{V}}(\log\lambda) + o(1), \tag{3.11}$$

$$\frac{\operatorname{Cov}\big(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big)}{\lambda} = \chi + \frac{1}{H}\psi_{\mathsf{C}}(\log\lambda) + o(1), \tag{3.12}$$

*where* $\psi_{\mathsf{X}}$, *for* $\mathsf{X} = \mathsf{E}, \mathsf{V}, \mathsf{C}$, *are bounded continuous functions defined as follows:*

(a) *If* $d_{\mathbf{p}} = 0$ *then* $\psi_{\mathsf{X}}$ *is constant: for all* $t$,

$$\psi_{\mathsf{X}}(t) := f^*_{\mathsf{X}}(-1). \tag{3.13}$$

(b) *If* $d = d_{\mathbf{p}} > 0$, *then* $\psi_{\mathsf{X}}$ *is a continuous* $d$*-periodic function having the Fourier series*

$$\psi_{\mathsf{X}}(t) \sim \sum_{m=-\infty}^{\infty} f^*_{\mathsf{X}}\Bigl(-1 - \frac{2\pi m}{d}\mathrm{i}\Bigr) e^{2\pi \mathrm{i} m t/d}. \tag{3.14}$$

*Furthermore,*

$$\psi_{\mathsf{X}}(t) = d \sum_{k=-\infty}^{\infty} e^{kd-t} f_{\mathsf{X}}\bigl(e^{t-kd}\bigr). \tag{3.15}$$

*Moreover, if* $\mathsf{X} = \mathsf{E}$, *or if* $f'_{\mathsf{X}}(\lambda) = O(\lambda^{-\varepsilon_1})$ *as* $\lambda \to \infty$ *for some* $\varepsilon_1 > 0$, *then the Fourier series* (3.14) *converges absolutely, and thus* $\sim$ *may be replaced by* $=$ *in* (3.14).

(iii) *If* $\varphi(T) \geq 0$ *for every trie* $T$ *and* $\varphi(T') > 0$ *for some trie* $T'$, *then* $\inf_t\bigl(H^{-1}\psi_{\mathsf{E}}(t) + \chi\bigr) > 0$, *and thus* $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \Theta(\lambda)$ *as* $\lambda \to \infty$.

**Remark 3.2.** When $d_{\mathbf{p}} > 0$, the constant term in (3.14) is $f^*_{\mathsf{X}}(-1)$. Thus we may regard the right-hand sides of (3.7)–(3.9) as "average asymptotic values" of the left-hand sides also when $d_{\mathbf{p}} > 0$, remembering that then the asymptotics really also include oscillations around these values. As is well known, the oscillation are numerically small in typical examples. □

**Remark 3.3.** It can be seen above, and in more detail later in the proof, that fringe subtrees • (leaves) play a special role; see also Section 2.9. The formulas in Theorem 3.1 simplify somewhat in the case $\varphi(\bullet) = 0$, where such fringe subtrees are ignored. (This case is very common in applications, see Section 4 for examples.) In particular, if $\varphi(\bullet) = 0$, then (3.4)–(3.6) simplify to

$$f_{\mathsf{E}}(\lambda) := \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda), \tag{3.16}$$

$$f_{\mathsf{V}}(\lambda) := 2\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) - \operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda), \tag{3.17}$$

$$f_{\mathsf{C}}(\lambda) := \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr). \tag{3.18}$$

□

**Remark 3.4.** It follows from the proof that $f_{\mathsf{E}}(\lambda), f_{\mathsf{V}}(\lambda), f_{\mathsf{C}}(\lambda)$ are finite for every $\lambda > 0$, and extend to entire functions, and that the Mellin transforms $f^*_{\mathsf{E}}(s), f^*_{\mathsf{V}}(s), f^*_{\mathsf{C}}(s)$ exist at least in the strip $-2 < \operatorname{Re} s < -1 + \varepsilon/2$, so the values in (3.7)–(3.9) and (3.14) are well defined. In fact, at least $f^*_{\mathsf{E}}(s)$ exists in the strip $-2 < \operatorname{Re} s < -1 + \varepsilon$, and (3.21) below shows that $f^*_{\mathsf{C}}$ extends analytically to the same strip, but we do not know whether (2.27) always converges absolutely there for $f_{\mathsf{C}}$. (The integral converges at least conditionally there by the proof of Lemma 3.6.) □

**Remark 3.5.** The results (3.8) and (3.11) in Theorem 3.1 extend immediately to the covariance $\operatorname{Cov}\bigl(\Phi_1(\widetilde{\mathcal{T}}_\lambda),\Phi_2(\widetilde{\mathcal{T}}_\lambda)\bigr)$ for two additive functionals with toll functions $\varphi_1,\varphi_2$ satisfying (3.1)–(3.2); the function $f_{\mathsf{V}}$ in (3.5) is replaced by, with $\chi_j:=\varphi_j(\bullet)$,

$$\begin{aligned}
f_{\mathsf{V},12}(\lambda):={}&\operatorname{Cov}\bigl(\varphi_1(\widetilde{\mathcal{T}}_\lambda),\Phi_2(\widetilde{\mathcal{T}}_\lambda)\bigr)+\operatorname{Cov}\bigl(\varphi_2(\widetilde{\mathcal{T}}_\lambda),\Phi_1(\widetilde{\mathcal{T}}_\lambda)\bigr)-\operatorname{Cov}\bigl(\varphi_1(\widetilde{\mathcal{T}}_\lambda),\varphi_2(\widetilde{\mathcal{T}}_\lambda)\bigr)\\
&+\chi_1\bigl(\mathbb{E}\,\Phi_2(\widetilde{\mathcal{T}}_\lambda)-\varphi_2(\widetilde{\mathcal{T}}_\lambda))\bigr)\lambda e^{-\lambda}+\chi_2\bigl(\mathbb{E}\,\Phi_1(\widetilde{\mathcal{T}}_\lambda)-\varphi_1(\widetilde{\mathcal{T}}_\lambda)\bigr)\lambda e^{-\lambda}\\
&-\chi_1\chi_2\bigl(1-\lambda e^{-\lambda}\bigr)\lambda e^{-\lambda}.
\end{aligned}\tag{3.19}$$

This follows by polarization, i.e., by considering $\varphi_1\pm\varphi_2$.

Note that taking $\varphi_2=\varphi_\bullet$ yields $\operatorname{Cov}\bigl(\Phi_1(\widetilde{\mathcal{T}}_\lambda),N_\lambda\bigr)$, so we can regard (3.9) and (3.12) as special cases of the bilinear versions of (3.8) and (3.11). Indeed, it is easily verified that if $\varphi_1=\varphi$ and $\varphi_2=\varphi_\bullet$, then (3.19) reduces to (3.6). □

We use in the sequel frequently the number $\chi$, the functions $f_{\mathsf{E}},f_{\mathsf{V}},f_{\mathsf{C}}$, their Mellin transforms $f^*_{\mathsf{E}},f^*_{\mathsf{V}},f^*_{\mathsf{C}}$, and the periodic functions $\psi_{\mathsf{E}},\psi_{\mathsf{V}},\psi_{\mathsf{C}}$ defined in Theorem 3.1; these have always the meanings above, for some given $\varphi$. (We say this explicitly sometimes, for emphasis, but not always.) We note a relation between $f_{\mathsf{E}}$ and $f_{\mathsf{C}}$.

**Lemma 3.6.** *Let $\varphi$ be as in Theorem 3.1. Then, for all $\lambda$ and $t$, and at least for* $\operatorname{Re}s\in(-2,-1+\varepsilon/2)$,

$$f_{\mathsf{C}}(\lambda)=\lambda f'_{\mathsf{E}}(\lambda),\tag{3.20}$$

$$f^*_{\mathsf{C}}(s)=-sf^*_{\mathsf{E}}(s),\tag{3.21}$$

$$\psi_{\mathsf{C}}(t)=\psi_{\mathsf{E}}(t)+\psi'_{\mathsf{E}}(t),\tag{3.22}$$

*In particular,*

$$f^*_{\mathsf{C}}(-1)=f^*_{\mathsf{E}}(-1).\tag{3.23}$$

**Remark 3.7.** The argument in the proof of (3.20) shows also that

$$\lambda\frac{\mathrm{d}}{\mathrm{d}\lambda}\Phi(\widetilde{\mathcal{T}}_\lambda)=\operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda,N_\lambda\bigr).\tag{3.24}$$

This derivative appears in the formula for the asymptotic variance of $\Phi(\mathcal{T}_n)$ already in [15]; we regard (3.24) as an explanation of this appearance.

Note also that (3.24) and (3.22) imply that (3.12) can be regarded as a formal derivative of (3.10). □

The next theorem might be regarded as our main result. It gives asymptotic normality of additive functionals of tries for both the Poisson and the fixed $n$ model. The theorem is easy to apply but still quite general; we will use it to show the results on fringe trees in Section 4. We have chosen to state this theorem here, because of its central role in the paper. However, as said above, we also later give some more general (and somewhat more technical) central limit theorems in Section 5; the proof of Theorem 3.9 combines some of these results from Sections 5 and Theorem 3.1. For simplicity, and convenience in many applications, we consider in the remainder of this section only toll function that are bounded.

**Remark 3.8.** The proof of Theorem 3.9 shows that the assumption on boundedness can be relaxed to the moment conditions (3.1), (3.2) and (5.5) (for any $r > 2$) for $\varphi$ and $\varphi_\pm$. The same applies to Theorem 3.12. □

**Theorem 3.9.** *Let $\varphi$ be a bounded toll function and let $\Phi$ be the corresponding additive functional given by* (2.16)*. Suppose further that $\varphi = \varphi_+ - \varphi_-$ for some bounded toll functions $\varphi_\pm$ such that the corresponding functionals $\Phi_\pm$ are increasing. Then, with notation as in Theorem 5.3,* (3.3)*–*(3.6) *and* (3.13)*–*(3.15)*:*

(i) *If $d_{\mathbf{p}} = 0$, then, as $\lambda \to \infty$ and $n \to \infty$,*

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}} \xrightarrow{\mathrm{d}} N\bigl(0, \sigma^2\bigr), \tag{3.25}$$

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{n}} \xrightarrow{\mathrm{d}} N\bigl(0, \widehat{\sigma}^2\bigr), \tag{3.26}$$

*with all [absolute] moments, where*

$$\sigma^2 = \chi^2 + H^{-1} f_{\mathsf{V}}^*(-1), \tag{3.27}$$

$$\widehat{\sigma}^2 = H^{-1} f_{\mathsf{V}}^*(-1) - H^{-2} f_{\mathsf{C}}^*(-1)^2 - 2\chi H^{-1} f_{\mathsf{C}}^*(-1). \tag{3.28}$$

(ii) *For any $d_{\mathbf{p}} \geq 0$, as $\lambda \to \infty$ and $n \to \infty$,*

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}} \overset{\mathrm{d}}{\approx} N\bigl(0, \sigma^2(\lambda)\bigr), \tag{3.29}$$

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{n}} \overset{\mathrm{d}}{\approx} N\bigl(0, \widehat{\sigma}^2(n)\bigr), \tag{3.30}$$

*with all [absolute] moments, where*

$$\sigma^2(\lambda) = \chi^2 + H^{-1} \psi_{\mathsf{V}}(\log \lambda), \tag{3.31}$$

$$\widehat{\sigma}^2(n) = H^{-1} \psi_{\mathsf{V}}(\log n) - H^{-2} \psi_{\mathsf{C}}(\log n)^2 - 2\chi H^{-1} \psi_{\mathsf{C}}(\log n), \tag{3.32}$$

*with continuous $d$-periodic functions $\psi_{\mathsf{V}}, \psi_{\mathsf{C}}$.*

(iii) *We have*

$$\mathbb{E}\,\Phi(\mathcal{T}_n) = \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_n) + o\bigl(\sqrt{n}\bigr) \tag{3.33}$$

*and may thus replace $\mathbb{E}\,\Phi(\mathcal{T}_n)$ by $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_n)$ in* (3.26) *and* (3.30)*.*

(iv) *If $\liminf_{n\to\infty} \operatorname{Var} \Phi(\mathcal{T}_n)/n > 0$, then, for any $d_{\mathbf{p}} \geq 0$,*

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\operatorname{Var} \Phi(\widetilde{\mathcal{T}}_\lambda)}} \xrightarrow{\mathrm{d}} N(0, 1), \tag{3.34}$$

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{\operatorname{Var} \Phi(\mathcal{T}_n)}} \xrightarrow{\mathrm{d}} N(0, 1), \tag{3.35}$$

*with convergence of all [absolute] moments.*

(v) *The means* $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$ *and* $\mathbb{E}\,\Phi(\mathcal{T}_n)$ *satisfy*

$$\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \lambda\bigl(\chi + H^{-1}\psi_{\mathsf{E}}(\log\lambda)\bigr) + o(\lambda), \tag{3.36}$$

$$\mathbb{E}\,\Phi(\mathcal{T}_n) = n\bigl(\chi + H^{-1}\psi_{\mathsf{E}}(\log n)\bigr) + o(n). \tag{3.37}$$

**Remark 3.10.** Theorem 3.9(i)–(iii) extend in an obvious way to multivariate limits for several functionals $\Phi_k$; this follows by the Cramér–Wold device (or by modifying the proof). □

**Remark 3.11.** We can in (3.25), (3.29) and (3.34) *not* replace $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$ by its asymptotic value $\lambda\bigl(\chi + H^{-1}\psi_{\mathsf{E}}(\log\lambda)\bigr)$ in (3.36). The reason is that that when $d_{\mathbf{p}} = 0$, the $o(1)$ error term in (3.10) typically is larger than $\lambda^{-1/2}$; in fact, this error term is in general not $O(\lambda^{-\varepsilon})$ for any $\varepsilon > 0$. When $d_{\mathbf{p}} > 0$, the error is $O(\lambda^{-\varepsilon})$ for some $\varepsilon > 0$ depending on the probabilities $\mathbf{p}$, but this $\varepsilon$ may be arbitrarily small; in particular, also in the case $d_{\mathbf{p}} > 0$, the error is in general not $o(\lambda^{-1/2})$. Thus the error term in (3.36) is in general not $o(\lambda^{1/2})$. These error estimates is implicit in Flajolet, Roux and Vallée [8]; see Appendix C for details.

The same holds for $\mathcal{T}_n$ and (3.26), (3.30), (3.35), as a consequence of these results for $\widetilde{\mathcal{T}}_\lambda$ and (3.33). □

As a corollary we obtain a weak law of large numbers. This is much weaker than the central limit theorem in Theorem 3.9, and presumably holds under weaker conditions (with a more direct proof), but we do not pursue this here.

**Theorem 3.12.** *Let* $\varphi$ *be a toll function satisfying the conditions of Theorem 3.9. Let* $f^*_{\mathsf{E}}(s)$, $\psi_{\mathsf{E}}(t)$ *and* $\chi$ *be as in Theorem 3.1.*

(i) *Then, as* $\lambda \to \infty$ *and* $n \to \infty$,

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} - H^{-1}\psi_{\mathsf{E}}(\log\lambda) - \chi \xrightarrow{\mathrm{p}} 0, \tag{3.38}$$

$$\frac{\Phi(\mathcal{T}_n)}{n} - H^{-1}\psi_{\mathsf{E}}(\log n) - \chi \xrightarrow{\mathrm{p}} 0. \tag{3.39}$$

*In particular, if* $d_{\mathbf{p}} = 0$, *then, as* $n \to \infty$,

$$\frac{\Phi(\mathcal{T}_n)}{n} \xrightarrow{\mathrm{p}} H^{-1}f^*_{\mathsf{E}}(-1) + \chi = H^{-1}\int_0^\infty \mathbb{E}[\varphi(\widetilde{\mathcal{T}}_\lambda)]\lambda^{-2}\,\mathrm{d}\lambda + \chi. \tag{3.40}$$

(ii) *If furthermore* $\varphi \geq 0$ *and* $\mathbb{P}(\varphi(\mathcal{T}_n) > 0) > 0$ *for some* $n \geq 1$, *then* $\inf_t\bigl(H^{-1}\psi_{\mathsf{E}}(t) + \chi\bigr) > 0$, *and thus, for some* $c > 0$, *as* $n \to \infty$,

$$\mathbb{P}\bigl(\Phi(\mathcal{T}_n) \geq cn\bigr) \to 1. \tag{3.41}$$

**Problem 3.13.** Do the limits (3.38)–(3.40) hold a.s.?

We give one case where the condition in Theorem 3.9(iv) holds; it holds in many other cases too, but see Example 3.17 for a counterexample.

**Lemma 3.14.** *Let $\Phi$ be an additive functional with bounded toll function $\varphi$ and suppose that there exists $n_0 \in \mathbb{N}$ such that $\varphi(\mathcal{T}_n) = a_n$ (a.s.) for $n \geq n_0$ and some constants $a_n$. Suppose also that* $\operatorname{Var}\Phi(\mathcal{T}_n) \neq 0$ *for some* $n \geq 1$. *Then* $\operatorname{Var}\Phi(\mathcal{T}_n) = \Omega(n)$ *as* $n \to \infty$.

Our formulas for variance asymptotics and asymptotic variances, (3.8), (3.11) and (3.31)–(3.32), use $f^*_{\mathsf{V}}$ and $\psi_{\mathsf{V}}$ which are defined using $f_{\mathsf{V}}(\lambda)$. The definition (3.5) of $f_{\mathsf{V}}(\lambda)$ is less useful for explicit calculations. We therefore give also an alternative formula, which will be used in the applications in Section 4. For simplicity, we consider only the case $\chi = 0$.

We use for convenience the special notation

$$\sum_{\boldsymbol{\alpha}}{}^{*} := \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|\geq 0} + \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|>0} \tag{3.42}$$

where thus every $\boldsymbol{\alpha} \in \mathcal{A}^*$ except $\boldsymbol{\alpha} = \boldsymbol{\epsilon}$ is counted twice.

**Lemma 3.15.** *Let $\varphi$ be a bounded toll function with $\varphi(\bullet) = 0$, and let $\Phi$ be the corresponding additive functional. Then, for $\lambda > 0$ and (at least)* $\operatorname{Re} s \in (-2, -\frac{1}{2})$,

$$f_{\mathsf{V}}(\lambda) = \sum_{\boldsymbol{\alpha}}{}^{*} \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), \varphi(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\bigr), \tag{3.43}$$

$$f^*_{\mathsf{V}}(s) = \sum_{\boldsymbol{\alpha}}{}^{*} \int_0^\infty \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), \varphi(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\bigr)\lambda^{s-1}\,\mathrm{d}\lambda, \tag{3.44}$$

*with sums and integrals absolutely convergent.*

We give also another useful formula for $f^*_{\mathsf{E}}$.

**Lemma 3.16.** *Let $\varphi$ be a bounded toll function. Then, at least for* $-2 < \operatorname{Re} s < 0$,

$$f^*_{\mathsf{E}}(s) = \sum_{n=2}^\infty \frac{\Gamma(n+s)}{n!}\,\mathbb{E}\,\varphi(\mathcal{T}_n). \tag{3.45}$$

*In particular,*

$$f^*_{\mathsf{E}}(-1) = f^*_{\mathsf{C}}(-1) = \sum_{n=2}^\infty \frac{\mathbb{E}\,\varphi(\mathcal{T}_n)}{(n-1)n}. \tag{3.46}$$

**Example 3.17.** The following example is in a sense negative, since it shows how trivial results can be derived by non-trivial calculations from the theorems above. However, the example serves both as an illustration of the formulas above, and as a counterexample and warning that there may be cancellations that are not obvious, leading to, for example, vanishing asymptotic variance or absence of expected oscillations.

Consider the toll function

$$\varphi(T) := \sum_{\alpha\in\mathcal{A}} \mathbf{1}\{|T^\alpha|_{\mathsf{e}} = 1\}. \tag{3.47}$$

Then, if $v$ is a leaf in $T$, then $\varphi(T^v) = 1$, while if $v$ is an internal node, then $\varphi(T^v)$ equals the number of children that are leaves. Since every leaf is a child of some internal node, except in the case $|T|_{\mathsf{e}} = 1$, it follows from (2.16) that if $|T|_{\mathsf{e}} > 1$, then $\Phi(T)$ is twice the number of leaves. In general, using the notation (2.19),

$$\Phi(T) = 2|T|_{\mathsf{e}} - \mathbf{1}\{|T|_{\mathsf{e}} = 1\} = 2|T|_{\mathsf{e}} - \varphi_{\bullet}(T). \tag{3.48}$$

In particular, $\Phi(\mathcal{T}_n) = 2n$ for $n \geq 2$, and $\operatorname{Var}\Phi(\mathcal{T}_n) = 0$. Also, $\Phi(\widetilde{\mathcal{T}}_\lambda) = 2N_\lambda + O(1)$ and $\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = 4\lambda + (4\lambda^2 - 3\lambda)e^{-\lambda} - \lambda^2 e^{-2\lambda} \sim 4\lambda$.

The additive functional $\Phi$ is increasing and the toll function $\varphi$ is bounded, so Theorem 3.9 applies.

We have $\chi = \varphi(\bullet) = 1$ and, by (3.4),

$$f_{\mathsf{E}}(\lambda) = \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) - \lambda e^{-\lambda} = \sum_{\alpha\in\mathcal{A}} p_\alpha \lambda e^{-p_\alpha\lambda} - \lambda e^{-\lambda}. \tag{3.49}$$

Thus, when $\operatorname{Re} s > -1$, using (2.27) and (2.4),

$$\begin{aligned} f^*_{\mathsf{E}}(s) &= \sum_{\alpha\in\mathcal{A}} \int_0^\infty p_\alpha \lambda^s e^{-p_\alpha\lambda}\,\mathrm{d}s - \int_0^\infty \lambda^s e^{-\lambda}\,\mathrm{d}\lambda \\ &= \sum_{\alpha\in\mathcal{A}} p_\alpha^{-s}\Gamma(s+1) - \Gamma(s+1) = \bigl(\rho(-s) - 1\bigr)\Gamma(s+1). \end{aligned} \tag{3.50}$$

By analytic continuation, (3.50) holds for $\operatorname{Re} s > -2$, with a removable singularity at $s = -1$. Letting $s \to -1$ yields, using (2.8),

$$f^*_{\mathsf{E}}(-1) = \lim_{s\to -1} \frac{\rho(-s) - 1}{s+1}\Gamma(s+2) = \frac{\mathrm{d}}{\mathrm{d}s}\rho(-s)\big|_{s=-1} = H. \tag{3.51}$$

Note that in the periodic case $d_{\mathbf{p}} > 0$, the sum (3.14) is over roots $\zeta_m := -1 - 2\pi m\mathrm{i}/d$ of $\rho(-s) = 1$, and (3.50) shows that $f^*_{\mathsf{E}}(\zeta_m) = 0$ for each such root $\zeta_m \neq -1$. Hence, $\psi_{\mathsf{E}}(t)$ is constant also in the periodic case, and for any $\mathbf{p}$,

$$\psi_{\mathsf{E}}(t) = f^*_{\mathsf{E}}(-1) = H. \tag{3.52}$$

In other words, the oscillations that usually occur vanish in this example. (This is not so obvious from (3.15).) Hence, for any $\mathbf{p}$, Theorem 3.9(v) gives

$$\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)/\lambda \to \chi + H^{-1} f^*_{\mathsf{E}}(-1) = 2 \qquad \text{as } \lambda \to \infty, \tag{3.53}$$

and similarly for $\mathbb{E}\,\Phi(\mathcal{T}_n)$. Of course, this is trivial from (3.48).

By (3.21) and (3.50), also $f^*_{\mathsf{C}}(\zeta_m) = 0$ for the roots $\zeta_m \neq -1$ of $\rho(-s) = 1$, and thus (3.14), (3.23) and (3.51) yield that, for any $\mathbf{p}$,

$$\psi_{\mathsf{C}}(t) = f^*_{\mathsf{C}}(-1) = H, \tag{3.54}$$

so this too is constant even in the periodic case.

Similarly, (3.5) yields, after some calculations,

$$f_{\mathsf{V}}(\lambda) = \sum_{\alpha\in\mathcal{A}} (3p_\alpha\lambda - 4p_\alpha^2\lambda^2)e^{-p_\alpha\lambda} + \sum_{\alpha\in\mathcal{A}} p_\alpha^2\lambda^2 e^{-2p_\alpha\lambda}$$

$$-(3\lambda-4\lambda^2)e^{-\lambda}-\lambda^2e^{-2\lambda}, \tag{3.55}$$

and thus

$$f^*_{\mathsf{V}}(\lambda)=\bigl(\rho(-s)-1\bigr)\bigl(3\Gamma(s+1)-4\Gamma(s+2)+2^{-s-2}\Gamma(s+2)\bigr). \tag{3.56}$$

Thus also $f^*_{\mathsf{V}}(\zeta_m)=0$ for the roots $\zeta_m\neq-1$, and (3.56) leads to, for any $\mathbf{p}$,

$$\psi_{\mathsf{V}}(t)=f^*_{\mathsf{V}}(-1)=3H. \tag{3.57}$$

Note that (3.31) and (3.32) yield, using (3.54) and (3.57), $\sigma^2(\lambda)=4$ and $\widehat{\sigma}^2(n)=0$. Of course, (3.29) and (3.30) with these variances are trivial from (3.48).

This example has $\chi=1$, and we see how $\chi$ and the functions $\psi_{\mathsf{X}}$ interact in (3.31)–(3.32) and (3.36)–(3.37). Consider now the modification

$$\varphi_*(T):=\varphi(T)-\varphi_\bullet(T). \tag{3.58}$$

This equals the number of children of the root that are external nodes. By (3.48),

$$\Phi_*(T)=|T|_{\mathsf{e}}-\mathbf{1}\{|T|_{\mathsf{e}}=1\}=|T|_{\mathsf{e}}-\varphi_\bullet(T). \tag{3.59}$$

In particular, again $\Phi_*(\mathcal{T}_n)$ is deterministic. Similar calculations, or simpler the general (7.51) and (7.50) in the proof of Theorem 5.3, yield $f_{\mathsf{E},*}(\lambda)=f_{\mathsf{E}}(\lambda)$, $f_{\mathsf{C},*}(\lambda)=f_{\mathsf{C}}(\lambda)$ given by (3.20) and (3.49), and

$$f_{\mathsf{V},*}(\lambda)=f_{\mathsf{V}}(\lambda)-2f_{\mathsf{C}}(\lambda)=\sum_{\alpha\in\mathcal{A}}(p_\alpha\lambda-2p_\alpha^2\lambda^2)e^{-p_\alpha\lambda}+\sum_{\alpha\in\mathcal{A}}p_\alpha^2\lambda^2e^{-2p_\alpha\lambda}$$
$$-(\lambda-2\lambda^2)e^{-\lambda}-\lambda^2e^{-2\lambda}. \tag{3.60}$$

Hence,

$$\psi_{\mathsf{E},*}(t)=\psi_{\mathsf{E}}(t)=\psi_{\mathsf{C},*}(t)=\psi_{\mathsf{C}}(t)=H, \tag{3.61}$$
$$\psi_{\mathsf{V},*}(t)=\psi_{\mathsf{V}}(t)-2\psi_{\mathsf{C}}(t)=H, \tag{3.62}$$

and (3.31)–(3.32) yield $\sigma^2_*(\lambda)=1$ and $\sigma^2_*(n)=0$. Again, Theorem 3.9(ii) and (v) hold trivially.

Finally, consider the modification

$$\varphi_{**}(T):=\varphi(T)-2\varphi_\bullet(T). \tag{3.63}$$

By (3.48), this toll function yields the additive functional

$$\Phi_{**}(T)=-\mathbf{1}\{|T|_{\mathsf{e}}=1\}=-\varphi_\bullet(T). \tag{3.64}$$

Hence $\Phi_{**}(\mathcal{T}_n)=0$ for $n\geq2$, and $\Phi_{**}(\widetilde{\mathcal{T}}_\lambda)=-\mathbf{1}\{N_\lambda=1\}$ converges rapidly to 0. This additive functional is thus essentially 0, although the toll function in (3.63) looks non-trivial. Both (3.34) and (3.35) obviously fail. The other parts of Theorem 3.9 apply also to this degenerate case. We have $\chi_{**}=-1$ and, for example using (7.51) and (7.50) again, $f_{\mathsf{E},**}(\lambda)=f_{\mathsf{E}}(\lambda)$, $f_{\mathsf{C},**}(\lambda)=f_{\mathsf{C}}(\lambda)$, and

$$f_{\mathsf{V},**}(\lambda)=f_{\mathsf{V},*}(\lambda)-2f_{\mathsf{C}}(\lambda)=-\sum_{\alpha\in\mathcal{A}}p_\alpha\lambda e^{-p_\alpha\lambda}+\sum_{\alpha\in\mathcal{A}}p_\alpha^2\lambda^2e^{-2p_\alpha\lambda}$$

$$+ \lambda e^{-\lambda} - \lambda^2 e^{-2\lambda}, \tag{3.65}$$

and thus

$$\psi_{\mathsf{E},**}(t) = \psi_{\mathsf{E}}(t) = \psi_{\mathsf{C},**}(t) = \psi_{\mathsf{C}}(t) = H, \tag{3.66}$$

$$\psi_{\mathsf{V},**}(t) = \psi_{\mathsf{V},*}(t) - 2\psi_{\mathsf{C}}(t) = -H. \tag{3.67}$$

Thus (3.31)–(3.32) yield $\sigma^2_{**}(\lambda) = 0$ and $\widehat{\sigma}^2_{**}(n) = 0$. Again Theorem 3.9(ii) and (v) hold trivially.

□

## 4. Central limit theorems for fringe tries

We give some applications of the general results above, including applications to the distribution of random fringe trees.

We often state results only for the fixed $n$ model $\mathcal{T}_n$; similar results for the Poisson model $\widetilde{\mathcal{T}}_\lambda$ follow similarly, but are only sometimes stated explicitly.

We use the notation of Section 3, in particular $\chi, f_{\mathsf{X}}, f^*_{\mathsf{X}}, \psi_{\mathsf{X}}$ defined in Theorem 3.1; recall also $\sum^*_{\boldsymbol{\alpha}}$ defined in (3.42). We will distinguish different additive functionals by subscripts, and we sometimes use these subscripts in an obvious way also for $\chi$, $f_{\mathsf{X}}$ and so on, but we often omit subscripts when there is no risk of confusion.

In all examples below, asymptotics for means and variances are given by (3.36)–(3.37) and (3.31)–(3.32), using $\psi_{\mathsf{E}}, \psi_{\mathsf{V}}, \psi_{\mathsf{C}}$ that are given by the Mellin transforms $f^*_{\mathsf{E}}, f^*_{\mathsf{V}}, f^*_{\mathsf{C}}$ and (3.14) (absolutely convergent in all cases). We calculate these Mellin transforms in several cases, but usually omit stating explicitly the formulas for asymptotic means and variances that they lead to.

### 4.1. **The size.**

As a warm-up, we consider first the size of the trie, measured as $\Phi_{\mathsf{i}}(T) := |T|_{\mathsf{i}}$, the number of internal nodes. This example has been studied by many authors. In particular, asymptotic normality was shown already by Jacquet and Régnier [15]; see also [23, Section 5.4]. Variance asymptotics is also studied there and in several other papers, see the detailed analysis by Fuchs, Hwang and Zacharovas [10] and the many references given there. We show here how these results follows by our methods.

The functional $\Phi_{\mathsf{i}}(T)$ is an additive functional with toll function

$$\varphi_{\mathsf{i}}(T) = \mathbf{1}\{\text{the root is an internal node}\}, \tag{4.1}$$

and thus

$$\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda) = \mathbf{1}\{N_\lambda \geq 2\}. \tag{4.2}$$

In this case, $\Phi_{\mathsf{i}}$ is an increasing functional, so Theorems 3.9 and 3.12 apply with $\varphi_+ = \varphi_{\mathsf{i}}$ and $\varphi_- = 0$.

Lemma 3.14 shows that $\operatorname{Var} \Phi_{\mathsf{i}}(\mathcal{T}_n) = \Theta(n)$, and thus Theorem 3.9(iv) applies; consequently, Theorem 3.9 shows immediately that both $\Phi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)$ and $\Phi_{\mathsf{i}}(\mathcal{T}_n)$ are asymptotically normal; more precisely, the following holds. (For the means, recall also Remark 3.11.)

**Theorem 4.1.** *[Jacquet and Régnier [15]] Consider the size* $\Phi_{\mathsf{i}}(T) = |T|_{\mathsf{i}}$*. Then, the central limit theorems* (3.29)–(3.30) *and* (3.34)–(3.35) *hold, with all [absolute] moments, and with asymptotic variances given by* (3.31)–(3.32) *(and thus by* (3.27)–(3.28) *when* $d_{\mathbf{p}} = 0$*). Furthermore, the means satisfy* (3.36) *and* (3.37)*, and the laws of large numbers* (3.38)–(3.39) *hold.* □

We have $\chi = 0$ (so the formulas simplify a little) and, by (4.2),

$$f_{\mathsf{E}}(\lambda) = \mathbb{E}\,\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda) = \mathbb{P}\bigl(N_\lambda \ge 2\bigr) = 1 - (1+\lambda)e^{-\lambda} \tag{4.3}$$

and thus

$$f^*_{\mathsf{E}}(s) = \int_0^\infty \bigl(1-(1+\lambda)e^{-\lambda}\bigr)\lambda^{s-1}\,\mathrm{d}\lambda = -\frac{\Gamma(s+2)}{s}, \qquad -2 < \operatorname{Re} s < 0; \tag{4.4}$$

where the integral can be evaluated e.g. using integration by parts, cf. [18, Proof of Theorem 5.3]. In particular, or by (3.46),

$$f^*_{\mathsf{E}}(-1) = 1, \tag{4.5}$$

so if $d_{\mathbf{p}} = 0$, then $\Phi_{\mathsf{i}}(\mathcal{T}_n)/n \overset{\mathrm{p}}{\longrightarrow} 1/H$ by (3.40).

By (3.21) and (4.4),

$$f^*_{\mathsf{C}}(s) = \Gamma(s+2), \qquad \operatorname{Re} s > -2. \tag{4.6}$$

For any trie $T$ and any $\boldsymbol{\alpha} \in \mathcal{A}^*$, if $\varphi_{\mathsf{i}}(T) = 0$ then $\varphi_{\mathsf{i}}(T^{\boldsymbol{\alpha}}) = 0$. Hence, $\varphi_{\mathsf{i}}(T)\varphi_{\mathsf{i}}(T^{\boldsymbol{\alpha}}) = \varphi_{\mathsf{i}}(T^{\boldsymbol{\alpha}})$, and thus, using (4.3),

$$\begin{aligned}\operatorname{Cov}\bigl(\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda), \varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\bigr) &= \bigl(1 - \mathbb{E}\,\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)\bigr)\,\mathbb{E}\,\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\\ &= (1+\lambda)e^{-\lambda}\bigl(1-(1+P(\boldsymbol{\alpha})\lambda)e^{-P(\boldsymbol{\alpha})\lambda}\bigr).\end{aligned} \tag{4.7}$$

Consequently, (3.44) yields

$$f^*_{\mathsf{V}}(s) = \sum_{\boldsymbol{\alpha}}{}^{*} \int_0^\infty (1+\lambda)e^{-\lambda}\bigl(1-(1+P(\boldsymbol{\alpha})\lambda)e^{-P(\boldsymbol{\alpha})\lambda}\bigr)\lambda^{s-1}\,\mathrm{d}\lambda. \tag{4.8}$$

For $s > 0$, the right-hand side is, by standard Gamma integrals, evaluated as

$$\begin{aligned}&\sum_{\boldsymbol{\alpha}}{}^{*}\Bigl(\Gamma(s)+\Gamma(s+1)-(1+P(\boldsymbol{\alpha}))^{-s}\Gamma(s)-(1+P(\boldsymbol{\alpha}))^{1-s-1}\Gamma(s+1)\\ &\qquad\qquad - P(\boldsymbol{\alpha})(1+P(\boldsymbol{\alpha}))^{-s-2}\Gamma(s+2)\Bigr)\\ &= \sum_{\boldsymbol{\alpha}}{}^{*}\frac{(1+s)\Gamma(s)}{(1+P(\boldsymbol{\alpha}))^{s+2}}\Bigl((1+P(\boldsymbol{\alpha}))^{s+2}-(1+P(\boldsymbol{\alpha}))^2 - sP(\boldsymbol{\alpha})\Bigr).\end{aligned} \tag{4.9}$$

The terms in the final sum are, by Taylor expansions, $O(P(\boldsymbol{\alpha})^2)$ for fixed $s$, and thus the sum converges for every $s > 0$ by (2.7); hence the Mellin transform $f^*_{\mathsf{V}}(s)$ is finite for $s > 0$ and equals (4.9). (Note that the expression in (4.7) is positive; hence we may interchange the order of summation and integration in (3.44) for real $s$.) Since the domain of existence of the Mellin transform always is a vertical strip, this shows that $f^*_{\mathsf{V}}(s)$ exists in the half-plane $\operatorname{Re} s > -2$, and analytic continuation yields that it equals (4.9); hence, for all such $s \neq 0$, rewriting $(s+1)\Gamma(s) = \Gamma(s+2)/s$,

$$f^*_{\mathsf{V}}(s) = \sum_{\boldsymbol{\alpha}}^{*} \frac{\Gamma(s+2)}{s(1+P(\boldsymbol{\alpha}))^{s+2}} \Big( (1+P(\boldsymbol{\alpha}))^{s+2} - (1+P(\boldsymbol{\alpha}))^2 - sP(\boldsymbol{\alpha}) \Big). \tag{4.10}$$

In particular,

$$f^*_{\mathsf{V}}(-1) = \sum_{\boldsymbol{\alpha}}^{*} \frac{P(\boldsymbol{\alpha})^2}{1+P(\boldsymbol{\alpha})}. \tag{4.11}$$

Using (4.6) and (4.11), we obtain expressions for $\psi_{\mathsf{C}}$ and $\psi_{\mathsf{V}}$ from (3.14), leading to (somewhat complicated) formulas for $\sigma^2(\lambda)$ and $\widehat{\sigma}^2(n)$ by (3.27)–(3.28) and (3.31) and (3.32). This yields the results found by Jacquet and Régnier [15; 30], Fuchs, Hwang and Zacharovas [10] and others by somewhat different methods.

4.2. **Size of fringe tries.** We turn to the fringe (sub)trees of a random trie. We first consider their sizes, in this section measured as their number of external nodes (leaves). (Note the difference from Section 4.1.)

Let $k \geq 1$ and let

$$\varphi_k(T) := \mathbf{1}\{|T|_{\mathsf{e}} = k\}. \tag{4.12}$$

Then, the corresponding additive functional $\Phi_k$ counts the number of fringe trees with exactly $k$ leaves. Note that $\varphi_1 = \varphi_\bullet$ in Example 2.1, and thus $\Phi_1(\mathcal{T}_n) = n$. In the sequel we mainly consider $k \geq 2$.

The functional $\Phi_k$ is not increasing, but the functional $\Phi_{\geq k} := \sum_{j \geq k} \Phi_j$ is, and $\Phi_k = \Phi_{\geq k} - \Phi_{\geq k+1}$; furthermore, $\Phi_{\geq k}$ has a bounded toll function $\varphi_{\geq k} := \sum_{j \geq k} \varphi_j$. Hence Theorems 3.9 and 3.12 apply (with $\varphi_+ = \varphi_{\geq k}$ and $\varphi_- = \varphi_{\geq k+1}$) and yield, using also Remark 3.10 and Lemma 3.14, the following.

**Theorem 4.2.** *Let $k \geq 2$ and consider $\Phi_k$, the number of fringe trees with $k$ leaves. Then, the central limit theorems* (3.29)–(3.30) *and* (3.34)–(3.35) *hold, with all [absolute] moments, and with asymptotic variances given by* (3.31)–(3.32) *(and thus by* (3.27)–(3.28) *when $d_{\mathbf{p}} = 0$); this extends to joint convergence for several $k$. Furthermore, the means satisfy* (3.36) *and* (3.37)*, and the laws of large numbers* (3.38)–(3.39) *hold.* □

Suppose that $k \geq 2$. We then have

$$f_{\mathsf{E},k}(\lambda) := \mathbb{E}\,\varphi_k(\widetilde{\mathcal{T}}_\lambda) = \mathbb{P}(N_\lambda = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{4.13}$$

Hence, or by Lemma 3.16, the Mellin transform $f^*_{\mathsf{E},k}(s)$ exists for $\operatorname{Re} s > -k$, and

$$f^*_{\mathsf{E},k}(s) = \frac{\Gamma(k+s)}{k!}. \tag{4.14}$$

In particular,

$$f^*_{\mathsf{E},k}(-1) = \frac{1}{k(k-1)}. \tag{4.15}$$

If $\varphi_k(\widetilde{\mathcal{T}}_\lambda) = 1$ and $\boldsymbol{\alpha} \in \mathcal{A}^*$, then $\varphi_k(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda) = 1$ only if all $k$ strings are passed to $\boldsymbol{\alpha}$, which has (conditional) probability $P(\boldsymbol{\alpha})^k$. Hence, recalling (4.13),

$$\operatorname{Cov}\bigl(\varphi_k(\widetilde{\mathcal{T}}_\lambda), \varphi_k(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\bigr) = \frac{\lambda^k}{k!}e^{-\lambda}\Bigl(P(\boldsymbol{\alpha})^k - \frac{(P(\boldsymbol{\alpha})\lambda)^k}{k!}e^{-P(\boldsymbol{\alpha})\lambda}\Bigr). \tag{4.16}$$

Consequently, by (3.44), for $\operatorname{Re} s > -k$,

$$\begin{aligned} f^*_{\mathsf{V},k}(s) &= \sum_{\boldsymbol{\alpha}}^* \int_0^\infty \frac{\lambda^k}{k!}e^{-\lambda}\Bigl(P(\boldsymbol{\alpha})^k - \frac{(P(\boldsymbol{\alpha})\lambda)^k}{k!}e^{-P(\boldsymbol{\alpha})\lambda}\Bigr)\lambda^{s-1}\,\mathrm{d}\lambda \\ &= \frac{\Gamma(s+k)}{k!}\sum_{\boldsymbol{\alpha}}^* P(\boldsymbol{\alpha})^k - \frac{\Gamma(s+2k)}{k!^2}\sum_{\boldsymbol{\alpha}}^* \frac{P(\boldsymbol{\alpha})^k}{(1+P(\boldsymbol{\alpha}))^{s+2k}}, \end{aligned} \tag{4.17}$$

where the sums converge since $k \geq 2$, see (2.7). In particular, this easily yields, using $\rho(k) := \sum_{\alpha\in\mathcal{A}} p_\alpha^k < 1$ as in (2.4)–(2.6),

$$f^*_{\mathsf{V},k}(-1) = \frac{1}{k(k-1)}\frac{1+\rho(k)}{1-\rho(k)} - \frac{(2k-2)!}{k!^2}\sum_{\boldsymbol{\alpha}}^* \frac{P(\boldsymbol{\alpha})^k}{(1+P(\boldsymbol{\alpha}))^{2k-1}}. \tag{4.18}$$

The asymptotic normality in Theorem 4.2 holds, as stated there, jointly for different $k$. Furthermore, still by Remark 3.10, it holds jointly with the asymptotic normality of $\Phi_{\mathsf{i}}(\mathcal{T}_n) = |\mathcal{T}_n|_{\mathsf{i}}$ in Theorem 4.1. Asymptotic covariances can be calculated by similar arguments as above. We illustrate this for the asymptotic covariance between $\Phi_k(\mathcal{T}_n)$ and $|\mathcal{T}_n|_{\mathsf{i}}$ for a given $k \geq 2$. (Calculations for other covariances are slightly more complicated, but the principle is the same.)

The bivariate version of (3.30) and (3.32) (cf. Remark 3.10) yields

$$n^{-1}\operatorname{Cov}\bigl(\Phi_k(\mathcal{T}_n), \Phi_{\mathsf{i}}(\mathcal{T}_n)\bigr) = \widehat{\sigma}_{k\mathsf{i}}(n) + o(1), \tag{4.19}$$

where

$$\widehat{\sigma}_{k\mathsf{i}}(n) = H^{-1}\psi_{\mathsf{V},k\mathsf{i}}(\log n) - H^{-2}\psi_{\mathsf{C},k}(\log n)\psi_{\mathsf{C},\mathsf{i}}(\log n) \tag{4.20}$$

where $\psi_{\mathsf{V},k\mathsf{i}}$ is given by (3.14) with $f^*_{\mathsf{X}} = f^*_{\mathsf{V},k\mathsf{i}}$, the Mellin transform of $f_{\mathsf{V},k\mathsf{i}}$ which by (3.19) is given by (noting that $\chi_k = \chi_{\mathsf{i}} = 0$)

$$f_{\mathsf{V},k\mathsf{i}}(\lambda) = \operatorname{Cov}\bigl(\varphi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)\bigr) + \operatorname{Cov}\bigl(\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda), \Phi_k(\widetilde{\mathcal{T}}_\lambda) - \varphi_k(\widetilde{\mathcal{T}}_\lambda)\bigr). \tag{4.21}$$

We note that

$$\begin{aligned} \mathbb{E}\bigl[\varphi_k(\widetilde{\mathcal{T}}_\lambda)\Phi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)\bigr] &= \mathbb{P}\bigl(|\widetilde{\mathcal{T}}_\lambda|_{\mathsf{e}} = k\bigr)\,\mathbb{E}\bigl[\Phi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda) \bigm| |\widetilde{\mathcal{T}}_\lambda|_{\mathsf{e}} = k\bigr] \\ &= f_{\mathsf{E},k}(\lambda)\,\mathbb{E}\bigl[\Phi_{\mathsf{i}}(\mathcal{T}_k)\bigr]. \end{aligned} \tag{4.22}$$

Furthermore, if $\varphi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)=0$, then $\varphi_k(\widetilde{\mathcal{T}}_\lambda)=\Phi_k(\widetilde{\mathcal{T}}_\lambda)=0$. Hence, using also (4.3) and (4.13), (4.21) yields, with $E_k:=\mathbb{E}\,\Phi_{\mathsf{i}}(\mathcal{T}_k)=\mathbb{E}\,|\mathcal{T}_k|_{\mathsf{i}}$,

$$\begin{aligned} f_{\mathsf{V},k\mathsf{i}}(\lambda)&=f_{\mathsf{E},k}(\lambda)\Bigl(E_k-\mathbb{E}\bigl[\Phi_{\mathsf{i}}(\widetilde{\mathcal{T}}_\lambda)\bigr]\Bigr)+\bigl(1-f_{\mathsf{E},\mathsf{i}}(\lambda)\bigr)\Bigl(\mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_\lambda)-\mathbb{E}\,\varphi_k(\widetilde{\mathcal{T}}_\lambda)\Bigr)\\ &=\frac{\lambda^k}{k!}e^{-\lambda}\Bigl(E_k-\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}f_{\mathsf{E},\mathsf{i}}(P(\boldsymbol{\alpha})\lambda)\Bigr)+(1+\lambda)e^{-\lambda}\sum_{|\boldsymbol{\alpha}|\ge1}f_{\mathsf{E},k}(P(\boldsymbol{\alpha})\lambda). \end{aligned}\tag{4.23}$$

This yields after simple calculations, partly arguing as for (4.8)–(4.9),

$$\begin{aligned} f^*_{\mathsf{V},k\mathsf{i}}(s)=\frac{\Gamma(k+s)}{k!}\Biggl(E_k&-\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}\frac{(1+P(\boldsymbol{\alpha}))^{k+s+1}-1-(k+s+1)P(\boldsymbol{\alpha})}{(1+P(\boldsymbol{\alpha}))^{k+s+1}}\\ &+\sum_{|\boldsymbol{\alpha}|\ge1}\frac{k+s+1+P(\boldsymbol{\alpha})}{(1+P(\boldsymbol{\alpha}))^{k+s+1}}P(\boldsymbol{\alpha})^k\Biggr). \end{aligned}\tag{4.24}$$

Furthermore, $f^*_{\mathsf{C},\mathsf{i}}=\Gamma(s+2)$ by (4.6) and

$$f^*_{\mathsf{C},k}=-s\Gamma(k+s)/k!\tag{4.25}$$

by Lemma 3.6 and (4.14). Finally, as said above, $\psi_{\mathsf{V},k\mathsf{i}},\psi_{\mathsf{C},\mathsf{i}},\psi_{\mathsf{C},k}$ are given by (3.14), and (4.20) yields $\widehat{\sigma}_{k\mathsf{i}}(n)$. In the aperiodic case, $\widehat{\sigma}_{k\mathsf{i}}$ is constant and the formulas simplify:

$$\begin{aligned} \widehat{\sigma}_{k\mathsf{i}}&=H^{-1}f^*_{\mathsf{V},k\mathsf{i}}(-1)-H^{-2}f^*_{\mathsf{C},k}(-1)f^*_{\mathsf{C},\mathsf{i}}(-1)\\ &=H^{-1}f^*_{\mathsf{V},k\mathsf{i}}(-1)-H^{-2}f^*_{\mathsf{E},k}(-1)f^*_{\mathsf{E},\mathsf{i}}(-1)\\ &=\frac{H^{-1}}{k(k-1)}\Biggl(E_k-\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}\frac{(1+P(\boldsymbol{\alpha}))^k-1-kP(\boldsymbol{\alpha})}{(1+P(\boldsymbol{\alpha}))^k}\\ &\qquad\qquad+\sum_{|\boldsymbol{\alpha}|\ge1}\frac{k+P(\boldsymbol{\alpha})}{(1+P(\boldsymbol{\alpha}))^k}P(\boldsymbol{\alpha})^k-H^{-1}\Biggr). \end{aligned}\tag{4.26}$$

4.2.1. *Asymptotic distributions.* We use these results to study the distribution of the size of a (uniformly) random fringe subtree $\mathcal{T}_n^*$ of $\mathcal{T}_n$, defined as in Section 2.7 as $\mathcal{T}_n^v$ for a uniformly random node $v$ in $\mathcal{T}_n$. Note that we allow both internal and external nodes $v$.

**Remark 4.3.** Alternatively, one might consider a random internal fringe tree by taking only internal nodes $v$. This is equivalent to conditioning the fringe tree $\mathcal{T}_n^v$ on $v$ being an internal node. Since $v$ is external if and only if $\mathcal{T}_n^v=\bullet$, this random internal fringe tree equals the random fringe tree $\mathcal{T}_n^*$ (defined as above) conditioned on $\mathcal{T}_n^*\neq\bullet$. The results below are easily transferred to this version. □

The total number of nodes in $\mathcal{T}_n$ is

$$|\mathcal{T}_n|=|\mathcal{T}_n|_{\mathsf{i}}+|\mathcal{T}_n|_{\mathsf{e}}=\Phi_{\mathsf{i}}(\mathcal{T}_n)+n.\tag{4.27}$$

Hence, by Theorem 4.1 and (3.39),

$$|\mathcal{T}_n|/n=H^{-1}\psi_{\mathsf{E},\mathsf{i}}(\log n)+1+o_{\mathrm{p}}(1).\tag{4.28}$$

Similarly, by Theorem 4.2, for $k \geq 2$,

$$\Phi_k(\mathcal{T}_n)/n = H^{-1}\psi_{\mathsf{E},k}(\log n) + o_{\mathrm{p}}(1). \tag{4.29}$$

This implies the following result, using also (4.5) and (4.13).

**Theorem 4.4.** *The fringe tree size distribution of $\mathcal{T}_n$ satisfies*

$$\mathbb{P}\bigl(|\mathcal{T}_n^*|_{\mathsf{e}} = k \mid \mathcal{T}_n\bigr) = \frac{\Phi_k(\mathcal{T}_n)}{|\mathcal{T}_n|} = \begin{cases} \frac{\psi_{\mathsf{E},k}(\log n)}{\psi_{\mathsf{E,i}}(\log n)+H} + o_{\mathrm{p}}(1), & k \geq 2, \\ \frac{H}{\psi_{\mathsf{E,i}}(\log n)+H} + o_{\mathrm{p}}(1), & k = 1. \end{cases} \tag{4.30}$$

*In particular, if $d_{\mathbf{p}} = 0$, the distribution converges in probability:*

$$\mathbb{P}\bigl(|\mathcal{T}_n^*|_{\mathsf{e}} = k \mid \mathcal{T}_n\bigr) \overset{\mathrm{p}}{\longrightarrow} \begin{cases} \frac{1}{(1+H)k(k-1)}, & k \geq 2, \\ \frac{H}{1+H}, & k = 1. \end{cases} \tag{4.31}$$

□

We thus have convergence in probability in the aperiodic case, but (as usual) oscillations in the periodic case. It is well-known that the oscillations seen for various properties of tries tend to be numerically small; hence, the limits in (4.31) can be regarded as approximations also in the periodic case. Note that the limits in (4.31) depend on the letter probabilities $\mathbf{p}$ only through the entropy $H$, and that this limit distribution conditioned on being $\neq 1$ is independent of $\mathbf{p}$. In the periodic case, the asymptotics in (4.30) depend also on $d_{\mathbf{p}}$; as always, $\psi_{\mathsf{E},k}$ and $\psi_{\mathsf{E,i}}$ are given by (3.14) with the corresponding $f_{\mathsf{E}}^*$ in (4.4) and (4.14).

**Remark 4.5.** The result in Theorem 4.4 is of the quenched type, where we condition on the random tree $\mathcal{T}_n$ and obtain approximation or convergence in probability of the conditional distribution. By unconditioning, this immediately implies the corresponding annealed result, for the distribution of $|\mathcal{T}_n^*|_{\mathsf{e}}$ where we consider the combined random experiment of first choosing $\mathcal{T}_n$ at random and then a random fringe subtree of it. □

**Remark 4.6.** The asymptotic distribution in (4.31) has probabilities, say $\pi_k$, decaying as $k^{-2}$ for large $k$. This is similar to the distribution of the size (now defined as the number of nodes) of fringe trees in, for example, the random recursive tree (with $\pi_k = 1/(k(k+1))$, $k \geq 1$) and the binary search tree (with $\pi_k = 2/((k+1)(k+2))$, $k \geq 1$); see [1; 13; 14]. Recall that for conditioned Galton–Watson trees (with finite offspring variance), the probabilities decay more slowly, as $k^{-3/2}$, see [1; 19; 20]. □

The convergence in probability in Theorem 4.4 can be refined to asymptotic normality of the conditional probabilities. In order to include the case $k = 1$ in a notationally convenient way, we (re)define in the rest of this subsection

$$\psi_{\mathsf{E},1}(t) := H, \quad \psi_{\mathsf{C},1}(t) := H, \quad \psi_{\mathsf{V},1}(t) := H, \quad \psi_{\mathsf{V},1\mathsf{i}}(x) := \psi_{\mathsf{C,i}}(x). \tag{4.32}$$

Thus the first case in (4.30) holds also for $k = 1$. (Our main justification for the fudge (4.32) is that it works. One interpretation, and perhaps explanation, is that we replace $\Phi_1$ by the almost identical $\Phi_*$ in (3.59), which has $\chi_* = 0$ and $\psi_{\mathsf{X}}(t)$ as in (4.32), see (3.61)–(3.62).)

**Theorem 4.7.** *The conditional fringe tree size distribution of $\mathcal{T}_n$, given $\mathcal{T}_n$, has asymptotically normal fluctuations, in the following sense. Let $k \geq 1$ and let either $a_{kn} := \mathbb{P}\big(|\mathcal{T}_n^*|_{\mathsf{e}} = k\big) = \mathbb{E}\,\frac{\Phi_k(\mathcal{T}_n)}{|\mathcal{T}_n|}$, or $a_{kn} := \frac{\mathbb{E}\,\Phi_k(\mathcal{T}_n)}{\mathbb{E}\,|\mathcal{T}_n|}$. Then, with all moments, as $n \to \infty$,*

$$n^{1/2}\Big(\mathbb{P}\big(|\mathcal{T}_n^*|_{\mathsf{e}} = k \mid \mathcal{T}_n\big) - a_{kn}\Big) = n^{1/2}\Big(\frac{\Phi_k(\mathcal{T}_n)}{|\mathcal{T}_n|} - a_{kn}\Big) \overset{\mathrm{d}}{\approx} N\big(0, \widetilde{\sigma}_k^2(n)\big), \tag{4.33}$$

*where, with $t = \log n$ and $\psi_{\mathsf{E},+}(t) := \psi_{\mathsf{E},\mathsf{i}}(t) + H$,*

$$\begin{aligned}\widetilde{\sigma}_k^2(n) := \frac{H}{\psi_{\mathsf{E},+}(t)^2}\Big(\psi_{\mathsf{V},k}(t) - 2\frac{\psi_{\mathsf{E},k}(t)}{\psi_{\mathsf{E},+}(t)}\psi_{\mathsf{V},k\mathsf{i}}(t) + \frac{\psi_{\mathsf{E},k}(t)^2}{\psi_{\mathsf{E},+}(t)^2}\psi_{\mathsf{V},\mathsf{i}}(t)\Big) \\ - \frac{1}{\psi_{\mathsf{E},+}(t)^4}\Big(\psi_{\mathsf{E},+}(t)\psi_{\mathsf{C},k}(t) - \psi_{\mathsf{E},k}(t)\psi_{\mathsf{C},\mathsf{i}}(t)\Big)^2.\end{aligned} \tag{4.34}$$

*In particular, if $d_{\mathbf{p}} = 0$, then $\widetilde{\sigma}_k^2(n)$ is constant and,*

$$\begin{aligned}\widetilde{\sigma}_k^2(n) = \frac{H}{(1+H)^4}\Big((1+H)^2 f_{\mathsf{V},k}^*(-1) - \frac{2(1+H)}{k(k-1)} f_{\mathsf{V},k\mathsf{i}}^*(-1) \\ + \frac{f_{\mathsf{V},\mathsf{i}}^*(-1) - H}{k^2(k-1)^2}\Big), \qquad k \geq 2,\end{aligned} \tag{4.35}$$

$$\widetilde{\sigma}_1^2(n) = (1+H)^{-4}\big(H^3 f_{\mathsf{V},\mathsf{i}}^*(-1) - H^2\big). \tag{4.36}$$

*Moreover, the approximation in distribution (4.33) holds jointly for any finite number of $k$, with a multivariate normal distribution $N\big(0, (\widetilde{\sigma}_{k\ell}(n))_{k,\ell}\big)$.*

The asymptotic covariances $\widetilde{\sigma}_{k\ell}$ can be expressed similarly to the case $\ell = k$ in (4.34); we leave the details to the reader.

Note that in the periodic case $d_{\mathbf{p}} > 0$, the asymptotic variance (4.34) is a continuous periodic function of $\log n$. However there is no easy way to find its mean or other Fourier coefficients.

Theorem 4.7 follows from joint convergence in Theorems 4.1 and 4.2 by standard methods. We prove first a general lemma of standard type.

**Lemma 4.8.** *Let $(X_n, Y_n)$ be a sequence of random vectors, and assume that, as $n \to \infty$,*

$$n^{-1/2}\big(X_n - \mathbb{E}\,X_n, Y_n - \mathbb{E}\,Y_n\big) \approx N\Big(0, \Big(\begin{smallmatrix}\sigma_{XX}(n) & \sigma_{XY}(n) \\ \sigma_{XY}(n) & \sigma_{YY}(n)\end{smallmatrix}\Big)\Big), \tag{4.37}$$

*where $\mathbb{E}\,X_n = O(n)$, $\mathbb{E}\,Y_n = \Theta(n)$ and $\sigma_{XX}(n), \sigma_{XY}(n), \sigma_{YY}(n) = O(1)$.*

(i) *Then, with* $x_n := \mathbb{E}\,X_n$ *and* $y_n := \mathbb{E}\,Y_n$*,*

$$n^{1/2}\Bigl(\frac{X_n}{Y_n}-\frac{\mathbb{E}\,X_n}{\mathbb{E}\,Y_n}\Bigr)\approx N\Bigl(0,\frac{n^2}{y_n^2}\Bigl(\sigma_{XX}(n)-2\frac{x_n}{y_n}\sigma_{XY}(n)+\frac{x_n^2}{y_n^2}\sigma_{YY}(n)\Bigr)\Bigr), \tag{4.38}$$

(ii) *If, moreover,* (4.37) *holds with all moments, and* $Y_n \geq cn$ *a.s., for some* $c>0$ *and all* $n$*, then* (4.38) *holds with all moments. Furthermore, we then may replace* $\mathbb{E}\,X_n/\,\mathbb{E}\,Y_n$ *by* $\mathbb{E}\bigl(X_n/Y_n\bigr)$ *in* (4.38)*.*

*Proof.* (i): Denote the left-hand side of (4.37) by $(X_n',Y_n')$. Then

$$\begin{aligned}n^{1/2}\Bigl(\frac{X_n}{Y_n}-\frac{\mathbb{E}\,X_n}{\mathbb{E}\,Y_n}\Bigr)&=n^{1/2}\Bigl(\frac{x_n+n^{1/2}X_n'}{y_n+n^{1/2}Y_n'}-\frac{x_n}{y_n}\Bigr)=\frac{ny_nX_n'-nx_nY_n'}{y_n(y_n+n^{1/2}Y_n')}\\&=\frac{y_n}{y_n+n^{1/2}Y_n'}\cdot\frac{n}{y_n}\Bigl(X_n'-\frac{x_n}{y_n}Y_n'\Bigr),\end{aligned} \tag{4.39}$$

and (4.38) follows since $y_n/(y_n+n^{1/2}Y_n')\xrightarrow{\text{p}}1$. (By the subsequence principle, it suffices to consider subsequences such that $x_n/n$, $y_n/n$ and $\sigma_{XX}(n),\sigma_{XY}(n),\sigma_{YY}(n)$ converge.)

(ii): Let $r$ be a positive integer. The assumptions imply that the sequences $|X_n'|^r$ and $|Y_n'|^r$ are uniformly integrable, and then it follows that the $r$th absolute powers of the variables (4.39) are uniformly integrable. Hence the $r$th moment converges in (4.38).

In particular, (4.38) holds with the first moment, and thus

$$n^{1/2}\Bigl(\mathbb{E}\,\frac{X_n}{Y_n}-\frac{\mathbb{E}\,X_n}{\mathbb{E}\,Y_n}\Bigr)=\mathbb{E}\Bigl[n^{1/2}\Bigl(\frac{X_n}{Y_n}-\frac{\mathbb{E}\,X_n}{\mathbb{E}\,Y_n}\Bigr)\Bigr]\to 0. \tag{4.40}$$

Hence we may replace $\mathbb{E}\,X_n/\,\mathbb{E}\,Y_n$ by $\mathbb{E}\bigl(X_n/Y_n\bigr)$ in (4.38). ☐

*Proof of Theorem 4.7.* We apply Lemma 4.8 with $X_n := \Phi_k(\mathcal{T}_n)$ and $Y_n := |\mathcal{T}_n| = n+\Phi_{\mathsf{i}}(\mathcal{T}_n)$. As noted above, (4.37) then holds (with all moments) by Theorem 3.9 (or Theorems 4.1 and 4.2) together with Remark 3.10, if we define, using (3.32) and (4.20), with $t=\log n$,

$$\sigma_{XX}(n) := \widehat{\sigma}_k^2(n) = H^{-1}\psi_{\mathsf{V},k}(t)-H^{-2}\psi_{\mathsf{C},k}(t)^2, \tag{4.41}$$

$$\sigma_{XY}(n) := \widehat{\sigma}_{k\mathsf{i}}(n) = H^{-1}\psi_{\mathsf{V},k\mathsf{i}}(t)-H^{-2}\psi_{\mathsf{C},k}(t)\psi_{\mathsf{C},\mathsf{i}}(t), \tag{4.42}$$

$$\sigma_{YY}(n) := \widehat{\sigma}_{\mathsf{i}}^2(n) = H^{-1}\psi_{\mathsf{V},\mathsf{i}}(t)-H^{-2}\psi_{\mathsf{C},\mathsf{i}}(t)^2. \tag{4.43}$$

Furthermore, (3.37) yields

$$x_n/n=\mathbb{E}\,\Phi_k(\mathcal{T}_n)/n=H^{-1}\psi_{\mathsf{E},k}(t)+o(1), \tag{4.44}$$

$$y_n/n=1+\mathbb{E}\,\Phi_{\mathsf{i}}(\mathcal{T}_n)/n=1+H^{-1}\psi_{\mathsf{E},\mathsf{i}}(t)+o(1)=H^{-1}\psi_{\mathsf{E},+}(t)+o(1). \tag{4.45}$$

Note that, as required by Lemma 4.8, $x_n/n=O(1)$ and $y_n/n=\Theta(1)$. Note further that (4.41), (4.42) and (4.44) hold also for $k=1$ by our special definition (4.32). (Trivially, with $\sigma_{XX}(n) = \sigma_{XY}(n) = 0$ and $x_n = n$; recall that $\Phi_1(\mathcal{T}_n) = n$ is deterministic.) We have, by (4.41)–(4.45),

$$\begin{aligned}
&\frac{n^2}{y_n^2}\Bigl(\sigma_{XX}(n) - 2\frac{x_n}{y_n}\sigma_{XY}(n) + \frac{x_n^2}{y_n^2}\sigma_{YY}(n)\Bigr) \\
&= \frac{H}{\psi_{\mathsf{E},+}(t)^2}\Bigl(\psi_{\mathsf{V},k}(t) - 2\frac{\psi_{\mathsf{E},k}(t)}{\psi_{\mathsf{E},+}(t)}\psi_{\mathsf{V},k\mathsf{i}}(t) + \frac{\psi_{\mathsf{E},k}(t)^2}{\psi_{\mathsf{E},+}(t)^2}\psi_{\mathsf{V},\mathsf{i}}(t)\Bigr) \\
&\quad - \frac{1}{\psi_{\mathsf{E},+}(t)^2}\Bigl(\psi_{\mathsf{C},k}(t)^2 - 2\frac{\psi_{\mathsf{E},k}(t)}{\psi_{\mathsf{E},+}(t)}\psi_{\mathsf{C},k}(t)\psi_{\mathsf{C},\mathsf{i}}(t) + \frac{\psi_{\mathsf{E},k}(t)^2}{\psi_{\mathsf{E},+}(t)^2}\psi_{\mathsf{C},\mathsf{i}}(t)^2\Bigr) + o(1) \\
&= \frac{H}{\psi_{\mathsf{E},+}(t)^2}\Bigl(\psi_{\mathsf{V},k}(t) - 2\frac{\psi_{\mathsf{E},k}(t)}{\psi_{\mathsf{E},+}(t)}\psi_{\mathsf{V},k\mathsf{i}}(t) + \frac{\psi_{\mathsf{E},k}(t)^2}{\psi_{\mathsf{E},+}(t)^2}\psi_{\mathsf{V},\mathsf{i}}(t)\Bigr) \\
&\quad - \frac{1}{\psi_{\mathsf{E},+}(t)^4}\Bigl(\psi_{\mathsf{E},+}(t)\psi_{\mathsf{C},k}(t) - \psi_{\mathsf{E},k}(t)\psi_{\mathsf{C},\mathsf{i}}(t)\Bigr)^2 + o(1)
\end{aligned} \tag{4.46}$$

which equals $\widetilde{\sigma}_k^2(n) + o(1)$ as defined in (4.34). Thus, Lemma 4.8 yields (4.33) with all moments. (Note that $Y_n \geq n$ a.s., so Lemma 4.8(ii) applies.)

When $d_{\mathbf{p}} = 0$, $\psi_{\mathsf{E},+}(t) = f^*_{\mathsf{E},\mathsf{i}}(-1) + H = 1 + H$ by (3.13) and (4.5), and (4.34) reduces to (4.35)–(4.36), using also (4.15) and (3.23). □

4.3. **Distribution of fringe tries.** The previous subsection studied the sizes of fringe tries. For a more detailed study of the distribution of the fringe trees of the random trie $\mathcal{T}_n$, let $T$ be a fixed trie, and consider the toll function

$$\varphi_T(T') := \mathbf{1}\{T' = T\} \tag{4.47}$$

and the corresponding additive functional $\Phi_T$ which counts the number of fringe trees equal to $T$. Let $k = |T|_{\mathsf{e}}$, and let $p_T := \mathbb{P}(\mathcal{T}_k = T)$. Note that $\varphi_\bullet$ is as defined in Example 2.1, and coincides with $\varphi_1$ in Section 4.2, so we are mainly interested in the case $k \geq 2$; then $\chi_T := \varphi_T(\bullet) = 0$. For completeness, we include below also the case $T = \bullet$, but in this case we use the special definitions (4.32); thus $\psi_{\mathsf{X},\bullet} := \psi_{\mathsf{X},1}$.

The functional $\Phi_T$ is not increasing, but with $\Phi_{>k} := \Phi_{\geq k+1}$ defined in Section 4.2, $\Phi_T + \Phi_{>k}$ is increasing, and thus Theorems 3.9 and 3.12 apply to $\Phi_T = (\Phi_T + \Phi_{>k}) - \Phi_{>k}$. Furthermore, Lemma 3.14 applies (with $n_0 = k+1$ and $a_n = 0$). Consequently, the arguments in Section 4.2 yield the following analogues of Theorems 4.2, 4.4, and 4.7, using also (4.54) which we postpone until after the theorems.

**Theorem 4.9.** *Let $T$ be a fixed trie and consider $\Phi_T$, the number of fringe trees equal to $T$ (as ordered trees). Then, the central limit theorems* (3.29)–(3.30) *and* (3.34)–(3.35) *hold, with all [absolute] moments, and with asymptotic variances given by* (3.31)–(3.32) *(and thus by* (3.27)–(3.28) *when $d_{\mathbf{p}} = 0$); this extends to joint convergence for several tries $T$. Furthermore, the means satisfy* (3.36) *and* (3.37)*, and the laws of large numbers* (3.38)–(3.39) *hold.* □

**Theorem 4.10.** *The fringe tree distribution of $\mathcal{T}_n$ satisfies*

$$\mathbb{P}\bigl(\mathcal{T}_n^* = T \mid \mathcal{T}_n\bigr) = \frac{\Phi_T(\mathcal{T}_n)}{|\mathcal{T}_n|} = \frac{\psi_{\mathsf{E},T}(\log n)}{\psi_{\mathsf{E},\mathsf{i}}(\log n) + H} + o_{\mathrm{p}}(1). \tag{4.48}$$

*In particular, if $d_{\mathbf{p}} = 0$, then the distribution converges in probability:*

$$\mathbb{P}\bigl(|\mathcal{T}_n^*|_{\mathsf{e}} = T \mid \mathcal{T}_n\bigr) \overset{\mathrm{p}}{\longrightarrow} \begin{cases} \frac{p_T}{(1+H)k(k-1)}, & |T| \geq 2, \\ \frac{H}{1+H}, & T = \bullet. \end{cases} \tag{4.49}$$

□

**Theorem 4.11.** *The conditional fringe tree distribution of $\mathcal{T}_n$, given $\mathcal{T}_n$, has asymptotically normal fluctuations, in the following sense. Let $T$ be a fixed trie and let either $a_{T,n} := \mathbb{P}(\mathcal{T}_n = k) = \mathbb{E}\,\frac{\Phi_T(\mathcal{T}_n)}{|\mathcal{T}_n|}$, or $a_{T,n} := \frac{\mathbb{E}\,\Phi_T(\mathcal{T}_n)}{\mathbb{E}\,|\mathcal{T}_n|}$. Then, with all moments, as $n \to \infty$,*

$$n^{1/2}\Bigl(\mathbb{P}\bigl(\mathcal{T}_n^* = T \mid \mathcal{T}_n\bigr) - a_{T,n}\Bigr) = n^{1/2}\Bigl(\frac{\Phi_T(\mathcal{T}_n)}{|\mathcal{T}_n|} - a_{T,n}\Bigr) \overset{\mathrm{d}}{\approx} N\bigl(0, \widetilde{\sigma}_T^2(n)\bigr), \tag{4.50}$$

*where, with $t = \log n$ and $\psi_{\mathsf{E},+}(t) := \psi_{\mathsf{E},\mathsf{i}}(t) + H$,*

$$\begin{aligned} \widetilde{\sigma}_T^2(n) := {} & \frac{H}{\psi_{\mathsf{E},+}(t)^2}\Bigl(\psi_{\mathsf{V},T}(t) - 2\frac{\psi_{\mathsf{E},T}(t)}{\psi_{\mathsf{E},+}(t)}\psi_{\mathsf{V},T\mathsf{i}}(t) + \frac{\psi_{\mathsf{E},T}(t)^2}{\psi_{\mathsf{E},+}(t)^2}\psi_{\mathsf{V},\mathsf{i}}(t)\Bigr) \\ & - \frac{1}{\psi_{\mathsf{E},+}(t)^4}\Bigl(\psi_{\mathsf{E},+}(t)\psi_{\mathsf{C},T}(t) - \psi_{\mathsf{E},T}(t)\psi_{\mathsf{C},\mathsf{i}}(t)\Bigr)^2. \end{aligned} \tag{4.51}$$

*In particular, if $d_{\mathbf{p}} = 0$, then $\widetilde{\sigma}_T^2(n)$ is constant. Moreover, the approximation in distribution* (4.50) *holds jointly for any finite number of $T$, with a multivariate normal distribution $N\bigl(0, (\widetilde{\sigma}_{TT'}(n))_{T,T'}\bigr)$.*

Asymptotic means, variances and covariances may be calculated as in Section 4.2. Suppose $k := |T|_{\mathsf{e}} \geq 2$. Then, recalling (4.13),

$$\begin{aligned} f_{\mathsf{E},T}(\lambda) &= \mathbb{P}(\widetilde{\mathcal{T}}_\lambda = T) = \mathbb{P}(N_\lambda = k)\,\mathbb{P}\bigl(\widetilde{\mathcal{T}}_\lambda = T \mid N_\lambda = k\bigr) \\ &= \mathbb{P}(N_\lambda = k)\,\mathbb{P}(\mathcal{T}_k = T) = p_T f_{\mathsf{E},k}(\lambda) = p_T \frac{\lambda^k}{k!}e^{-\lambda}. \end{aligned} \tag{4.52}$$

Hence, using (4.14)–(4.15), for $\operatorname{Re} s > -k$,

$$f_{\mathsf{E},T}^*(s) = p_T f_{\mathsf{E},k}^*(s) = p_T \frac{\Gamma(s+k)}{k!}. \tag{4.53}$$

and

$$f_{\mathsf{E},T}^*(-1) = \frac{p_T}{k(k-1)}. \tag{4.54}$$

Furthermore, if $|\boldsymbol{\alpha}| > 0$, then $\varphi_T(\widetilde{\mathcal{T}}_\lambda)\varphi_T(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = 0$. Hence, cf. (4.16),

$$\begin{aligned} \operatorname{Cov}\bigl(\varphi_T(\widetilde{\mathcal{T}}_\lambda), \varphi_T(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\bigr) &= f_{\mathsf{E},T}(\lambda)\mathbf{1}\{|\boldsymbol{\alpha}| = 0\} - f_{\mathsf{E},T}(\lambda)f_{\mathsf{E},T}(P(\boldsymbol{\alpha})\lambda) \\ &= f_{\mathsf{E},T}(\lambda)\mathbf{1}\{|\boldsymbol{\alpha}| = 0\} - p_T^2\frac{\lambda^k(P(\boldsymbol{\alpha})\lambda)^k}{k!^2}e^{-(1+P(\boldsymbol{\alpha}))\lambda}. \end{aligned} \tag{4.55}$$

Consequently, by (3.44) and (4.53), cf. (4.17), for $\operatorname{Re} s > -k$,

$$f^*_{\mathsf{V},T}(s) = f^*_{\mathsf{E},T}(s) - p_T^2 \sum_{\boldsymbol{\alpha}}^* \int_0^\infty \frac{P(\boldsymbol{\alpha})^k \lambda^{2k}}{k!^2} e^{-(1+P(\boldsymbol{\alpha}))\lambda} \lambda^{s-1} \,\mathrm{d}\lambda$$
$$= p_T \frac{\Gamma(k+s)}{k!} - p_T^2 \frac{\Gamma(s+2k)}{k!^2} \sum_{\boldsymbol{\alpha}}^* \frac{P(\boldsymbol{\alpha})^k}{(1+P(\boldsymbol{\alpha}))^{s+2k}}. \tag{4.56}$$

In particular,

$$f^*_{\mathsf{V},T}(-1) = \frac{p_T}{k(k-1)} - p_T^2 \frac{(2k-2)!}{k!^2} \sum_{\boldsymbol{\alpha}}^* \frac{P(\boldsymbol{\alpha})^k}{(1+P(\boldsymbol{\alpha}))^{2k-1}}. \tag{4.57}$$

We leave further calculations of variances and covariances to the reader.

**Example 4.12** (Cherries)**.** A cherry is the tree $T_{\mathrm{ch}}$ with one internal node (the root) and two external nodes. This is a trie generated by two strings with different first letters. Suppose for simplicity that $\mathcal{A} = \{0,1\}$, and write $p_0 = p$, $p_1 = q$. Then $p_{T_{\mathrm{ch}}} = \mathbb{P}(\mathcal{T}_2 = T_{\mathrm{ch}}) = 2pq$. Hence, (4.53) and (4.56) yield

$$f^*_{\mathsf{E},T_{\mathrm{ch}}}(s) = pq\Gamma(s+2), \tag{4.58}$$
$$f^*_{\mathsf{V},T_{\mathrm{ch}}}(s) = pq\Gamma(s+2) - p^2q^2\Gamma(s+4) \sum_{\boldsymbol{\alpha}}^* \frac{P(\boldsymbol{\alpha})^2}{(1+P(\boldsymbol{\alpha}))^{s+4}}. \tag{4.59}$$

□

4.4. **Protected nodes.** The *rank* of a node $v$ in a rooted tree is the minimum distance to a descendant of $v$ that is a leaf. (In particular, leaves are the nodes with rank 0.) For a trie $T$ and a node $\boldsymbol{\alpha} \in T$, we thus have, recalling that $\nu_{\boldsymbol{\alpha}}$ is the number of the generating strings that have $\boldsymbol{\alpha}$ as a prefix, cf. (2.13),

$$\operatorname{rank}(\boldsymbol{\alpha}) := \min\bigl\{|\boldsymbol{\beta}| : \boldsymbol{\alpha\beta} \text{ is a leaf in } T\bigr\} = \min\bigl\{|\boldsymbol{\beta}| : \nu_{\boldsymbol{\alpha\beta}} = 1\bigr\}. \tag{4.60}$$

Nodes with rank $\geq k$ are called *k-protected.* Here $k \geq 0$; the interesting cases are $k \geq 2$. (For $k = 1$ we get just the internal nodes. The results below then reduce to corresponding results in Section 4.1.)

Let $\Phi_{k\text{-prot}}(T)$ be the number of $k$-protected nodes in $T$. This is an additive functional with toll function, for $T \neq \emptyset$,

$$\varphi_{k\text{-prot}}(T) := \mathbf{1}\{\text{the root } \boldsymbol{\epsilon} \text{ of } T \text{ is } k\text{-protected}\} = \mathbf{1}\{\operatorname{rank}(\boldsymbol{\epsilon}) \geq k\}. \tag{4.61}$$

$\Phi_{k\text{-prot}}$ is not an additive functional, since adding a new leaf may make some nodes unprotected. However, the only nodes that may lose protection are the $k-1$ nearest ancestors of the new leaf, and thus $\Phi_{k\text{-prot}} + k\Phi_\bullet$ is an increasing functional. Hence, Theorems 3.9 and 3.12 apply to $\Phi_{k\text{-prot}}$, and we obtain analogues of Theorems 4.2–4.7 and 4.9–4.11 yielding asymptotic normal distributions of the number and proportion of $k$-protected nodes. (We omit detailed statements.)

At least the asymptotic mean is rather easily calculated. For a trie $T$ we have, using (4.61) and (4.60), for $k \geq 1$ and including the case $T = \emptyset$,

$$\varphi_{k\text{-prot}}(T) = \mathbf{1}\{\nu_{\boldsymbol{\beta}} \neq 1 \ \forall \boldsymbol{\beta} \in \mathcal{A}^{k-1}\} - \mathbf{1}\{\nu_{\boldsymbol{\epsilon}} = 0\}. \tag{4.62}$$

In particular, for the Poisson random trie $\widetilde{\mathcal{T}}_\lambda$, where $\nu_{\boldsymbol{\beta}} = N_{\lambda,\boldsymbol{\beta}}$,

$$\varphi_{k\text{-prot}}(\widetilde{\mathcal{T}}_\lambda) = \mathbf{1}\{N_{\lambda,\boldsymbol{\alpha}} \neq 1 \ \forall \boldsymbol{\alpha} \in \mathcal{A}^{k-1}\} - \mathbf{1}\{N_\lambda = 0\}. \tag{4.63}$$

Hence, since $N_{\lambda,\boldsymbol{\alpha}} \sim \operatorname{Po}(P(\boldsymbol{\alpha})\lambda)$ are independent for $\boldsymbol{\alpha} \in \mathcal{A}^{k-1}$,

$$f_{\mathsf{E}}(\lambda) = \mathbb{E}\,\varphi_{k\text{-prot}}(\widetilde{\mathcal{T}}_\lambda) = \prod_{\boldsymbol{\alpha}\in\mathcal{A}^{k-1}} \bigl(1 - P(\boldsymbol{\alpha})\lambda e^{P(\boldsymbol{\alpha})\lambda}\bigr) - e^{-\lambda} \tag{4.64}$$

$$= \sum_{\emptyset\neq S\subseteq\mathcal{A}^{k-1}} (-1)^{|S|} \prod_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha}) \cdot e^{-\sum_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha})\lambda} \lambda^{|S|} - \bigl(e^{-\lambda} - 1\bigr). \tag{4.65}$$

For $-1 < \operatorname{Re} s < 0$, the Mellin transform $f^*_{\mathsf{E}}(s)$ can be calculated using (4.65) in (2.27) and integrating termwise, yielding

$$f^*_{\mathsf{E}}(s) = \sum_{\emptyset\neq S\subseteq\mathcal{A}^{k-1}} (-1)^{|S|} \prod_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha}) \cdot \Bigl(\sum_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha})\Bigr)^{-|S|-s} \Gamma(|S|+s) - \Gamma(s). \tag{4.66}$$

We know that $f^*_{\mathsf{E}}$ is analytic in the strip $-2 < \operatorname{Re} s < 0$, see Remark 3.4, and the right-hand side of (4.66) is analytic for $-2 < \operatorname{Re} s < 0$ except possibly at $s = -1$. Hence, (4.66) holds in this strip, with a removable singularity at $-1$. To find $f^*_{\mathsf{E}}(-1)$, let $g(s)$ be the sum over $|S| \geq 2$ in (4.66); then, using (2.5),

$$\begin{aligned} f^*_{\mathsf{E}}(s) &= -\sum_{\boldsymbol{\alpha}\in\mathcal{A}^{k-1}} P(\boldsymbol{\alpha})^{-s}\Gamma(s+1) + g(s) - \Gamma(s) \\ &= -\Gamma(s+1)\rho(-s)^{k-1} - \Gamma(s) + g(s) \\ &= -\frac{\Gamma(s+2)}{s} \cdot \frac{s\rho(-s)^{k-1}+1}{s+1} + g(s) \end{aligned} \tag{4.67}$$

and thus, letting $s \to -1$ and recalling (2.8),

$$\begin{aligned} f^*_{\mathsf{E}}(-1) &= \frac{\mathrm{d}}{\mathrm{d}s}\bigl(s\rho(-s)^{k-1}\bigr)\big|_{s=-1} + g(-1) \\ &= 1 - (k-1)\frac{\mathrm{d}}{\mathrm{d}s}\rho(-s)\big|_{s=-1} + g(-1) \\ &= 1 - (k-1)H + \sum_{S\subseteq\mathcal{A}^{k-1},|S|\geq 2} (-1)^{|S|} \frac{\prod_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha})}{\bigl(\sum_{\boldsymbol{\alpha}\in S} P(\boldsymbol{\alpha})\bigr)^{|S|-1}} (|S|-2)!. \end{aligned} \tag{4.68}$$

As in earlier applications, this yields asymptotics for the mean. $f^*_{\mathsf{V}}$ and variance asymptotics may be calculated by similar arguments, but the results are more complicated and we omit the details.

**Example 4.13.** For the number of 2-protected nodes in a binary trie we have $k = 2$ and $\mathcal{A} = \{0, 1\}$, and then (4.66) and (4.68) yield

$$f^*_{\mathsf{E}}(s) = -\bigl(p_0^{-s} + p_1^{-s}\bigr)\Gamma(s+1) + p_0p_1\Gamma(s+2) - \Gamma(s) \tag{4.69}$$

with

$$f^*_{\mathsf{E}}(-1) = 1 - H + p_0p_1. \tag{4.70}$$

In particular, for $p_0 = \frac{1}{2}$, $f^*_{\mathsf{E}}(-1) = \frac{5}{4} - \log 2 \doteq 0.55685$. Hence, by the analogue of Theorems 4.4 and 4.10, for a large random symmetric binary trie the proportion of 2-protected nodes is roughly (ignoring small oscillations),

$$\frac{f^*_{\mathsf{E}}(-1)}{1+H} = \frac{5/4 - \log 2}{1 + \log 2} \doteq 0.32888. \tag{4.71}$$

For comparison, the corresponding proportion in a binary search tree converges (in probability) to $11/30 \doteq 0.36667$ [24; 3; 5; 13]; in a uniformly random binary tree the proportion converges to $33/64 = 0.515625$ [5].

In this example, one can also use the easily verified fact that for any binary tree $T$ with $|T|_{\mathsf{e}} > 1$, with $T_{\mathsf{ch}}$ the cherry in Example 4.12,

$$\Phi_{2\text{-prot}}(T) = \Phi_{\mathsf{i}}(T) - \Phi_{\bullet} + \Phi_{T_{\mathsf{ch}}}. \tag{4.72}$$

Hence results in this case alternatively follow from results in the Sections 4.1–4.3. □

In general, the sums in (4.66) and (4.68) have almost $2^{|\mathcal{A}|^{k-1}}$ terms, which quickly becomes very large for larger $k$ or $|\mathcal{A}|$. However, in the symmetric case, the sums simplify by symmetry since the summands then depend only on $|S|$.

**Example 4.14.** Consider the symmetric case with $|\mathcal{A}| = r \geq 2$ and $p_\alpha = 1/r$ for all $\alpha \in \mathcal{A}$. We calculate $f^*_{\mathsf{E}}$, which we denote by $f^*_{\mathsf{E},k\text{-prot},r}$.

For $k = 2$, (4.66) and (4.68) yield

$$f^*_{\mathsf{E},2\text{-prot},r}(s) = \sum_{j=1}^{r} \binom{r}{j} (-1)^j r^s j^{-j-s} \Gamma(s+j) - \Gamma(s). \tag{4.73}$$

$$\begin{aligned} f^*_{\mathsf{E},2\text{-prot},r}(-1) &= 1 - \log r + \sum_{j=2}^{r} \binom{r}{j} (-1)^j r^{-1} j^{1-j} (j-2)! \\ &= 1 - \log r + \sum_{j=2}^{r} (-1)^j \frac{(r-1)!}{(r-j)!\,(j-1)j^j}. \end{aligned} \tag{4.74}$$

Furthermore, for general $k \geq 2$, (4.66) implies

$$f^*_{\mathsf{E},k\text{-prot},r}(s) = f^*_{\mathsf{E},2\text{-prot},r^{k-1}}(s). \tag{4.75}$$

For example, for the binary case and $k = 3, 4$,

$$f^*_{\mathsf{E},3\text{-prot},2}(-1) = f^*_{\mathsf{E},2\text{-prot},4}(-1) = \frac{1897}{1152} - 2\log 2 \doteq 0.26041, \tag{4.76}$$

| $k$ | $f^*_{\mathsf{E},k\text{-prot},2}$ | $f^*_{\mathsf{E},k\text{-prot},2}/(1+\log 2)$ |
|---|---|---|
| 1 | 1 | 0.59061 |
| 2 | 0.55685 | 0.32888 |
| 3 | 0.26040 | 0.15380 |
| 4 | 0.10884 | 0.06428 |
| 5 | 0.04718 | 0.02786 |
| 6 | 0.02182 | 0.01289 |
| 7 | 0.01039 | 0.00613 |
| 8 | 0.00502 | 0.00296 |
| 9 | 0.00244 | 0.00144 |
| 10 | 0.00120 | 0.00070 |

TABLE 1. Approximate proportions of $k$-protected nodes in symmetric random binary tries (right column).

$$f^*_{\mathsf{E},4\text{-prot},2}(-1) = f^*_{\mathsf{E},2\text{-prot},8}(-1) = \frac{13666493449090877}{6245298339840000} - 3\log 2 \doteq 0.10884. \tag{4.77}$$

Recall that the asymptotic proportion of $k$-protected nodes, ignoring the oscillations, equals $f^*_{\mathsf{E},k\text{-prot},2}(-1)/(1+H)$, where $H = \log 2$. Table 1 gives numerical values for small $k$. ☐

The numerical values in Table 1 suggest that the proportions decrease geometrically as $k \to \infty$. In fact, this holds for any $r$.

**Theorem 4.15.** *Consider symmetric tries as in Example 4.14, and assume $r, k \geq 2$. As $k \to \infty$ or $r \to \infty$ (or both),*

$$f^*_{\mathsf{E},k\text{-prot},r}(-1) \sim \frac{1}{2r^{k-1}}. \tag{4.78}$$

*In particular, for symmetric binary tries,*

$$f^*_{\mathsf{E},k\text{-prot},2}(-1) \sim 2^{-k}. \tag{4.79}$$

In other words, for large $k$ and much larger $n$, the proportion of $k$-protected nodes in a symmetric binary trie is roughly (again ignoring oscillations) $2^{-k}/(1+\log 2)$.

*Proof.* By (4.75), it suffices to consider $k = 2$. In this case, (4.64) yields

$$f^*_{\mathsf{E},2\text{-prot},r}(-1) = \int_0^\infty f_{\mathsf{E},2\text{-prot},r}(\lambda)\lambda^{-2}\,\mathrm{d}\lambda = \int_0^\infty \Bigl(\Bigl(1 - \frac{\lambda}{r}e^{-\lambda/r}\Bigr)^r - e^{-\lambda}\Bigr)\frac{\mathrm{d}\lambda}{\lambda^2}. \tag{4.80}$$

Let

$$g_r(x) := \Bigl(1 - \frac{x}{r}e^{-x/r}\Bigr)^r - e^{-x}. \tag{4.81}$$

Note first that as $r \to \infty$, by the change of variables $x = ru$ and dominated convergence,

$$r \int_r^\infty g_r(x) \frac{\mathrm{d}x}{x^2} = \int_1^\infty g_r(ru) \frac{\mathrm{d}u}{u^2} \le \int_1^\infty \bigl(1 - ue^{-u}\bigr)^r \frac{\mathrm{d}u}{u^2} \to 0. \tag{4.82}$$

Furthermore, for $x \in [0, r]$, write $y_1 := 1 - \frac{x}{r} e^{-x/r}$ and $y_0 := e^{-x/r}$, so $g_r(x) = y_1^r - y_0^r$. For $y \in [0, 1]$, we have $0 \le (1 - ye^{-y}) - e^{-y} \le y^2/2$, and

$$(1 - ye^{-y}) - e^{-y} = \tfrac{1}{2} y^2 + O(y^3), \tag{4.83}$$

and thus, by the mean value theorem, for some $\theta = \theta(x, r) \in [0, 1]$,

$$g_r(x) = y_1^r - y_0^r = (y_1 - y_0) r (y_0 + \theta(y_1 - y_0))^{r-1} \tag{4.84}$$

$$= r\Bigl(\frac{1}{2}\Bigl(\frac{x}{r}\Bigr)^2 + O\Bigl(\frac{x}{r}\Bigr)^3\Bigr)\Bigl(1 - \frac{x}{r} + O\Bigl(\frac{x}{r}\Bigr)^2\Bigr)^{r-1}. \tag{4.85}$$

Hence, for fixed $x \ge 0$, $r g_r(x) \to \frac{1}{2} x^2 e^{-x}$ as $r \to \infty$. Moreeover, again by (4.84), for $x \in [0, r]$ and $r \ge 2$,

$$r g_r(x) \le r^2 (y_1 - y_0) y_1^{r-1} \le r^2 \Bigl(\frac{x}{r}\Bigr)^2 \Bigl(1 - \frac{x}{r} e^{-1}\Bigr)^{r-1} \le x^2 e^{-x/(2e)}. \tag{4.86}$$

Consequently, as $r \to \infty$, dominated convergence yields

$$r \int_0^r g_r(x) \frac{\mathrm{d}x}{x^2} \to \int_0^\infty \frac{1}{2} x^2 e^{-x} \frac{\mathrm{d}x}{x^2} = \frac{1}{2}, \tag{4.87}$$

which together with (4.82) and (4.80)–(4.81) yields the result. □

**Problem 4.16.** Extend these results to the non-symmetric case. In particular, for a general $\mathbf{p}$, does $f^*_{\mathsf{E},k\text{-prot}}$ decrease geometrically as $k \to \infty$? If so, at which rate?

Some similar (but less complete) results for binary and $m$-ary search trees are given in [4] and [14, Section 10.1].

4.5. **Number of subtrees.** Let $s(T)$ be the number of subtrees of a tree $T$, and $s_1(T)$ the number of subtrees that contain the root. Then, as noted by Wagner [32, 33], $\Phi(T) := \log(1 + s_1(T))$ is an additive functional with toll function

$$\varphi(T) := \log\bigl(1 + 1/s_1(T)\bigr). \tag{4.88}$$

The functional $\varphi$ is bounded (by $\log 2$). Moreover, $\Phi(T)$ is an increasing functional, and thus Theorems 3.9 and 3.12 apply and yield asymptotic normality for $\Phi(\mathcal{T}_n)$. This time we do not see a simple argument showing $\operatorname{Var} \Phi(\mathcal{T}_n) = \Omega(n)$, so we cannot apply (3.34)–(3.35); nevertheless (3.29)–(3.30) hold, and we obtain the following theorem. (We conjecture that $\operatorname{Var} \Phi(\mathcal{T}_n) = \Omega(n)$ in this application too, but leave this as an open problem.)

**Theorem 4.17.** *As $n \to \infty$,*

$$\frac{\log s(\mathcal{T}_n) - \mathbb{E}[\log s(\mathcal{T}_n)]}{\sqrt{n}} \approx \frac{\log s_1(\mathcal{T}_n) - \mathbb{E}[\log s_1(\mathcal{T}_n)]}{\sqrt{n}} \approx N\bigl(0, \widehat{\sigma}^2(n)\bigr), \tag{4.89}$$

*with all [absolute] moments, where $\widehat{\sigma}^2(n)$ is a continuous bounded function given by* (3.32)*.*

*Proof.* We have $\Phi(T) = \log s_1(T) + O(1)$ and $s_1(T) \le s(T) \le |T| s_1(T)$ (see [32; 33]), and thus, recalling $|\mathcal{T}_n| = n + \Phi_{\mathsf{i}}(\mathcal{T}_n)$ and using (3.39) for $\Phi_{\mathsf{i}}$,

$$\log s(\mathcal{T}_n) = \log s_1(\mathcal{T}_n) + O\bigl(\log|\mathcal{T}_n|\bigr) = \Phi(\mathcal{T}_n) + O\bigl(\log n\bigr) + O_{\mathrm{p}}(1), \tag{4.90}$$

where as usual $O_{\mathrm{p}}(1)$ denotes a random variable (depending on $n$) that is bounded in probability. Furthermore, for any fixed $m \ge 1$, by Theorem 4.1,

$$\mathbb{E}\bigl[\log^m |\mathcal{T}_n|\bigr] \le C_m \, \mathbb{E}\bigl[|\mathcal{T}_n|^{m/4}\bigr] \le C_m n^{m/4} = o\bigl(n^{m/2}\bigr). \tag{4.91}$$

Taking $m = 1$, we obtain from (4.90) and (4.91),

$$\mathbb{E}\bigl[\log s(\mathcal{T}_n)\bigr] = \mathbb{E}\bigl[\log s_1(\mathcal{T}_n)\bigr] + o\bigl(n^{1/2}\bigr) = \mathbb{E}\,\Phi(\mathcal{T}_n) + o\bigl(n^{1/2}\bigr), \tag{4.92}$$

The asymptotic normality (3.30) in Theorem 3.9 together with (4.90) and (4.91) yields (4.89), with [absolute] moments. ☐

Cf. similar results for some other classes of random trees in [32; 33] and [20].

4.6. **Shape parameter.** The *shape parameter* is defined as the logarithm of the product of all fringe tree sizes; this is thus an additive functional $\Phi(T)$ with toll function $\varphi(T) = \log|T|$. The shape functional $\Phi(T)$ is increasing. However, $\varphi(T)$ is unbounded, so we cannot use Theorems 3.9 and 3.12 as stated. Nevertheless, we have by Theorem 4.1, as in (4.91), for any $r \ge 1$,

$$\mathbb{E}\bigl[\log^r |\widetilde{\mathcal{T}}_\lambda|\bigr] \le C_r \, \mathbb{E}\bigl[|\widetilde{\mathcal{T}}_\lambda|^{r/4}\bigr] \le C_r \lambda^{r/4}. \tag{4.93}$$

In particular, (3.1), (3.2) and (5.5) (for any $r$) hold, and thus by Remark 3.8, or using Theorem 5.5 below, we find, for example,

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{n}} \approx N\bigl(0, \widehat{\sigma}^2(n)\bigr), \tag{4.94}$$

with all moments.

Cf. similar results for some other classes of random trees in [25], [7] and [33].

4.7. **Bucket tries.** The results above are easily adapted to bucket tries for a fixed bucket size $b$, by noting that the internal nodes of a bucket trie are precisely the nodes $\boldsymbol{\alpha}$ of the corresponding trie with $\nu_{\boldsymbol{\alpha}} > b$. In particular, if the bucket tries corresponding to $\widetilde{\mathcal{T}}_\lambda$ and $\mathcal{T}_n$ are denoted $\widetilde{\mathcal{T}}_\lambda^{(b)}$ and $\mathcal{T}_n^{(b)}$, then $|\widetilde{\mathcal{T}}_\lambda^{(b)}|_{\mathsf{i}} = \Phi_{>b}(\widetilde{\mathcal{T}}_\lambda)$ and $|\mathcal{T}_n^{(b)}|_{\mathsf{i}} = \Phi_{>b}(\mathcal{T}_n)$, and it follows that Theorem 4.1

holds for $\widetilde{\mathcal{T}}_\lambda^{(b)}$ and $\mathcal{T}_n^{(b)}$ too. We have, generalizing the case $b=1$ in (4.3)–(4.5),

$$f_{\mathsf{E}}(\lambda) = f_{\mathsf{E},>b}(\lambda) = \mathbb{P}(N_\lambda > b) = 1 - \sum_{i=0}^{b} \frac{\lambda^i}{i!} e^{-\lambda} \tag{4.95}$$

and thus,

$$f^*_{\mathsf{E}}(\lambda) = \int_0^\infty \Bigl(1 - \sum_{i=0}^{b} \frac{\lambda^i}{i!} e^{-\lambda}\Bigr) \lambda^{s-1} \,\mathrm{d}\lambda = -\frac{\Gamma(s+b+1)}{b!\, s}, \tag{4.96}$$

$$f^*_{\mathsf{E}}(-1) = \frac{1}{b}. \tag{4.97}$$

Consider now the number of buckets containing exactly $k$ strings, for some fixed $k \in \{1,\dots,b\}$. If we assume $n > b$ (so the root is internal), this equals the additive functional $\Phi_{b;k}$ with toll function

$$\varphi_{b;k}(T) = \sum_{\alpha\in\mathcal{A}} \mathbf{1}\{\nu_\epsilon > b,\, \nu_\alpha = k\}. \tag{4.98}$$

$\Phi_{b;k}$ is not increasing, but $\Phi_{b;k} + \Phi_\bullet$ is, so Theorems 3.9 and 3.12 apply to $\Phi_{b;k} = (\Phi_{b;k} + \Phi_\bullet) - \Phi_\bullet$.

Since $N_\lambda - N_{\lambda,\alpha}$ and $N_{\lambda,\alpha}$ are independent,

$$\begin{aligned} f_{\mathsf{E},b;k}(\lambda) = \mathbb{E}\,\varphi_{b;k}(\widetilde{\mathcal{T}}_\lambda) &= \sum_{\alpha\in\mathcal{A}} \mathbb{P}\bigl(N_\lambda > b,\, N_{\lambda,\alpha} = k\bigr) \\ &= \sum_{\alpha\in\mathcal{A}} \mathbb{P}\bigl(N_\lambda - N_{\lambda,\alpha} > b-k\bigr)\, \mathbb{P}\bigl(N_{\lambda,\alpha} = k\bigr) \\ &= \sum_{\alpha\in\mathcal{A}} \Bigl(1 - \sum_{i=0}^{b-k} \frac{(1-p_\alpha)^i \lambda^i}{i!} e^{-(1-p_\alpha)\lambda}\Bigr) \frac{p_\alpha^k \lambda^k}{k!} e^{-p_\alpha \lambda}. \end{aligned} \tag{4.99}$$

Hence,

$$\begin{aligned} f^*_{\mathsf{E},b;k}(s) &= \frac{1}{k!} \sum_{\alpha\in\mathcal{A}} p_\alpha^{-s} \Gamma(s+k) - \sum_{\alpha\in\mathcal{A}} \sum_{i=0}^{b-k} \frac{p_\alpha^k}{k!} \frac{(1-p_\alpha)^i}{i!} \Gamma(s+k+i) \\ &= \frac{1}{k!} \bigl(\rho(-s) - \rho(k)\bigr) \Gamma(s+k) - \sum_{i=1}^{b-k} \sum_{\alpha\in\mathcal{A}} p_\alpha^k (1-p_\alpha)^i \frac{\Gamma(s+k+i)}{k!\, i!}. \end{aligned} \tag{4.100}$$

In particular, for $k \ge 2$,

$$f^*_{\mathsf{E},b;k}(-1) = \frac{1-\rho(k)}{k(k-1)} - \sum_{i=1}^{b-k} \sum_{\alpha\in\mathcal{A}} p_\alpha^k (1-p_\alpha)^i \frac{(k+i-2)!}{k!\, i!}. \tag{4.101}$$

For $k=1$, we obtain by taking the limit as $s \to -1$, (4.101) with the first fraction (now undefined) replaced by $H$, cf. (3.51).

We leave calculations of $f^*_{\mathsf{V}}$ and (co)variances to the reader.

## 5. General central limit theorems

We state here several related general central limit theorems for additive functionals on tries; proofs are given in Section 6. As said in the introduction, the theorems use conditions on moments of the additive functionals and their toll functions; we will later obtain Theorem 3.9 as a special case of the results below by using Theorem 3.1 to verify these moment conditions.

In the statements of the theorems below, we use several functions $a(\lambda)$, $b(\lambda)$ and $c(\lambda)$, (with indices in the multivariate versions). This might seem frightening, but is intended to be friendly and flexible for applications; the meaning of these functions is as follows.

First, $a(\lambda)$ is an approximation of the mean $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$, and $b(\lambda)$ and $c(\lambda)$ are approximations of variances and covariances, see e.g. (5.1), (5.2), (5.11). We may choose $a(\lambda) := \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$, $b(\lambda) := \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)$, and $c(\lambda) := \operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr)$, and then (5.1), (5.2) and (5.11) are trivial, but in applications it is often preferable to use simpler approximations of the means and (co)variances, which is precisely what these functions are intended to be. Note that here the means and (co)variances are for the Poisson model, also in the theorems for the model with fixed $n$; this is both because of our proofs, and because in applications, the moments typically are easier to compute for the Poisson model. However, the mean for fixed $n$ is asymptotically the same as for the Poisson model, and the variances are related; see e.g. (5.12)–(5.14).

**Remark 5.1.** The conditions on these functions in the theorems below are asymptotic, as $\lambda \to \infty$. Hence the values of these functions for small $\lambda$ are irrelevant, and it is enough that they are defined for large $\lambda$. □

**Remark 5.2.** In the theorems below we assume that the assumptions hold for arbitrary real $\lambda$. (Or at least for sufficiently large $\lambda$, see Remark 5.1.) However, the results hold (by the same proofs) also if we consider only a given sequence $\lambda_n \to \infty$. □

In general, there there are oscillations in the variance. We therefore state many of the results as approximations (in distribution) using the notation $\overset{\mathrm{d}}{\approx}$ defined in Section 2.12. (This is especially important in the multivariate versions.) Note that we then include rather trivial cases when the normalized variable (e.g. the left-hand side of (5.6) or (5.7)) converges to 0 (in probability).

We begin with a general central limit theorem for the Poisson model.

**Theorem 5.3.** *Let $\varphi$ be a toll function and let $\Phi$ be the corresponding additive functional given by* (2.16)*. Let $a(\lambda)$ and $b(\lambda)$ be real-valued functions and suppose that for some $r > 2$, as $\lambda \to \infty$,*

$$\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = a(\lambda) + o\bigl(\sqrt{\lambda}\bigr) \tag{5.1}$$

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = b(\lambda) + o(\lambda), \tag{5.2}$$

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = O(\lambda), \tag{5.3}$$

$$\operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) = o(\lambda), \tag{5.4}$$

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)|^r = O\bigl(\lambda^{r/2}\bigr). \tag{5.5}$$

(i) *Then, as* $\lambda \to \infty$,

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - a(\lambda)}{\sqrt{\lambda}} \overset{\mathrm{d}}{\approx} N\bigl(0, b(\lambda)/\lambda\bigr) \tag{5.6}$$

*or, equivalently,*

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}} \overset{\mathrm{d}}{\approx} N\bigl(0, \operatorname{Var}[\Phi(\widetilde{\mathcal{T}}_\lambda)]/\lambda\bigr), \tag{5.7}$$

*in both cases with all [absolute] moments of order* $s < r$.

(ii) *Suppose further that*

$$b(\lambda) = \Omega(\lambda). \tag{5.8}$$

*Then, as* $\lambda \to \infty$,

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - a(\lambda)}{\sqrt{b(\lambda)}} \overset{\mathrm{d}}{\longrightarrow} N(0,1) \tag{5.9}$$

*and*

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)}} \overset{\mathrm{d}}{\longrightarrow} N(0,1), \tag{5.10}$$

*in both cases with convergence of all [absolute] moments of order* $s < r$.

**Remark 5.4.** We do not know whether (5.5) implies that also the $r$th moment converges in (5.9)–(5.10), and we leave this as an open problem. (The proof shows that this moment stays bounded, but this is not enough to imply convergence.) Nevertheless, the theorem shows that if (5.5) holds for all $r > 2$, then (5.6)–(5.7) hold with all [absolute] moments and that, if also (5.8) holds, then all [absolute] moments converge in (5.9)–(5.10). The same applies to the theorems below. □

We derive results for the model with fixed $n$ by conditioning. For this we assume that the functional $\Phi$ can be written as a difference between two increasing functionals with suitable conditions. (In particular, the theorem applies to increasing functionals $\Phi$.)

**Theorem 5.5.** *Let* $\varphi$ *be a toll function and let* $\Phi$ *be the corresponding additive functional given by* (2.16)*. Let* $b(\lambda)$ *be a real-valued function that satisfies* (5.2)*, and let* $c(\lambda)$ *be a function such that, as* $\lambda \to \infty$,

$$\operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) = c(\lambda) + o(\lambda). \tag{5.11}$$

*Suppose further that* $\varphi = \varphi_+ - \varphi_-$ *for some toll functions* $\varphi_\pm$ *such that the corresponding functionals* $\Phi_\pm$ *are increasing, and furthermore* (5.3), (5.4) *and* (5.5) *hold for* $\Phi_\pm$ *and* $\varphi_\pm$ *and some* $r > 2$.

(i) *Then, as* $n \to \infty$,

$$\mathbb{E}\,\Phi(\mathcal{T}_n) = \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_n) + o\big(\sqrt{n}\big) \tag{5.12}$$

$$\operatorname{Var}\Phi(\mathcal{T}_n) = b(n) - c(n)^2/n + o\big(n\big) \tag{5.13}$$

$$= \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_n) - \operatorname{Cov}\big(\Phi(\widetilde{\mathcal{T}}_n), N_n\big)^2/n + o\big(n\big), \tag{5.14}$$

*and*

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{n}} \overset{\mathrm{d}}{\approx} N\Big(0, \frac{\operatorname{Var}\Phi(\mathcal{T}_n)}{n}\Big) \overset{\mathrm{d}}{\approx} N\Big(0, \frac{b(n)}{n} - \frac{c(n)^2}{n^2}\Big), \tag{5.15}$$

*with all [absolute] moments of order* $s < r$.

(ii) *Suppose further that* $a(\lambda)$ *is a function satisfying* (5.1), *and that, as* $n \to \infty$,

$$b(n) - c(n)^2/n = \Omega\big(n\big). \tag{5.16}$$

*Then, as* $n \to \infty$,

$$\frac{\Phi(\mathcal{T}_n) - a(n)}{\sqrt{b(n) - c(n)^2/n}} \overset{\mathrm{d}}{\longrightarrow} N(0,1) \tag{5.17}$$

*and, equivalently,*

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\mathcal{T}_n)}{\sqrt{\operatorname{Var}\Phi(\mathcal{T}_n)}} \overset{\mathrm{d}}{\longrightarrow} N(0,1), \tag{5.18}$$

*in both cases with convergence of all [absolute] moments of order* $s < r$.

These theorems are easily extended to multivariate versions. This can essentially be done by the standard Cramér–Wold device, with a (minor) technical complication because of the possibility of oscillations in the covariance matrix, and thus no straightforward limit distribution. We begin with a multivariate extension of Theorem 5.3. For later convenience, we give two equivalent versions of this extension, using functions $a_k$ and $b_{k\ell}$ as discussed above in Corollary 5.7 but not in Theorem 5.6.

**Theorem 5.6.** *Let* $\varphi_1, \dots, \varphi_K$ *be toll functions, for some* $K \geq 1$, *let* $\Phi_k$ *be the corresponding additive functionals given by* (2.16), *and assume that, as* $\lambda \to \infty$, (5.3), (5.4) *and* (5.5) *hold for each* $\varphi_k$ *and some* $r > 2$. *Then, as* $\lambda \to \infty$,

$$\Big(\frac{\Phi_k(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}}\Big)_{k=1}^K \overset{\mathrm{d}}{\approx} N\big(0, \Sigma(\lambda)\big), \tag{5.19}$$

*where the covariance matrix* $\Sigma(\lambda) = \big(\sigma_{k\ell}(\lambda)\big)_{k,\ell=1}^K$ *is given by*

$$\sigma_{k\ell}(\lambda) := \frac{\operatorname{Cov}\big(\Phi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_\ell(\widetilde{\mathcal{T}}_\lambda)\big)}{\lambda}. \tag{5.20}$$

*Furthermore,* (5.19) *holds with all [absolute] moments of order* $s < r$.

**Corollary 5.7.** *Suppose in addition to the assumptions of Theorem 5.6 that $a_k(\lambda)$ and $b_{k\ell}(\lambda)$, for $k,\ell = 1,\dots,K$, are real-valued functions such that, as $\lambda \to \infty$,* (5.1) *holds for each $\Phi_k$ (with $a_k(\lambda)$), and* (5.2) *holds in the form*

$$\operatorname{Cov}\bigl(\Phi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_\ell(\widetilde{\mathcal{T}}_\lambda)\bigr) = b_{k\ell}(\lambda) + o(\lambda). \tag{5.21}$$

*Then, as $\lambda \to \infty$,*

$$\Bigl(\frac{\Phi_k(\widetilde{\mathcal{T}}_\lambda) - a_k(\lambda)}{\sqrt{\lambda}}\Bigr)_{k=1}^K \overset{\mathrm{d}}{\approx} N\bigl(0, \Sigma(\lambda)\bigr), \tag{5.22}$$

*where the covariance matrix $\Sigma(\lambda) = \bigl(\sigma_{k\ell}(\lambda)\bigr)_{k,\ell=1}^K$ is given by*

$$\sigma_{k\ell}(\lambda) := \frac{b_{k\ell}(\lambda)}{\lambda}. \tag{5.23}$$

*Furthermore,* (5.22) *holds with all [absolute] moments of order $s < r$.*

We state also a corresponding multivariate extension of Theorem 5.5 for the model with fixed $n$.

**Theorem 5.8.** *Let $\varphi_1,\dots,\varphi_K$ be toll functions, for some $K \ge 1$, let $\Phi_k$ be the corresponding additive functionals given by* (2.16)*, and let $a_k(\lambda)$, $b_{k\ell}(\lambda)$ and $c_k(\lambda)$ be real-valued functions such that* (5.1) *and* (5.11) *hold for each $\Phi_k$ (with $a_k(\lambda)$ and $c_k(\lambda)$), and* (5.21) *holds.*

*Suppose further that each $\varphi_k = \varphi_{k+} - \varphi_{k-}$ for some toll functions $\varphi_{k\pm}$ such that the corresponding functionals $\Phi_{k\pm}$ are increasing, and furthermore* (5.3)*,* (5.4) *and* (5.5) *hold for $\Phi_{k\pm}$ and $\varphi_{k\pm}$ and some $r > 2$,*

*Then, as $n \to \infty$,*

$$\Bigl(\frac{\Phi_k(\mathcal{T}_n) - a_k(n)}{\sqrt{n}}\Bigr)_{k=1}^K \overset{\mathrm{d}}{\approx} N\bigl(0, \widehat{\Sigma}(n)\bigr), \tag{5.24}$$

*where the covariance matrix $\widehat{\Sigma}(n) = \bigl(\widehat{\sigma}_{k\ell}(n)\bigr)_{k,\ell=1}^K$ is given by*

$$\widehat{\sigma}_{k\ell}(n) := \frac{b_{k\ell}(n) - c_k(n)c_\ell(n)/n}{n} = \frac{b_{k\ell}(n)}{n} - \frac{c_k(n)}{n}\frac{c_\ell(n)}{n}. \tag{5.25}$$

*Moreover,* (5.24) *holds with all [absolute] moments of order $s < r$; in particular,*

$$\mathbb{E}\,\Phi_k(\mathcal{T}_n) = a_k(n) + o\bigl(\sqrt{n}\bigr), \tag{5.26}$$

$$\operatorname{Cov}\bigl(\Phi_k(\mathcal{T}_n), \Phi_\ell(\mathcal{T}_n)\bigr) = b_{k\ell}(n) - \frac{c_k(n)c_\ell(n)}{n} + o\bigl(n\bigr). \tag{5.27}$$

**Remark 5.9.** In (5.24), we may replace $a_k(n)$ by either $\mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_n)$ (since we may choose $a_k(n) := \mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_n)$), or by $\mathbb{E}\,\Phi_k(\mathcal{T}_n)$ (by (5.12)). In these cases (5.1) holds automatically and does not have to be verified. □

## 6. Proofs of general central limit theorems

We first note that if $\varphi$ is a functional of tries and either $\varphi \geq 0$ or $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)| < \infty$, then, since $N_\lambda \sim \mathrm{Po}(\lambda)$, and $\varphi(\emptyset) = 0$,

$$\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) = e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} a_n, \tag{6.1}$$

with $a_n := \mathbb{E}\,\varphi(\mathcal{T}_n)$.

**Lemma 6.1.** *Let $0 \leq \lambda_1 \leq \lambda_2$.*

(i) *If $\varphi : \mathfrak{T} \to \mathbb{R}$ is an arbitrary functional, then*

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda_1})| \leq e^{\lambda_2}\,\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda_2})|. \tag{6.2}$$

(ii) *Moreover, if $m$ is such that $\varphi(T) = 0$ when $|T|_{\mathsf{e}} < m$, then*

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda_1})| \leq \Bigl(\frac{\lambda_1}{\lambda_2}\Bigr)^m e^{\lambda_2}\,\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda_2})|. \tag{6.3}$$

*Proof.* By (6.1) applied to $|\varphi|$,

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)| = e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} a_n, \tag{6.4}$$

where $a_n := \mathbb{E}\,|\varphi(\mathcal{T}_n)| \geq 0$, and both (6.2) and (6.3) follow. (The latter because $a_n = 0$ for $n < m$.) □

**Lemma 6.2.** *Let $\varphi$ be a toll function and let $\Phi$ be the corresponding additive functional given by (2.16). Let $r \geq 1$ and assume that $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$ for some $\lambda > 0$. Then $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$.*

*Proof.* We consider first three special cases.

*Case 1: $\varphi(T) = 0$ unless $|T|_{\mathsf{e}} = 1$.* Then, using Example 2.1, if $a := \varphi(\bullet)$, we have $\varphi = a\varphi_\bullet$, and $\Phi(\widetilde{\mathcal{T}}_\lambda) = a|\widetilde{\mathcal{T}}_\lambda|_{\mathsf{e}} = aN_\lambda$. Hence, $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$ for every $r < \infty$.

*Case 2: There exists $m \geq 2$ such that $\varphi(T) = 0$ unless $|T|_{\mathsf{e}} = m$.* Consider first the random trie $\mathcal{T}_m$ constructed from $m$ strings $\Xi^{(1)}, \dots, \Xi^{(m)}$. Note that $\mathcal{T}_m \overset{\mathrm{d}}{=} (\widetilde{\mathcal{T}}_\lambda \mid N_\lambda = m)$ and thus

$$\mathbb{E}\,|\varphi(\mathcal{T}_m)|^r = \mathbb{E}\bigl(|\varphi(\widetilde{\mathcal{T}}_\lambda)|^r \mid N_\lambda = m\bigr) \leq \mathbb{P}(N_\lambda = m)^{-1}\,\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty. \tag{6.5}$$

Let $\boldsymbol{\alpha} \in \mathcal{A}^*$ and consider the fringe tree $\mathcal{T}_m^{\boldsymbol{\alpha}}$. If not all $m$ strings $\Xi^{(j)}$ have the prefix $\boldsymbol{\alpha}$, then this fringe tree has less than $m$ leaves, and thus, by our assumption, $\varphi(\mathcal{T}_m^{\boldsymbol{\alpha}}) = 0$. Furthermore, if we condition on the opposite event, i.e., that all $m$ strings have prefix $\boldsymbol{\alpha}$, then the fringe tree $\mathcal{T}_m^{\boldsymbol{\alpha}}$ has the same distribution as the unconditioned $\mathcal{T}_m$. Hence,

$$\mathbb{E}\,|\varphi(\mathcal{T}_m^{\boldsymbol{\alpha}})|^r = \mathbb{P}\bigl(\Xi \succ \boldsymbol{\alpha}\bigr)^m\,\mathbb{E}\,|\varphi(\mathcal{T}_m)|^r = C\,\mathbb{P}\bigl(\Xi \succ \boldsymbol{\alpha}\bigr)^m = CP(\boldsymbol{\alpha})^m. \tag{6.6}$$

Moreover, for every $\ell \geq 0$, there exists at most one $\boldsymbol{\alpha} \in \mathcal{A}^\ell$ such that $\varphi(\mathcal{T}_m^{\boldsymbol{\alpha}}) \neq 0$. Hence, if we let

$$X_\ell := \sum_{|\boldsymbol{\alpha}|=\ell} \varphi(\mathcal{T}_m^{\boldsymbol{\alpha}}), \tag{6.7}$$

then, by (6.6) and (2.5), where by (2.6) $\rho(m) := \sum_{\alpha\in\mathcal{A}} p_\alpha^m < 1$,

$$\mathbb{E}\,|X_\ell|^r = \mathbb{E}\sum_{|\boldsymbol{\alpha}|=\ell} |\varphi(\mathcal{T}_m^{\boldsymbol{\alpha}})|^r = C\sum_{|\boldsymbol{\alpha}|=\ell} P(\alpha)^m = C\rho(m)^\ell. \tag{6.8}$$

Thus $\|X_\ell\|_r \leq C\rho(m)^{\ell/r}$. Hence, (2.17) and Minkowski's inequality yield

$$\|\Phi(\mathcal{T}_m)\|_r = \Big\|\sum_{\ell\geq 0} X_\ell\Big\|_r \leq \sum_{\ell\geq 0} C\rho(m)^{\ell/r} < \infty. \tag{6.9}$$

Now return to the random trie $\widetilde{\mathcal{T}}_\lambda$ in the Poisson model. Consider the bucket trie with bucket size $m$, based on the same strings. As said in Section 2.6, the trie $\widetilde{\mathcal{T}}_\lambda$ is obtained from the bucket trie by letting a small trie grow from each bucket. By our assumption, the only non-zero contributions to $\Phi(\widetilde{\mathcal{T}}_\lambda)$ in (2.16) then comes from the small tries grown from the buckets that contain exactly $m$ strings. Condition on the bucket trie, and let $M$ be the number of buckets with $m$ strings. Then the small tries grown from them are $M$ independent copies of $\mathcal{T}_m$. Hence, if $W_1, W_2, \dots,$ are i.i.d. copies of $\Phi(\mathcal{T}_m)$, we have

$$\big(\Phi(\widetilde{\mathcal{T}}_\lambda) \mid M\big) \overset{\mathrm{d}}{=} \sum_{j=1}^M W_j. \tag{6.10}$$

Consequently, by Minkowski's inequality and (6.9),

$$\big\|\big(\Phi(\widetilde{\mathcal{T}}_\lambda) \mid M\big)\big\|_r = \Big\|\sum_{j=1}^M W_j\Big\|_r \leq M\big\|W_1\big\|_r = M\big\|\Phi(\mathcal{T}_m)\big\|_r = CM. \tag{6.11}$$

Furthermore, since the sets of strings in the buckets are disjoint, $M \leq N_\lambda/m \leq N_\lambda$. Consequently,

$$\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r = \mathbb{E}\big[\mathbb{E}\big(|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r \mid M\big)\big] \leq \mathbb{E}(CM^r) \leq C\,\mathbb{E}\,N_\lambda^r < \infty. \tag{6.12}$$

*Case 3: $\varphi(T) = 0$ if $|T|_{\mathsf{e}} \leq r$.* Then, in particular, $\varphi(\bullet) = 0$.

Let $m := \lfloor r\rfloor + 1$. Then (2.25) and Lemma 6.1(ii) (applied to $|\varphi(T)|^r$) yield, for any $\boldsymbol{\alpha} \in \mathcal{A}^*$,

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})|^r = \mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})|^r \leq P(\boldsymbol{\alpha})^m e^\lambda\,\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^r. \tag{6.13}$$

Hence, for some $C_\lambda < \infty$,

$$\|\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\|_r \leq C_\lambda P(\boldsymbol{\alpha})^{m/r}. \tag{6.14}$$

Consequently, (2.17), Minkowski's inequality and (2.7) yield, since $m/r > 1$,

$$\|\Phi(\widetilde{\mathcal{T}}_\lambda)\|_r \leq \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \|\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\|_r \leq \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} C_\lambda P(\boldsymbol{\alpha})^{m/r} < \infty, \tag{6.15}$$

and thus $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$.

*Case 4: The general case.* Decompose

$$\varphi = \sum_{1\le j\le \lfloor r\rfloor} \varphi_j + \varphi', \tag{6.16}$$

where $\varphi_j(T) := \varphi(T)\mathbf{1}\{|T|_{\mathsf{e}} = j\}$ and $\varphi'(T) := \varphi(T)\mathbf{1}\{|T|_{\mathsf{e}} > r\}$. Then Case 1 applies to $\varphi_1$, Case 2 to $\varphi_j$ for $2 \le j \le \lfloor r\rfloor$, and Case 3 to $\varphi'$. Consequently, the corresponding additive functionals $\Phi_j$ and $\Phi'$ satisfy $\mathbb{E}\,|\Phi_j(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$ and $\mathbb{E}\,|\Phi'(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$, and the result follows by Minkowski's inequality since $\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{j=1}^m \Phi_j(\widetilde{\mathcal{T}}_\lambda) + \Phi'(\widetilde{\mathcal{T}}_\lambda)$. □

**Lemma 6.3.** *Let $\varphi$ be a toll function and let $\Phi$ be the corresponding additive functional given by* (2.16)*. Let $r > 2$ and assume that, as $\lambda \to \infty$,*

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = O(\lambda), \tag{6.17}$$

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)|^r = O\big(\lambda^{r/2}\big). \tag{6.18}$$

*Then, $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$ for all $\lambda \ge 0$ and*

$$\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)|^r = O\big(\lambda^{r/2}\big), \qquad \lambda \ge 1. \tag{6.19}$$

*Proof.* Note first that in the special case $\varphi(T) = \varphi_\bullet(T) := \mathbf{1}\{T = \bullet\}$ in Example 2.1, $\Phi(\widetilde{\mathcal{T}}_\lambda) = N_\lambda \sim \operatorname{Po}(\lambda)$, and (6.17)–(6.19) hold; for (6.19), this is because as $\lambda \to \infty$, $(N_\lambda - \lambda)/\lambda^{1/2} \overset{\mathrm{d}}{\longrightarrow} N(0,1)$ with all absolute moments. (This follows e.g. first for integer $\lambda$ from [11, Theorem 7.5.1], and then in general using Minkowski's inequality.) Hence, by subtracting a suitable multiple of $\varphi_\bullet$ from $\varphi$, and using Minkowski's inequality for each of (6.17)–(6.19), we may in the remainder of the proof assume that $\varphi(\bullet) = 0$. Then also $\Phi(\bullet) = 0$.

By (2.18) and (2.25) (for $\Phi$, using $\Phi(\bullet) = 0$), we have the decomposition

$$\Phi(\widetilde{\mathcal{T}}_\lambda) = \varphi(\widetilde{\mathcal{T}}_\lambda) + \sum_{\alpha\in\mathcal{A}} \Phi(\widetilde{\mathcal{T}}_\lambda^\alpha) = \varphi(\widetilde{\mathcal{T}}_\lambda) + \sum_{\alpha\in\mathcal{A}} \Phi(\widetilde{\mathcal{T}}_\lambda^{\alpha+}). \tag{6.20}$$

Define, for $\alpha \in \mathcal{A}$,

$$X_{\lambda,\alpha} := \Phi\big(\widetilde{\mathcal{T}}_\lambda^{\alpha+}\big) - \mathbb{E}\,\Phi\big(\widetilde{\mathcal{T}}_\lambda^{\alpha+}\big). \tag{6.21}$$

Then, by (6.20),

$$\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) + \sum_{\alpha\in\mathcal{A}} X_{\lambda,\alpha}. \tag{6.22}$$

In the Poisson model, the different modified branches $\widetilde{\mathcal{T}}_\lambda^{\alpha+}$, $\alpha \in \mathcal{A}$, are independent random tries, and thus the random variables $X_{\lambda,\alpha}$, $\alpha \in \mathcal{A}$, are independent. Furthermore, $\mathbb{E}\,X_{\lambda,\alpha} = 0$ by (6.21). Hence, we may apply the version of Rosenthal's inequality in Lemma 6.4 below, and conclude that, if we fix any $K > 1$ (this will be chosen later), there exists $C_1 = C_1(r, K)$ such that

$$\mathbb{E}\Bigl|\sum_{\alpha\in\mathcal{A}} X_{\lambda,\alpha}\Bigr|^r \le K \sum_{\alpha\in\mathcal{A}} \mathbb{E}\,|X_{\lambda,\alpha}|^r + C_1\Bigl(\sum_{\alpha\in\mathcal{A}} \mathbb{E}\,X_{\lambda,\alpha}^2\Bigr)^{r/2}. \tag{6.23}$$

Let

$$g(\lambda) := \|\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)\|_r = \bigl(\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)|^r\bigr)^{1/r}. \tag{6.24}$$

Since $\Phi(\widetilde{\mathcal{T}}_\lambda^{\alpha+}) \overset{\mathrm{d}}{=} \Phi(\widetilde{\mathcal{T}}_{p_\alpha\lambda})$ by (2.25), recalling that $P(\alpha) = p_\alpha$ for $\alpha \in \mathcal{A}$, it follows from (6.21) that

$$\mathbb{E}\,|X_{\lambda,\alpha}|^r = \mathbb{E}\bigl|\Phi(\widetilde{\mathcal{T}}_{p_\alpha\lambda}) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_{p_\alpha\lambda})\bigr|^r = g(p_\alpha\lambda)^r. \tag{6.25}$$

By (6.17) and (6.18), there exists $\lambda_0 \ge 1$ such that for all $\lambda \ge \lambda_0$, and all $\alpha \in \mathcal{A}$,

$$\mathbb{E}\,X_{\lambda,\alpha}^2 = \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda^{\alpha+}) = \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_{p_\alpha\lambda}) \le C_2 p_\alpha\lambda \le C_2\lambda, \tag{6.26}$$

$$\|\varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)\|_r \le C_3\lambda^{1/2}. \tag{6.27}$$

Hence, by (6.24), (6.22), Minkowski's inequality, (6.23), (6.25), and (6.26)–(6.27), for $\lambda \ge \lambda_0$,

$$\begin{aligned} g(\lambda) &\le \Bigl\|\sum_{\alpha\in\mathcal{A}} X_{\lambda,\alpha}\Bigr\|_r + \bigl\|\varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)\bigr\|_r \\ &\le \Bigl(K\sum_{\alpha\in\mathcal{A}} g\bigl(p_\alpha\lambda\bigr)^r + C_4\lambda^{r/2}\Bigr)^{1/r} + C_3\lambda^{1/2} \\ &\le \Bigl(K\sum_{\alpha\in\mathcal{A}} g\bigl(p_\alpha\lambda\bigr)^r\Bigr)^{1/r} + C_5\lambda^{1/2}. \end{aligned} \tag{6.28}$$

Let $\lambda_1 := \lambda_0/p_{\min}$, and $\lambda_k := \lambda_1/p_{\max}^{k-1}$ for $k \ge 2$. We show by induction on $k$ that for some large $A < \infty$,

$$g(\lambda) \le A\lambda^{1/2}, \qquad \lambda \in [1, \lambda_k]. \tag{6.29}$$

First, $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda_1})|^r < \infty$ by (6.27), and thus Lemma 6.2 yields $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_{\lambda_1})|^r < \infty$. Hence, by Lemma 6.1(i),

$$g(\lambda) \le 2\|\Phi(\widetilde{\mathcal{T}}_\lambda)\|_r \le C_6\|\Phi(\widetilde{\mathcal{T}}_{\lambda_1})\|_r \le C_7, \qquad \lambda \le \lambda_1. \tag{6.30}$$

Thus (6.29) holds in the base case $k = 1$ if $A \ge C_7$.

For the induction step, assume (6.29). It suffices to consider $\lambda \in (\lambda_k, \lambda_{k+1}]$, and then $p_\alpha\lambda \in [\lambda_0, \lambda_k]$ for every $\alpha \in \mathcal{A}$. Hence, (6.28) and the induction hypothesis (6.29) yield, recalling (2.4),

$$\begin{aligned} g(\lambda) &\le \Bigl(K\sum_{\alpha\in\mathcal{A}} A^r(p_\alpha\lambda)^{r/2}\Bigr)^{1/r} + C_5\lambda^{1/2} \\ &= \Bigl(\Bigl(K\sum_{\alpha\in\mathcal{A}} p_\alpha^{r/2}\Bigr)^{1/r} + C_5A^{-1}\Bigr)A\lambda^{1/2} \end{aligned}$$

$$= \Big( \big(K\rho(r/2)\big)^{1/r} + C_5 A^{-1} \Big) A\lambda^{1/2}. \tag{6.31}$$

By (2.6), $\rho(r/2) < 1$. We now assume that $K$ was chosen such that $1 < K < \rho(r/2)^{-1}$. Then $K\rho(r/2) < 1$, and we may choose $A$ so large that

$$\big(K\rho(r/2)\big)^{1/r} + C_5 A^{-1} \le 1. \tag{6.32}$$

Then (6.31) shows that (6.29) holds also for $\lambda \in (\lambda_k, \lambda_{k+1}]$, showing the induction step.

We have shown that (6.29) hold for every $k \ge 1$, and thus $g(\lambda) \le A\lambda^{1/2}$ for all $\lambda \ge 1$, which by the definition (6.24) is the same as (6.19). We have also shown in (6.30) that $g(\lambda) = O(1)$ for $\lambda \le 1$. Hence, $\mathbb{E}\,|\Phi(\widetilde{\mathcal{T}}_\lambda)|^r < \infty$ for every $\lambda$. □

The proof above used the following version of Rosenthal's inequality. The standard version of Rosenthal's inequality, see e.g. [11, Theorem 3.9.1], is (6.33) with $K = C = C(r)$ (growing with $r$); the fact needed here that one can choose $K$ arbitrarily close to 1 (at the expense of increasing $C$) is due to Pinelis [28], see also [29] for a sharper result.

**Lemma 6.4** (Rosenthal, Pinelis [28])**.** *For every $r > 2$ and every $K > 1$, there exists a constant $C = C(r, K)$ such that for any independent random variables $X_1, \dots, X_n$ with means $\mathbb{E}\,X_i = 0$,*

$$\mathbb{E}\Big|\sum_{i=1}^n X_i\Big|^r \le K \sum_{i=1}^n \mathbb{E}\,|X_i|^r + C\Big(\sum_{i=1}^n \mathbb{E}\,|X_i|^2\Big)^{r/2}. \tag{6.33}$$

**Remark 6.5.** Note that in the special case $r = 4$, a simple calculation shows (6.33) directly, with $K = 1$ and $C = 3$. Similarly, when $r$ is any even integer, (6.33) (with any $K > 1$) is easily shown by elementary calculations and Hölder's inequality. (These cases suffice for most applications of our theorems.) □

*Proof.* Pinelis [28, Corollary and (4)] (with $t = r$) yields the inequality (6.33) with $K$ replaced by

$$\sum_{j=0}^{\lfloor r/2\rfloor-1} c_j(r) a_j^{-2j}/j! = c_0(r) + \sum_{j=1}^{\lfloor r/2\rfloor-1} c_j(r) a_j^{-2j}/j! \tag{6.34}$$

and $C$ given by an explicit formula (involving $c_j(t)$ and $a_j$) that we ignore; here $a_j > 0$ are arbitrary and $c_j(r)$ are some numbers defined from some other numbers $p(s)$ and $q(s)$, $s \in [2, r]$, that can be chosen freely under the conditions $p(s) \ge 1$, $q(s) \ge 1$ and, when $s > 3$,

$$p(s)^{-1/(s-3)} + q(s)^{-1/(s-3)} \le 1; \tag{6.35}$$

in particular, $c_0(r) = q(r)$. (See [28] for further details.)

Given $K > 0$ we can choose first $q(r)$ with $1 < q(r) < K$ and then $p(r) \ge 1$ so large that if $r > 3$, (6.35) holds for $s = r$. We choose also, for example, $p(s) = q(s) = \max\{1, 2^{s-3}\}$ for $s \in [2, r)$. This defines the numbers $c_j(r)$

for $j = 0, \dots, \lfloor r/2 \rfloor - 1$ with $c_0(r) = q(r) < K$, and we then can choose $a_j$, $j \geq 1$, so large that the sum in (6.34) is $\leq K$. ☐

We may now prove Theorem 5.3, the general central limit theorem for the Poisson model.

*Proof of Theorem 5.3.* All limits and asymptotic notions below are as $\lambda \to \infty$.

Let $m \geq 1$. Using the decomposition (2.18) recursively $m$ times on the tree $\widetilde{\mathcal{T}}_\lambda$, we obtain,

$$\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{|\boldsymbol{\alpha}|<m} \varphi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}\bigr) + \sum_{|\boldsymbol{\alpha}|=m} \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}\bigr) \tag{6.36}$$

$$= \sum_{|\boldsymbol{\alpha}|<m} \varphi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}\bigr) + \sum_{|\boldsymbol{\alpha}|=m} \bigl(\Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}\bigr) - \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr)\bigr) + \sum_{|\boldsymbol{\alpha}|=m} \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr) \tag{6.37}$$

$$=: R'_m + R''_m + \sum_{|\boldsymbol{\alpha}|=m} \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr), \tag{6.38}$$

defining $R'_m$ and $R''_m$ as the first two sums in (6.37). By (2.23), $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+} \overset{\mathrm{d}}{=} \widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}$, and thus (5.4) implies that for every fixed $\boldsymbol{\alpha} \in \mathcal{A}^*$,

$$\operatorname{Var} \varphi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr) = \operatorname{Var} \varphi\bigl(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}\bigr) = o\bigl(\lambda P(\boldsymbol{\alpha})\bigr) = o(\lambda). \tag{6.39}$$

By (2.24) and Minkowski's inequality, this implies

$$\bigl(\operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\bigr)^{1/2} \leq \bigl(\operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+})\bigr)^{1/2} + O(1) = o\bigl(\lambda^{1/2}\bigr). \tag{6.40}$$

Hence, Minkowski's inequality again yields, for any fixed $m$,

$$\bigl(\operatorname{Var} R'_m\bigr)^{1/2} \leq \sum_{|\boldsymbol{\alpha}|<m} \bigl(\operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\bigr)^{1/2} = \sum_{|\boldsymbol{\alpha}|<m} o\bigl(\lambda^{1/2}\bigr) = o\bigl(\lambda^{1/2}\bigr). \tag{6.41}$$

Similarly, by (2.24) (applied to $\Phi$), $\Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}\bigr) - \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr) = O(1)$, and thus, still for fixed $m$, $R''_m = O(1)$ and thus $\operatorname{Var} R''_m = O(1)$. Hence, defining $R_m := R'_m + R''_m$,

$$\operatorname{Var} R_m \leq 2 \operatorname{Var} R'_m + 2 \operatorname{Var} R''_m = o(\lambda). \tag{6.42}$$

Consequently, $R_m$ is negligible, and the major term in (6.38) is the last sum. We subtract the expectations, and obtain from (6.38)

$$\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = R_m - \mathbb{E}\,R_m + \sum_{|\boldsymbol{\alpha}|=m} X_{\lambda,\boldsymbol{\alpha}}, \tag{6.43}$$

where

$$X_{\lambda,\boldsymbol{\alpha}} := \Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr) - \mathbb{E}\,\Phi\bigl(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}\bigr). \tag{6.44}$$

Lemma 6.3 applies, since (6.17) and (6.18) are our assumptions (5.3) and (5.5); thus, for $\lambda \geq 1$,

$$\mathbb{E}\bigl|\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)\bigr|^r \leq C_1 \lambda^{r/2}. \tag{6.45}$$

Hence, using again (2.23), for any $m \geq 0$ and all $\boldsymbol{\alpha} \in \mathcal{A}^m$, at least for $\lambda \geq p_{\min}^{-m}$,

$$\mathbb{E}\,|X_{\lambda,\boldsymbol{\alpha}}|^r = \mathbb{E}\bigl|\Phi\bigl(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}\bigr) - \mathbb{E}\,\Phi\bigl(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}\bigr)\bigr|^r \leq C_1(\lambda P(\boldsymbol{\alpha}))^{r/2}, \tag{6.46}$$

where the constant $C_1$ does not depend on $m$.

The random modified fringe trees $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ for $|\boldsymbol{\alpha}| = m$ are independent; hence the random variables $X_{\lambda,\boldsymbol{\alpha}}$ in (6.43) are independent. Furthermore, by the definition (6.44), $\mathbb{E}\,X_{\lambda,\boldsymbol{\alpha}} = 0$. Moreover, (6.46) and (2.5) imply that for $\lambda \geq p_{\min}^{-m}$,

$$\sum_{|\boldsymbol{\alpha}|=m} \frac{\mathbb{E}\,|X_{\lambda,\boldsymbol{\alpha}}|^r}{\lambda^{r/2}} \leq \sum_{|\boldsymbol{\alpha}|=m} C_1 P(\boldsymbol{\alpha})^{r/2} = C_1\rho(r/2)^m. \tag{6.47}$$

We have so far kept $m$ fixed, and shown that (6.47) holds for large $\lambda$, and also that (6.42) holds, and thus, for example, for large $\lambda$,

$$\operatorname{Var} R_m \leq \frac{1}{m}\lambda. \tag{6.48}$$

In other words, there exist $\lambda(m) < \infty$ such that (6.47) and (6.48) hold for $\lambda \geq \lambda(m)$. We may also assume $\lambda(m+1) \geq \lambda(m)+1$. Now define (for large $\lambda$) $m(\lambda) := \max\{m : \lambda(m) \leq \lambda\}$, and take in the remainder of the proof $m := m(\lambda)$. Then, $m = m(\lambda) \to \infty$ as $\lambda \to \infty$, and by definition, (6.47) and (6.48) hold with $m = m(\lambda)$. Since $m \to \infty$ and $\rho(r/2) < 1$, see (2.6), we have $\rho(r/2)^m \to 0$, and thus (6.47) shows that

$$\sum_{|\boldsymbol{\alpha}|=m} \frac{\mathbb{E}\,|X_{\lambda,\boldsymbol{\alpha}}|^r}{\lambda^{r/2}} \to 0. \tag{6.49}$$

Furthermore, (6.48) implies

$$\operatorname{Var} R_m = o(\lambda). \tag{6.50}$$

By the subsequence principle in Remark 2.5, it suffices to show that for any given sequence $\lambda_n \to \infty$, the results hold for some subsequence. By (5.3), $\operatorname{Var}(\Phi(\widetilde{\mathcal{T}}_\lambda))/\lambda = O(1)$, and thus we may, by selecting a subsequence $(\lambda'_n)$ of the given sequence $(\lambda_n)$, assume that $\operatorname{Var}(\Phi(\widetilde{\mathcal{T}}_\lambda))/\lambda \to \gamma$ for some $\gamma \geq 0$, and thus also, by (5.2), $b(\lambda)/\lambda \to \gamma$.

If $\gamma = 0$, i.e., $\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = o(\lambda)$ along the subsequence, then the left-hand side of (5.7) tends to 0 in probability, and (5.7) holds trivially (along the subsequence). The same holds for (5.6) by (5.1)–(5.2). (Note also that $\gamma = 0$ is impossible in (ii) since we there assume (5.8).)

Now suppose $\gamma > 0$, and consider only the selected subsequence $(\lambda'_n)$. Then (5.8) holds, and thus (6.50) yields

$$\operatorname{Var} R_m = o(\lambda) = o\bigl(b(\lambda)\bigr). \tag{6.51}$$

Hence, (5.2) and Minkowski's inequality yield

$$\operatorname{Var}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda) - R_m\bigr) = b(\lambda) + o(\lambda) = b(\lambda) + o\bigl(b(\lambda)\bigr). \tag{6.52}$$

Now use the decomposition (6.43) with $m = m(\lambda)$. Note that (6.43) and (6.52) yield

$$\sum_{|\boldsymbol{\alpha}|=m} \operatorname{Var}\bigl(X_{\lambda,\boldsymbol{\alpha}}\bigr) = \operatorname{Var}\Bigl(\sum_{|\boldsymbol{\alpha}|=m} X_{\lambda,\boldsymbol{\alpha}}\Bigr) = \operatorname{Var}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda) - R_m\bigr) \sim b(\lambda). \quad (6.53)$$

Hence, the central limit theorem (see e.g. [11, Theorem 7.2.4] or [21, Theorem 5.12]) applies to the sum $\sum_{|\boldsymbol{\alpha}|=m} X_{\lambda,\boldsymbol{\alpha}}/b(\lambda)^{1/2}$, with Lyapounov's condition (as in [11, Theorem 7.2.2]) verified by (6.49) and (5.8). Consequently,

$$b(\lambda)^{-1/2} \sum_{|\boldsymbol{\alpha}|=m} X_{\lambda,\boldsymbol{\alpha}} \overset{\mathrm{d}}{\longrightarrow} N(0,1). \quad (6.54)$$

Furthermore, (6.51) implies $b(\lambda)^{-1/2}(R_m - \mathbb{E}\, R_m) \overset{\mathrm{p}}{\longrightarrow} 0$. Thus (6.43), (6.54) and the Cramér–Slutsky theorem [11, Theorem 5.11.4] yield

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{b(\lambda)^{1/2}} \overset{\mathrm{d}}{\longrightarrow} N(0,1). \quad (6.55)$$

The conclusions (5.9) and (5.10) (along the subsequence) now follow using (5.1), (5.2), and (5.8). Moreover, by multiplying (5.9) by $\sqrt{b(\lambda)/\lambda} \to \gamma^{1/2}$, it follows that the left-hand side of (5.6) converges in distribution to $N(0,\gamma)$, which yields (5.6), see Remark 2.4. Similarly, (5.7) holds.

Combining the two cases above, we have shown, for any $\gamma \geq 0$, that (5.6)–(5.7) and (5.9)–(5.10) (assuming (5.8)) hold along the subsequence $(\lambda'_n)$, Since we started with an arbitrary subsequence $(\lambda_n)$, they hold for arbitrary $\lambda \to \infty$, see Remark 2.5. This proves (5.6)–(5.7) and (5.9)–(5.10).

It remains only to show that these hold with moments as stated. If (5.8) holds, then (6.45) implies, recalling also (5.1) and (5.2), that the $r$th absolute moments of the left-hand sides of (5.9) and (5.10) are bounded as $\lambda \to \infty$, which as is well-known implies that every power of lower order is uniformly integrable, and thus every [absolute] moment of lower order converges to the corresponding moment of $N(0,1)$. (See e.g. [11, Theorems 5.4.2 and 5.5.9].)

Similarly, if we write (5.6) or (5.7) as $X_\lambda \approx Y_\lambda$, then (6.45) implies that $\mathbb{E}\,|X_\lambda|^r = O(1)$ for $\lambda \geq 1$, and thus if $0 < s < r$, then the variables $|X_\lambda|^s$, $\lambda \geq 1$, are uniformly integrable. The same holds for $|Y_\lambda|^s$ (at least for large $\lambda$), since $Y_\lambda$ is normal with $\operatorname{Var} Y_\lambda = O(1)$ as $\lambda \to \infty$. Hence, Lemma 2.6 applies to $X_{\lambda_n}$ and $Y_{\lambda_n}$ for any sequence $\lambda_n \to \infty$, and it follows that $X_\lambda \overset{\mathrm{d}}{\approx} Y_\lambda$ with [absolute] moments of order $s$. □

We next prove the multivariate extensions of Theorem 5.3.

*Proof of Theorem 5.6.* By the subsequence principle in Remark 2.5, it suffices to show that for any given sequence $\lambda_n \to \infty$, the result holds for some subsequence.

The Cauchy–Schwarz inequality and (5.3) (for $\Phi_k, \Phi_\ell$) yield, as $\lambda \to \infty$,

$$\operatorname{Cov}\bigl(\Phi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_\ell(\widetilde{\mathcal{T}}_\lambda)\bigr) = O(\lambda). \quad (6.56)$$

By selecting a suitable subsequence $(\lambda'_n)$ of the given sequence $(\lambda_n)$, we may thus assume that

$$\sigma_{k\ell}(\lambda) := \frac{\operatorname{Cov}\bigl(\Phi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_\ell(\widetilde{\mathcal{T}}_\lambda)\bigr)}{\lambda} \to \beta_{k\ell} \tag{6.57}$$

as $\lambda \to \infty$ along the subsequence, for all $k, \ell$ and some real $\beta_{k\ell}$.

Let $(t_1, \dots, t_K)$ be an arbitrary vector in $\mathbb{R}^K$ and consider the linear combination

$$\Phi := \sum_{k=1}^{K} t_k \Phi_k. \tag{6.58}$$

This is an additive functional with toll function $\varphi := \sum_{k=1}^K t_k \varphi_k$. Then (5.3), (5.4) and (5.5) hold by the assumptions and Minkowski's inequality. Let $a(\lambda) := \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$ and $b(\lambda) := \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)$, so (5.1) and (5.2) hold trivially. Thus Theorem 5.3 applies, and (5.7) holds.

Furthermore, (6.58) yields

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{k,\ell=1}^{K} t_k t_\ell \operatorname{Cov}\bigl(\Phi_k(\widetilde{\mathcal{T}}_\lambda), \Phi_\ell(\widetilde{\mathcal{T}}_\lambda)\bigr). \tag{6.59}$$

Hence, (6.57) implies that, along the subsequence $(\lambda'_n)$, we have $\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)/\lambda \to \sum_{k,\ell} t_k t_\ell \beta_{k\ell}$, and thus (5.7) implies that

$$\sum_{k=1}^{K} t_k \frac{\Phi_k(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}} \overset{\mathrm{d}}{\longrightarrow} N\Bigl(0, \sum_{k,\ell} t_k t_\ell \beta_{k\ell}\Bigr). \tag{6.60}$$

Since the vector $(t_1, \dots, t_K)$ is arbitrary, it follows by the Cramér–Wold device that, along the subsequence,

$$\Bigl(\frac{\Phi_k(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}}\Bigr)_1^K \overset{\mathrm{d}}{\longrightarrow} N\bigl(0, (\beta_{k\ell})_{k,\ell=1}^K\bigr). \tag{6.61}$$

Combined with (6.57), this shows that (5.19) holds along the subsequence, see Remark 2.4. Since we started with an arbitrary subsequence $(\lambda_n)$, the subsequence principle shows that (5.19) holds in general, see Remark 2.5.

Finally, if we write (5.19) as $X_\lambda \overset{\mathrm{d}}{\approx} Y_\lambda$, then, as in the proof of Theorem 5.3, Lemma 6.3 implies that $\mathbb{E}\,|X_\lambda|^r = O(1)$ for $\lambda \geq 1$, and Lemma 2.6 shows that (5.19) holds with [absolute] moments of order $s$ for $s < r$. □

*Proof of Corollary 5.7.* Consider again a subsequence where (6.57) holds for some $\beta_{k\ell}$. By (5.21), we also have

$$\frac{b_{k\ell}(\lambda)}{\lambda} \to \beta_{k\ell}. \tag{6.62}$$

Hence the proof of Theorem 5.6 just given shows that (5.19) holds also with $\Sigma(\lambda)$ defined by (5.23) instead of (5.20).

Furthermore, (5.1) implies that we may replace $\mathbb{E}\,\Phi_k(\lambda)$ by $a_k(\lambda)$ in (6.61), and thus in (5.19). The result (5.22) follows. Finally, the same argument as in the proofs of Theorems 5.3 and 5.6 shows that (5.22) holds with [absolute] moments of order $s < r$. □

We turn to proofs of the theorems for the model $\mathcal{T}_n$ with a given number of leaves. This time we begin with the multivariate version.

*Proof of Theorem 5.8.* We consider the toll functions $\varphi_{k\pm}$, and also the toll function $\varphi_\bullet(T) := \mathbf{1}\{T = \bullet\}$ in Example 2.1; recall that $\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = N_\lambda$ by (2.20). Note that (5.4) and (5.5) are trivial for $\varphi_\bullet$, since $\varphi_\bullet(T) = O(1)$, and that (5.3) holds for $\Phi_\bullet$ because

$$\operatorname{Var}\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = \operatorname{Var} N_\lambda = \lambda. \tag{6.63}$$

Hence, Theorem 5.6 applies to the set of toll functions $\{\varphi_\bullet, \varphi_{k\pm}\}$. (We use $\bullet$ and $k\pm$ (for $k = 1, \dots, K$) as indices instead of $\{1, \dots, 2K+1\}$.)

Consider $\mathbb{R}^{2K}$ with the usual coordinate-wise partial order, i.e., $(x_i)_1^{2K} \le (y_i)_1^{2K}$ if $x_i \le y_i$ for every $i$. Since each $\Phi_{k\pm}$ by assumption is an increasing functional, and $\mathcal{T}_{n+1}$ is obtain by adding a new string to $\mathcal{T}_n$, it follows that if $n_1 \le n_2$, then

$$\bigl(\Phi_{k\pm}(\mathcal{T}_{n_1})\bigr)_{k\pm} \le \bigl(\Phi_{k\pm}(\mathcal{T}_{n_2})\bigr)_{k\pm} \qquad \text{in } \mathbb{R}^{2K}. \tag{6.64}$$

Furthermore, by the construction of the random trie $\widetilde{\mathcal{T}}_\lambda$, if we condition on $N_\lambda = n$, then we recover $\mathcal{T}_n$ (in distribution), i.e., $\bigl(\widetilde{\mathcal{T}}_\lambda \mid N_\lambda = n\bigr) \overset{\mathrm{d}}{=} \mathcal{T}_n$. It follows that the random vector $(\Phi_{k\pm}(\widetilde{\mathcal{T}}_\lambda))_{k\pm}$ is stochastically increasing in $\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = N_\lambda$ in the sense that for any $\mathbf{x} \in \mathbb{R}^{2K}$ and $n_1 \le n_2$,

$$\begin{aligned}\mathbb{P}\Bigl((\Phi_{k\pm}(\widetilde{\mathcal{T}}_\lambda))_{k\pm} \le \mathbf{x} \Bigm| \Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = n_1\Bigr) &= \mathbb{P}\bigl((\Phi_{k\pm}(\mathcal{T}_{n_1}))_{k\pm} \le \mathbf{x}\bigr)\\ \ge \mathbb{P}\bigl((\Phi_{k\pm}(\mathcal{T}_{n_2}))_{k\pm} \le \mathbf{x}\bigr) &= \mathbb{P}\Bigl((\Phi_{k\pm}(\widetilde{\mathcal{T}}_\lambda))_{k\pm} \le \mathbf{x} \Bigm| \Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = n_2\Bigr).\end{aligned} \tag{6.65}$$

Consider now the sequence $\lambda_n = n$, and take an arbitrary subsequence $(n_j)$ such that, for the set of functionals $\{\Phi_\bullet, \Phi_{k\pm}\}$, the covariances converge as in (6.57), and thus (6.61) holds by the proof of Theorem 5.6 above. Note that then, by (6.57) and (6.63),

$$\beta_{\bullet,\bullet} = \lim_{j\to\infty} \sigma_{\bullet,\bullet}(n_j) = \lim_{j\to\infty} \frac{\operatorname{Var}\Phi_\bullet(\widetilde{\mathcal{T}}_{n_j})}{n_j} = \lim_{j\to\infty} \frac{\operatorname{Var} N_{n_j}}{n_j} = 1, \tag{6.66}$$

and, similarly, $\mathbb{E}\,\Phi_\bullet(\widetilde{\mathcal{T}}_n) = \mathbb{E}\,N_n = n$. We may now apply a theorem by Nerman [27, Theorem 1], or (slightly more conveniently) its corollary [17, Theorem 2.3], which allows us to condition on $\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = n$ in (6.61) (under the stochastic monotonicity (6.65) just shown). Consequently, we obtain that, along the subsequence $(n_j)$,

$$\Bigl(\frac{\Phi_{k\pm}(\mathcal{T}_n) - \mathbb{E}\,\Phi_{k\pm}(\widetilde{\mathcal{T}}_n)}{\sqrt{n}}\Bigr)_{k\pm} \overset{\mathrm{d}}{=} \Bigl(\Bigl(\frac{\Phi_{k\pm}(\widetilde{\mathcal{T}}_n) - \mathbb{E}\,\Phi_{k\pm}(\widetilde{\mathcal{T}}_n)}{\sqrt{n}}\Bigr)_{k\pm} \Bigm| \Phi_\bullet(\widetilde{\mathcal{T}}_n) = n\Bigr)$$

$$\xrightarrow{\mathrm{d}} N\big(\bullet, (\hat\beta_{k\pm,\ell\pm})\big), \tag{6.67}$$

where, for $\eta_1, \eta_2 \in \{+, -\}$, recalling (6.66),

$$\hat\beta_{k\eta_1,\ell\eta_2} := \beta_{k\eta_1,\ell\eta_2} - \frac{\beta_{k\eta_1,\bullet}\beta_{\ell\eta_2,\bullet}}{\beta_{\bullet,\bullet}} = \beta_{k\eta_1,\ell\eta_2} - \beta_{k\eta_1,\bullet}\beta_{\ell\eta_2,\bullet}. \tag{6.68}$$

Note that in (6.67) we normalize $\Phi_{k\pm}(\mathcal{T}_n)$ using $\mathbb{E}\,\Phi_{k\pm}(\widetilde{\mathcal{T}}_n)$ for the Poisson model.

Since $\Phi_k = \Phi_{k+} - \Phi_{k-}$, it follows from (6.67) that, along the subsequence,

$$\Big(\frac{\Phi_k(\mathcal{T}_n) - \mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_n)}{\sqrt{n}}\Big)_{k=1}^K \xrightarrow{\mathrm{d}} N\big(0, (\hat\beta_{k\ell})_{k,\ell=1}^K\big), \tag{6.69}$$

where, using (6.68),

$$\begin{aligned}\hat\beta_{k\ell} &= \hat\beta_{k+,\ell+} - \hat\beta_{k+,\ell-} - \hat\beta_{k-,\ell+} + \hat\beta_{k-,\ell-} \\ &= \beta_{k+,\ell+} - \beta_{k+,\ell-} - \beta_{k-,\ell+} + \beta_{k-,\ell-} - (\beta_{k+,\bullet} - \beta_{k-,\bullet})(\beta_{\ell+,\bullet} - \beta_{\ell-,\bullet}).\end{aligned} \tag{6.70}$$

We are considering a subsequence such that (6.57) holds for the functionals $\{\Phi_\bullet, \Phi_\pm\}$ along the subsequence. It follows from (5.25), (5.21), (5.11), (6.57) (for the set $\{\Phi_\bullet, \Phi_\pm\}$), $\Phi_\bullet(\widetilde{\mathcal{T}}_n) = N_n$, linearity and (6.70), that, along the subsequence,

$$\begin{aligned}\widehat\sigma_{k\ell}(n) &= \frac{\operatorname{Cov}\big(\Phi_k(\widetilde{\mathcal{T}}_n), \Phi_\ell(\widetilde{\mathcal{T}}_n)\big)}{n} - \frac{\operatorname{Cov}\big(\Phi_k(\widetilde{\mathcal{T}}_n), N_n\big)}{n}\frac{\operatorname{Cov}\big(\Phi_\ell(\widetilde{\mathcal{T}}_n), N_n\big)}{n} + o(1) \\ &\to \hat\beta_{k\ell}.\end{aligned} \tag{6.71}$$

By (5.1), we may replace $\mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_n)$ by $a_k(n)$ in (6.69), and thus (6.71) shows that (5.24) holds along the subsequence. Hence, (5.24) holds in general by the subsequence principle.

Furthermore, the proof of Theorem 5.6 shows also that, along the subsequence $(n_j)$ above, (6.61) holds with absolute moments of order $s < r$. By [27, Section 4] (see also [17, Theorem 2.6]), the same holds after conditioning on $\Phi_\bullet(\widetilde{\mathcal{T}}_n) = n$, i.e., in (6.67). Since absolute moment convergence here is equivalent to uniform $s$th power integrability [11, Theorem 5.5.9], it follows that also (6.69) holds with uniform $s$th power integrability. We may again replace $\mathbb{E}\,\Phi_k(\widetilde{\mathcal{T}}_n)$ by $a_k(n)$, using (5.1). Hence, (5.24) holds along the subsequence with uniform $s$th power integrability, and thus with convergence of $s$th [absolute] moments. Hence, by the subsequence principle again, (5.24) holds with $s$th [absolute] moments.

In particular, (5.24) holds with moments of order 1 and 2, which gives (5.26) and (5.27). □

*Proof of Theorem 5.5.* This is essentially the special case $K = 1$ of Theorem 5.8. In part (i), we do not assume any function $a(\lambda)$. However, we may then define $a(\lambda) := \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)$, so (5.1) holds trivially. Thus we may throughout the proof assume that we have a function $a(\lambda)$ such that (5.1) holds. Then Theorem 5.8 applies with $K = 1$. In particular, (5.26)–(5.27) hold, which yields (5.12)–(5.13) using choice $a(\lambda) = \operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)$ just made for (i); then (5.14) follows by (5.2) and (5.11), noting that (5.3) implies $\operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) = O(\lambda)$ by the Cauchy–Schwarz inequality. The approximations (5.15) follow from (5.24) and (5.12)–(5.13), with [absolute] moments of order $s < r$.

For part (ii), we have by (5.24) (or (5.15)),

$$\frac{\Phi(\mathcal{T}_n) - a(n)}{\sqrt{n}} \overset{\mathrm{d}}{\approx} N\bigl(0, \widehat{\sigma}^2(n)\bigr), \tag{6.72}$$

with

$$\widehat{\sigma}^2(n) := \frac{b(n) - c(n)^2/n}{n}. \tag{6.73}$$

The assumptions (5.3), (5.2) and (5.16) imply $\widehat{\sigma}^2(n) = \Theta(1)$. Hence, (6.72) implies that for any subsequence such that $\widehat{\sigma}^2(n)$ converges, say $\widehat{\sigma}^2(n) \to \gamma$, we have $\gamma > 0$, and then (6.72) implies

$$\frac{\Phi(\mathcal{T}_n) - a(n)}{\sqrt{b(n) - c(n)^2/n}} = \frac{\Phi(\mathcal{T}_n) - a(n)}{\sqrt{n\widehat{\sigma}^2(n)}} \overset{\mathrm{d}}{\longrightarrow} N(0,1) \tag{6.74}$$

along the subsequence. By the subsequence principle, (6.74) holds in general, which is (5.17). This yields also (5.18), using (5.12)–(5.13), (5.1), and again (5.16). Moment convergence follows by the same argument. □

## 7. Proof of Theorem 3.1

Before proving Theorem 3.1, we give some lemmas. To begin with, we assume that $\varphi(\bullet) = 0$.

**Lemma 7.1.** *Suppose that $\varphi(\bullet) = 0$ and $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)| < \infty$ for some $\lambda > 0$. Then*

$$\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), \tag{7.1}$$

*where the sums have finite summands and converge absolutely. Moreover,*

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\bigl|\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\bigr| = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\bigl|\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})\bigr| < \infty. \tag{7.2}$$

*Proof.* By Lemma 6.1(ii), with $m = 2$,

$$\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})| \le P(\boldsymbol{\alpha})^2 e^{\lambda}\,\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)| = C_\lambda P(\boldsymbol{\alpha})^2. \tag{7.3}$$

Hence, using (2.7),

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})| \le C_\lambda \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} P(\boldsymbol{\alpha})^2 < \infty, \tag{7.4}$$

which proves the inequality in (7.2). The equality in (7.2) follows from (2.25).

Finally, the first equality in (7.1) follows by (2.17) and Fubini's theorem, using (7.2) which also implies absolute convergence of the sum. The second equality follows by (2.25). □

For the variance, we give in the next lemma several different formulas.

**Lemma 7.2.** *Suppose that* $\varphi(\bullet) = 0$ *and* $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^2 < \infty$ *for some* $\lambda > 0$. *Then*

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathcal{A}^*} \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}), \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\beta}})\bigr) \tag{7.5}$$

$$= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathcal{A}^*} \mathbb{E}\bigl[\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\beta}})\bigr] - \Bigl(\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\Bigr)^2 \tag{7.6}$$

$$= 2\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}), \Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})\bigr) - \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) \tag{7.7}$$

$$= \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \Bigl(2\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), \Phi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})\bigr) - \operatorname{Var}\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})\Bigr) \tag{7.8}$$

*and*

$$\operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), N_{\lambda P(\boldsymbol{\alpha})}\bigr) \tag{7.9}$$

$$= \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\bigl[\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})N_{\lambda P(\boldsymbol{\alpha})}\bigr] - \lambda \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} P(\boldsymbol{\alpha})\,\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), \tag{7.10}$$

*where all sums have finite summands and converge absolutely.*

**Remark 7.3.** Analoguous formulas for the covariance $\operatorname{Cov}\bigl(\Phi_1(\widetilde{\mathcal{T}}_\lambda), \Phi_2(\widetilde{\mathcal{T}}_\lambda)\bigr)$ for two toll functions $\varphi_1$ and $\varphi_2$ with $\varphi_1(\bullet) = \varphi_2(\bullet) = 0$ follow immediately by polarization in (7.5)–(7.8); the details are omitted.

However, note that these formulas for variance and covariance do not include the case $\varphi_\bullet(T) := \mathbf{1}\{T = \bullet\}$ with $\Phi_\bullet(T) = |T|_{\mathsf{e}}$, see Example 2.1. (It is easily checked that e.g. (7.8) and (7.9) fail for $\varphi_\bullet$.) Hence, separate formulas are given in Lemma 7.2 for the covariance with $\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = N_\lambda$. □

*Proof of Lemma 7.2.* Let, recalling (2.25),

$$X_{\boldsymbol{\alpha}} := \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}) \overset{\mathrm{d}}{=} \varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), \tag{7.11}$$

and define, for $k \geq 0$,

$$Y_k := \sum_{|\boldsymbol{\alpha}|=k} X_{\boldsymbol{\alpha}}. \tag{7.12}$$

By (7.11) and Lemma 6.1(ii), applied to $\varphi^2$ and with $m = 2$, cf. (7.3),

$$\mathbb{E}\,X_{\boldsymbol{\alpha}}^2 = \mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})|^2 \leq C_\lambda P(\boldsymbol{\alpha})^2. \tag{7.13}$$

Furthermore, for any $k \geq 0$, the random variables $X_{\boldsymbol{\alpha}}$ with $|\boldsymbol{\alpha}| = k$ are independent. Hence, using (7.13) and (2.5),

$$\operatorname{Var} Y_k = \sum_{|\boldsymbol{\alpha}|=k} \operatorname{Var} X_{\boldsymbol{\alpha}} \leq \sum_{|\boldsymbol{\alpha}|=k} \mathbb{E}\, X_{\boldsymbol{\alpha}}^2 \leq \sum_{|\boldsymbol{\alpha}|=k} C_\lambda P(\boldsymbol{\alpha})^2 = C_\lambda \rho(2)^k. \tag{7.14}$$

Since $\rho(2) < 1$ by (2.6), it follows from (2.17), (7.11) and (7.12) that

$$\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \bigl(X_{\boldsymbol{\alpha}} - \mathbb{E}\, X_{\boldsymbol{\alpha}}\bigr) = \sum_{k=0}^{\infty} \bigl(Y_k - \mathbb{E}\, Y_k\bigr), \tag{7.15}$$

where the sums converge absolutely in $L^1$ since $\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \|X_{\boldsymbol{\alpha}}\|_1 < \infty$ by (7.2), and the final sum converges absolutely in $L^2$ by (7.14), i.e., $\sum_{k=0}^\infty \|Y_k - \mathbb{E}\, Y_k\|_2 < \infty$. (The first sum does not always converge absolutely in $L^2$; this is why we introduce $Y_k$.) Hence,

$$\operatorname{Var} \Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{k,\ell\geq 0} \operatorname{Cov}(Y_k, Y_\ell) < \infty \tag{7.16}$$

with absolute convergence.

Suppose temporarily that $\varphi \geq 0$. Then $\sum_{k=0}^\infty \mathbb{E}\, Y_k = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} \mathbb{E}\, X_{\boldsymbol{\alpha}} < \infty$ by (7.12), (7.11), and (7.2), and thus (7.16) yields

$$\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathcal{A}^*} \mathbb{E}\bigl[X_{\boldsymbol{\alpha}} X_{\boldsymbol{\beta}}\bigr] = \sum_{k,\ell} \mathbb{E}\bigl[Y_k Y_\ell\bigr] = \sum_{k,\ell} \bigl(\operatorname{Cov}(Y_k, Y_\ell) + \mathbb{E}\, Y_k\, \mathbb{E}\, Y_\ell\bigr) < \infty. \tag{7.17}$$

Returning to a general $\varphi$, we apply (7.17) to $|\varphi|$, and find

$$\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathcal{A}^*} \mathbb{E}\bigl|X_{\boldsymbol{\alpha}} X_{\boldsymbol{\beta}}\bigr| < \infty. \tag{7.18}$$

We see from (7.18) and (7.2) that the sums in (7.6) are absolutely convergent. It follows that so is the sum in (7.5), and that it equals (7.6); furthermore, recalling (7.11) and (7.12), this sum equals $\sum_{k,\ell} \operatorname{Cov}(Y_k, Y_\ell)$. Hence (7.16) implies (7.5) and (7.6).

Next, rewrite (7.16) as

$$\operatorname{Var} \Phi(\widetilde{\mathcal{T}}_\lambda) = 2 \sum_{0\leq k\leq \ell} \operatorname{Cov}(Y_k, Y_\ell) - \sum_{k=0}^{\infty} \operatorname{Var} Y_k. \tag{7.19}$$

Let $k \leq \ell$. If $\boldsymbol{\alpha} \in \mathcal{A}^k$ and $\boldsymbol{\beta} \in \mathcal{A}^\ell$ and $\boldsymbol{\alpha}$ is not a prefix of $\boldsymbol{\beta}$, then $X_{\boldsymbol{\alpha}}$ and $X_{\boldsymbol{\beta}}$ are independent. Thus,

$$\operatorname{Cov}(Y_k, Y_\ell) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k,\, \boldsymbol{\beta}\in\mathcal{A}^\ell} \operatorname{Cov}(X_{\boldsymbol{\alpha}}, X_{\boldsymbol{\beta}}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k,\, \boldsymbol{\gamma}\in\mathcal{A}^{\ell-k}} \operatorname{Cov}(X_{\boldsymbol{\alpha}}, X_{\boldsymbol{\alpha\gamma}}). \tag{7.20}$$

Hence, for any $k \geq 0$, recalling (7.11) and absolute convergence in (7.5),

$$\sum_{\ell\geq k} \operatorname{Cov}(Y_k, Y_\ell) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k,\, \boldsymbol{\gamma}\in\mathcal{A}^*} \operatorname{Cov}(X_{\boldsymbol{\alpha}}, X_{\boldsymbol{\alpha\gamma}}), \tag{7.21}$$

with absolute convergence, also when summed over $k$. Furthermore, by (7.15) applied to $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}$,

$$\Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \sum_{j=0}^{\infty}\sum_{|\boldsymbol{\gamma}|=j}\bigl(X_{\boldsymbol{\alpha\gamma}} - \mathbb{E}\,X_{\boldsymbol{\alpha\gamma}}\bigr) \tag{7.22}$$

with the sum over $j$ converging in $L^2$. Thus, for each $\boldsymbol{\alpha}$,

$$\operatorname{Cov}(\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}), \Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})) = \sum_{j=0}^{\infty}\sum_{|\boldsymbol{\gamma}|=j}\operatorname{Cov}(X_{\boldsymbol{\alpha}}, X_{\boldsymbol{\alpha\gamma}}) = \sum_{\boldsymbol{\gamma}\in\mathcal{A}^*}\operatorname{Cov}(X_{\boldsymbol{\alpha}}, X_{\boldsymbol{\alpha\gamma}}). \tag{7.23}$$

Hence, (7.21) yields, with absolute convergence, also when summed over $k$,

$$\sum_{\ell\geq k}\operatorname{Cov}(Y_k, Y_\ell) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k}\operatorname{Cov}(\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}), \Phi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})). \tag{7.24}$$

Consequently, (7.7) follows from (7.19) and (7.14). Finally, (7.8) follows from (7.7) by (2.23).

For the covariance, let $\psi(T) := \varphi(T)|T|_{\mathsf{e}}$, so that

$$\mathbb{E}\,\psi(\widetilde{\mathcal{T}}_\lambda) = \mathbb{E}\bigl[\varphi(\widetilde{\mathcal{T}}_\lambda)|\widetilde{\mathcal{T}}_\lambda|_{\mathsf{e}}\bigr] = \mathbb{E}\bigl[\varphi(\widetilde{\mathcal{T}}_\lambda)N_\lambda\bigr]. \tag{7.25}$$

Since $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^2 < \infty$ and $\mathbb{E}\,N_\lambda^2 < \infty$, the Cauchy–Schwarz inequality implies that $\mathbb{E}\,|\psi(\widetilde{\mathcal{T}}_\lambda)| < \infty$. Hence Lemma 6.2 applies to both $\varphi$ and $\psi$, which shows that both sums in (7.10) converge absolutely. Thus so does the sum in (7.9).

For any $\boldsymbol{\alpha} \in \mathcal{A}^*$, by properties of the Poisson distribution, the two set of strings $\{\Xi^{(k)} : k \leq N_\lambda \text{ and } \Xi^{(k)} \not\succ \boldsymbol{\alpha}\}$ and $\{\Xi^{(k)} : k \leq N_\lambda \text{ and } \Xi^{(k)} \succ \boldsymbol{\alpha}\}$ are independent. Consequently, $N_\lambda - N_{\lambda,\boldsymbol{\alpha}}$ is independent of $N_{\lambda,\boldsymbol{\alpha}}$ and of $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ and thus of $X_{\boldsymbol{\alpha}} = \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+})$. Hence, recalling (2.22)–(2.23),

$$\begin{aligned}\operatorname{Cov}\bigl(X_{\boldsymbol{\alpha}}, N_\lambda\bigr) &= \operatorname{Cov}\bigl(X_{\boldsymbol{\alpha}}, N_{\lambda,\boldsymbol{\alpha}}\bigr) = \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}), |\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}|_{\mathsf{e}}\bigr)\\ &= \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}), |\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}|_{\mathsf{e}}\bigr)\\ &= \mathbb{E}\,\psi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})})\lambda P(\boldsymbol{\alpha}).\end{aligned} \tag{7.26}$$

We sum (7.26) first over $\boldsymbol{\alpha} \in \mathcal{A}^k$, and then over $k \geq 0$, using the $L^2$ convergence in (7.15), and obtain (7.9)–(7.10). ☐

**Lemma 7.4.** *Suppose that $\varphi$ is a toll function such that as $\lambda \to \infty$,*

$$\operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) = O\bigl(\lambda^{1-\varepsilon}\bigr) \tag{7.27}$$

*for some $\varepsilon > 0$. Then*

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = O\bigl(\lambda\bigr), \qquad \lambda \in (0,\infty). \tag{7.28}$$

*Proof.* By subtracting a suitable multiple of $\varphi_\bullet(T) := \mathbf{1}\{T = \bullet\}$ from $\varphi$, we may assume that $\varphi(\bullet) = 0$. (Because $\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = N_\lambda$ satisfies (7.28).)

By (7.27), there exist $\lambda_0$ and $C_1$ such that, for $\lambda \geq \lambda_0$,

$$\operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda) \leq C_1 \lambda^{1-\varepsilon} \tag{7.29}$$

Lemma 6.1(ii) applies to $\varphi^2$, with $\lambda_2 = \lambda_0$ and $m = 2$, and shows that, for $\lambda \leq \lambda_0$,

$$\operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda) \leq \mathbb{E}[\varphi(\widetilde{\mathcal{T}}_\lambda)^2] \leq C_2 \lambda^2. \tag{7.30}$$

It follows that, perhaps after increasing $C_1$ and $C_2$, (7.29) and (7.30) both hold for all $\lambda \in (0, \infty)$.

Let $J_k := (p_{\max}^{k+1}, p_{\max}^k]$, $k \geq 0$, and define, recalling (2.25),

$$Z_{\lambda,k} := \sum_{P(\boldsymbol{\alpha}) \in J_k} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}}) = \sum_{P(\boldsymbol{\alpha}) \in J_k} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}). \tag{7.31}$$

Thus, we have by (2.17) the decomposition

$$\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{k=0}^{\infty} Z_{\lambda,k}. \tag{7.32}$$

If $\boldsymbol{\alpha}$ is a prefix of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$, then $P(\boldsymbol{\beta}) \leq p_{\max} P(\boldsymbol{\alpha})$, and thus $P(\boldsymbol{\alpha})$ and $P(\boldsymbol{\beta})$ cannot both belong to the same $J_k$. Hence, the modified fringe tries $\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}$ are independent for all $\boldsymbol{\alpha}$ with $P(\boldsymbol{\alpha}) \in J_k$. Consequently, using (2.25),

$$\operatorname{Var} Z_{\lambda,k} = \sum_{P(\boldsymbol{\alpha}) \in J_k} \operatorname{Var} \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}+}) = \sum_{P(\boldsymbol{\alpha}) \in J_k} \operatorname{Var} \varphi(\widetilde{\mathcal{T}}_{\lambda P(\boldsymbol{\alpha})}). \tag{7.33}$$

By definition, $P(\boldsymbol{\alpha})$ is the probability that the random string $\Xi$ has $\boldsymbol{\alpha}$ as a prefix; hence $\sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha})$ is the expected number of prefixes $\boldsymbol{\alpha}$ in $\Xi$ with $P(\boldsymbol{\alpha}) \in J_k$. Since none of these strings $\boldsymbol{\alpha}$ is a prefix of another, as just seen, $\Xi$ can contain at most one such prefix. Hence,

$$\sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha}) \leq 1. \tag{7.34}$$

Combining (7.33) with (7.29) and (7.34), we obtain

$$\begin{aligned} \operatorname{Var} Z_{\lambda,k} &\leq \sum_{P(\boldsymbol{\alpha}) \in J_k} C_1 (\lambda P(\boldsymbol{\alpha}))^{1-\varepsilon} = C_1 \lambda^{1-\varepsilon} \sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha})^{1-\varepsilon} \\ &\leq C_1 \lambda^{1-\varepsilon} p_{\max}^{-\varepsilon(k+1)} \sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha}) \leq C_3 \lambda \bigl(\lambda p_{\max}^k\bigr)^{-\varepsilon}. \end{aligned} \tag{7.35}$$

Similarly, using instead (7.30),

$$\begin{aligned} \operatorname{Var} Z_{\lambda,k} &\leq \sum_{P(\boldsymbol{\alpha}) \in J_k} C_2 (\lambda P(\boldsymbol{\alpha}))^2 = C_2 \lambda^2 \sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha})^2 \\ &\leq C_2 \lambda^2 p_{\max}^k \sum_{P(\boldsymbol{\alpha}) \in J_k} P(\boldsymbol{\alpha}) \leq C_2 \lambda \bigl(\lambda p_{\max}^k\bigr). \end{aligned} \tag{7.36}$$

By (7.32) and Minkowski's inequality, (7.35)–(7.36) imply

$$\begin{aligned}\big(\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)\big)^{1/2} &\le \sum_{k=0}^{\infty}\big(\operatorname{Var} Z_{\lambda,k}\big)^{1/2}\\ &\le C_4\lambda^{1/2}\sum_{k=0}^{\infty}\min\big\{\big(\lambda p_{\max}^k\big)^{-\varepsilon/2},\big(\lambda p_{\max}^k\big)^{1/2}\big\}\\ &\le C_5\lambda^{1/2}, \end{aligned}\tag{7.37}$$

since the last sum is dominated by the sum of two convergent geometric series, uniformly in $\lambda$. This shows (7.28). □

**Remark 7.5.** We cannot take $\varepsilon = 0$ in (7.27) and assume only $\operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) = O(\lambda)$. A counter example is $\varphi(T) = |T|_{\mathsf{e}}\mathbf{1}\{|T|_{\mathsf{e}} \ge 2\}$; then $\Phi(T)$ is the external path length, and $\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)$ is of order $\lambda\log^2\lambda$, see [15, Lemma 12]. □

*Proof of Theorem 3.1.* We prove the theorem in two steps, first in the special case $\chi = \varphi(\bullet) = 0$, and then in general.

*Step 1:* $\chi = \varphi(\bullet) = 0$. First, (3.1)–(3.2) show that $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|$ and $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|^2$ are finite for large $\lambda$, and thus for all $\lambda > 0$ by Lemma 6.1. Hence, Lemmas 7.1 and 7.2 apply for any $\lambda$. By (3.16)–(3.18), we can write (7.1), (7.8) and (7.9) as

$$\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f_{\mathsf{E}}(\lambda P(\boldsymbol{\alpha})),\tag{7.38}$$

$$\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f_{\mathsf{V}}(\lambda P(\boldsymbol{\alpha})),\tag{7.39}$$

$$\operatorname{Cov}\big(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^*} f_{\mathsf{C}}(\lambda P(\boldsymbol{\alpha})),\tag{7.40}$$

with absolute convergence; in particular the left-hand sides are finite and so are (taking the term $\boldsymbol{\alpha} = \boldsymbol{\epsilon}$ in the sums) $f_{\mathsf{E}}(\lambda), f_{\mathsf{V}}(\lambda), f_{\mathsf{C}}(\lambda)$ for every $\lambda > 0$.

These equations are all instances of (A.1) in Theorem A.1 in Appendix A, and we verify the conditions of that theorem. First, we may write also $f_{\mathsf{V}}$ and $f_{\mathsf{C}}$ using only expectations of functionals of $\widetilde{\mathcal{T}}_\lambda$:

$$f_{\mathsf{V}}(\lambda) = 2\,\mathbb{E}\big[\varphi(\widetilde{\mathcal{T}}_\lambda)\Phi(\widetilde{\mathcal{T}}_\lambda)\big] - 2\,\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)\,\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)^2 + \big(\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda)\big)^2,\tag{7.41}$$

$$f_{\mathsf{C}}(\lambda) = \mathbb{E}\big[\varphi(\widetilde{\mathcal{T}}_\lambda)N_\lambda\big] - \lambda\,\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda).\tag{7.42}$$

All expectations in (3.16) and (7.41)–(7.42) are finite by (3.1)–(3.2) and Lemma 6.1, (7.38)–(7.39), and the Cauchy–Schwarz inequality. Hence, it follows from the general formula (6.1) that $f_{\mathsf{E}}, f_{\mathsf{V}}, f_{\mathsf{C}}$ are all continuous, and in fact, entire analytic. Furthermore, it follows from Lemma 6.1(ii) that the expectations in (3.16), (7.41) and (7.42) all are $O(\lambda^2)$ for $\lambda \le 1$, and thus

(A.2) holds for $f_{\mathsf{E}}, f_{\mathsf{V}}, f_{\mathsf{C}}$. Equivalently, the entire functions $f_{\mathsf{X}}$ satisfy

$$f_{\mathsf{X}}(0) = f'_{\mathsf{X}}(0) = 0. \tag{7.43}$$

Next, $f_{\mathsf{E}}$ satisfies (A.3) by the assumption (3.1). Furthermore, Lemma 7.4 applies by (3.2) and yields $\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda) = O(\lambda)$. Also, $\operatorname{Var} N_\lambda = \lambda$. Hence (3.17)–(3.18), (3.2) and the Cauchy–Schwarz inequality yield

$$f_{\mathsf{V}}(\lambda), f_{\mathsf{C}}(\lambda) = O\bigl(\lambda^{1-\varepsilon/2}\bigr) \tag{7.44}$$

for large $\lambda$, and thus for $\lambda \geq 1$ (since $f_{\mathsf{V}}$ and $f_{\mathsf{C}}$ are continuous and thus bounded on finite intervals). In other words, (A.3) holds for $f_{\mathsf{V}}$ and $f_{\mathsf{C}}$, with $\varepsilon$ replaced by $\varepsilon/2$.

Hence, Theorem A.1(i)–(iii) apply to $f_{\mathsf{E}}, f_{\mathsf{V}}, f_{\mathsf{C}}$. Furthermore, if $d_{\mathbf{p}} > 0$ and $f'_{\mathsf{X}}(\lambda) = O(\lambda^{-\varepsilon_1})$ as $\lambda \to \infty$, then also Theorem A.1(v) applies; note that $f'_{\mathsf{X}}(\lambda) = O(\lambda)$ as $\lambda \to 0$ by (7.43). In particular, this is always the case for $f_{\mathsf{E}}$ (when $d_{\mathbf{p}} > 0$), since it follows from (3.20) (which will be proved below) and (7.44) that $f'_{\mathsf{E}}(\lambda) = O\bigl(\lambda^{-\varepsilon/2}\bigr)$ for $\lambda \geq 1$.

Consequrently, the results in (i) and (ii) follow from Theorem A.1.

Finally, if $\varphi \geq 0$ and $\varphi(T') > 0$ for some trie $T'$, then $f_{\mathsf{E}}(\lambda) = \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) > 0$ for every $\lambda > 0$ by (6.1); hence (iii) follows by Theorem A.1(iv).

*Step 2: The general case.* Consider the toll function

$$\varphi_* := \varphi - \chi\varphi_\bullet \tag{7.45}$$

and the corresponding additive functional $\Phi_*$; note that $\varphi_*(\bullet) = 0$. Then

$$\Phi_*(\widetilde{\mathcal{T}}_\lambda) = \Phi(\widetilde{\mathcal{T}}_\lambda) - \chi\Phi_\bullet(\widetilde{\mathcal{T}}_\lambda) = \Phi(\widetilde{\mathcal{T}}_\lambda) - \chi N_\lambda. \tag{7.46}$$

Define, for $\mathsf{X} = \mathsf{E}, \mathsf{V}, \mathsf{C}$ as usual, $f_{\mathsf{X},*}$ by (3.16)–(3.18) for the functionals $\varphi_*$ and $\Phi_*$, and define $\psi_{\mathsf{X},*}$ by (3.13)–(3.14) with $f_{\mathsf{X},*}$. Then, by the case just proved, (3.10)–(3.12) hold for moments of $\Phi_*(\widetilde{\mathcal{T}}_\lambda)$, if we omit the terms $\chi$ or $\chi^2$ and replace $\psi_{\mathsf{X}}$ by $\psi_{\mathsf{X},*}$. Hence, using (7.46) and recalling $\mathbb{E}\,N_\lambda = \operatorname{Var} N_\lambda = \lambda$,

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \frac{\mathbb{E}\,\Phi_*(\widetilde{\mathcal{T}}_\lambda) + \chi\,\mathbb{E}\,N_\lambda}{\lambda} = \frac{1}{H}\psi_{\mathsf{E},*}(\log\lambda) + \chi + o(1), \tag{7.47}$$

$$\begin{aligned}\frac{\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} &= \frac{\operatorname{Var}\Phi_*(\widetilde{\mathcal{T}}_\lambda) + 2\chi\operatorname{Cov}(\Phi_*(\widetilde{\mathcal{T}}_\lambda), N_\lambda) + \chi^2\operatorname{Var} N_\lambda}{\lambda}\\ &= \frac{1}{H}\psi_{\mathsf{V},*}(\log\lambda) + 2\chi\frac{1}{H}\psi_{\mathsf{C},*}(\log\lambda) + \chi^2 + o(1),\end{aligned} \tag{7.48}$$

$$\begin{aligned}\frac{\operatorname{Cov}\bigl(\Phi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr)}{\lambda} &= \frac{\operatorname{Cov}\bigl(\Phi_*(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) + \chi\operatorname{Var} N_\lambda}{\lambda}\\ &= \frac{1}{H}\psi_{\mathsf{C},*}(\log\lambda) + \chi + o(1).\end{aligned} \tag{7.49}$$

This proves (3.10)–(3.12) (and thus (3.7)–(3.9) when $d_{\mathbf{p}} = 0$) if we define

$$\psi_{\mathsf{E}} := \psi_{\mathsf{E},*}, \qquad \psi_{\mathsf{V}} := \psi_{\mathsf{V},*} + 2\chi\psi_{\mathsf{C},*}, \qquad \psi_{\mathsf{C}} := \psi_{\mathsf{C},*}, \tag{7.50}$$

which agrees with (3.13)–(3.15) if we have

$$f_{\mathsf{E}} = f_{\mathsf{E},*}, \qquad f_{\mathsf{V}} = f_{\mathsf{V},*} + 2\chi f_{\mathsf{C},*}, \qquad f_{\mathsf{C}} = f_{\mathsf{C},*}. \tag{7.51}$$

It remains to verify that (7.51) agrees with the definitions (3.4)–(3.6). In fact, (7.51) yields, by (3.16)–(3.18) and (7.45)–(7.46),

$$f_{\mathsf{E}}(\lambda) = f_{\mathsf{E},*}(\lambda) = \mathbb{E}\,\varphi_*(\widetilde{\mathcal{T}}_\lambda) = \mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) - \chi\,\mathbb{E}\,\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), \tag{7.52}$$

$$\begin{aligned} f_{\mathsf{V}}(\lambda) &= f_{\mathsf{V},*}(\lambda) + 2\chi f_{\mathsf{C},*}(\lambda) \\ &= 2\operatorname{Cov}\bigl(\varphi_*(\widetilde{\mathcal{T}}_\lambda), \Phi_*(\widetilde{\mathcal{T}}_\lambda)\bigr) - \operatorname{Var}\bigl(\varphi_*(\widetilde{\mathcal{T}}_\lambda)\bigr) + 2\chi\operatorname{Cov}\bigl(\varphi_*(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) \\ &= 2\operatorname{Cov}\bigl(\varphi_*(\widetilde{\mathcal{T}}_\lambda), \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) - \operatorname{Var}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda) - \chi\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda)\bigr) \\ &= 2\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) - 2\chi\operatorname{Cov}\bigl(\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) \\ &\qquad - \operatorname{Var}\varphi(\widetilde{\mathcal{T}}_\lambda) + 2\chi\operatorname{Cov}\bigl(\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), \varphi(\widetilde{\mathcal{T}}_\lambda)\bigr) - \chi^2\operatorname{Var}\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), \end{aligned} \tag{7.53}$$

$$f_{\mathsf{C}}(\lambda) = f_{\mathsf{C},*}(\lambda) = \operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) - \chi\operatorname{Cov}\bigl(\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr). \tag{7.54}$$

Furthermore, recalling that $\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda) = \mathbf{1}\{N_\lambda = 1\}$, we have

$$\mathbb{E}\,\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda) = \mathbb{P}(N_\lambda = 1) = \lambda e^{-\lambda}, \tag{7.55}$$

$$\operatorname{Var}\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda) = \lambda e^{-\lambda}\bigl(1 - \lambda e^{-\lambda}\bigr), \tag{7.56}$$

$$\operatorname{Cov}\bigl(\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), N_\lambda\bigr) = \mathbb{P}(N_\lambda = 1) - \mathbb{P}(N_\lambda = 1)\lambda = (1-\lambda)\lambda e^{-\lambda}. \tag{7.57}$$

Also, since $\Phi(\bullet) = \varphi(\bullet)$ and thus $\varphi_\bullet(T)\bigl(\varphi(T) - \Phi(T)\bigr) = 0$ for every $T$, we have

$$\begin{aligned} \operatorname{Cov}\bigl(\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda), \varphi(\widetilde{\mathcal{T}}_\lambda) - \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) &= -\,\mathbb{E}\,\varphi_\bullet(\widetilde{\mathcal{T}}_\lambda)\,\mathbb{E}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda) - \Phi(\widetilde{\mathcal{T}}_\lambda)\bigr) \\ &= -\lambda e^{-\lambda}\bigl(\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)\bigr). \end{aligned} \tag{7.58}$$

Combining (7.52)–(7.54) and (7.55)–(7.58), we obtain (3.4)–(3.6) after simple calculations.

The assertion on absolute convergence of the Fourier series follows as in the case $\chi = 0$ from Theorem A.1(v), using (7.51) to verify $f_{\mathsf{X}}(\lambda) = O(\lambda^2)$ for $\lambda < 1$.

This completes the proof of (i)–(ii). For (iii), we note that this was proved in Step 1 if $\chi = 0$. If $\chi > 0$, we note that $\varphi_* \geq 0$ and thus $f_{\mathsf{E}}(\lambda) = \mathbb{E}\,\varphi_*(\widetilde{\mathcal{T}}_\lambda) \geq 0$; hence $\psi_{\mathsf{E}}(t) \geq 0$ by (3.15) and $\inf_t\bigl(H^{-1}\psi_{\mathsf{E}}(t) + \chi\bigr) \geq \chi > 0$. Alternatively, we may simply note that $\Phi_* \geq 0$ and thus (7.46) implies $\Phi(\widetilde{\mathcal{T}}_\lambda) \geq \chi N_\lambda$ and $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) \geq \chi\lambda$. □

Note that (7.43)–(7.44) and the comments between them justify the claims on existence of the Mellin transforms $f^*_{\mathsf{X}}(s)$ in Remark 3.4.

*Proof of Lemma 3.6.* By the proof of Theorem 3.1, we have $f_{\mathsf{E}} = f_{\mathsf{E},*}$ and $f_{\mathsf{C}} = f_{\mathsf{C},*}$, so it suffices to consider the case $\chi = 0$. In this case, it follows from (6.1) that the derivative of the entire analytic function $f_{\mathsf{E}}(\lambda)$ is, using (3.16) and (3.18) (or (7.42)),

$$\begin{aligned} f_{\mathsf{E}}'(\lambda) &= -f_{\mathsf{E}}(\lambda) + e^{-\lambda}\sum_{n=1}^{\infty}\frac{n\lambda^{n-1}}{n!}a_n = -f_{\mathsf{E}}(\lambda) + \lambda^{-1}\,\mathbb{E}\big(N_\lambda\varphi(\widetilde{\mathcal{T}}_\lambda)\big) \\ &= \frac{-\lambda\,\mathbb{E}\,\varphi(\widetilde{\mathcal{T}}_\lambda) + \mathbb{E}\big(\varphi(\widetilde{\mathcal{T}}_\lambda)N_\lambda\big)}{\lambda} = \frac{f_{\mathsf{C}}(\lambda)}{\lambda}, \end{aligned} \tag{7.59}$$

which proves (3.20). Furthermore, for $-2 < \operatorname{Re} s < -1 + \varepsilon/2$ (cf. Remark 3.4), (7.59) and an integration by parts gives, using (3.1) and $f_{\mathsf{E}}(\lambda) = O(\lambda^2)$ shown above,

$$f_{\mathsf{C}}^*(s) = \int_0^\infty f_{\mathsf{E}}'(\lambda)\lambda^s \,\mathrm{d}\lambda = -s\int_0^\infty f_{\mathsf{E}}(\lambda)\lambda^{s-1}\,\mathrm{d}\lambda = -sf_{\mathsf{E}}^*(s), \tag{7.60}$$

showing (3.21). Finally, (3.22) follows from (3.21) when $d_{\mathbf{p}} = 0$, and otherwise from either (3.15) and (3.20), or (3.14) and (3.21); we omit the details. ☐

## 8. Proof of Theorem 3.9 – Lemma 3.16

Finally, we prove Theorem 3.9 and the remaining other results in Section 3.

*Proof of Theorem 3.9.* Since $\varphi$ and $\varphi_\pm$ are bounded, (3.1)–(3.2) hold for $\varphi$ and $\varphi_\pm$ (with $\varepsilon = 1$), and thus Theorem 3.1 applies to these functionals. In particular, (3.11) (or Lemma 7.4) implies that (5.3) holds for $\Phi$ and $\Phi_\pm$. Furthermore, (5.4) and (5.5) (for any $r > 0$) hold trivially for $\varphi$ and $\varphi_\pm$, again because the functionals are bounded. Moreover, define

$$a(\lambda) := \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda), \tag{8.1}$$

$$b(\lambda) := \Big(\chi^2 + \frac{1}{H}\psi_{\mathsf{V}}(\log\lambda)\Big)\lambda, \tag{8.2}$$

Then (5.1) holds trivially, and (5.2) holds by (3.11). Consequently, Theorem 5.3 applies, with any $r > 2$, and (5.6) yields

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\sqrt{\lambda}} \overset{\mathrm{d}}{\approx} N\big(0, b(\lambda)/\lambda\big), \tag{8.3}$$

with all [absolute] moments, which by (8.2) is (3.29) with (3.31). In the special case $d_{\mathbf{p}} = 0$, this yields (3.25) with (3.27).

Moreover, define also

$$c(\lambda) := \Big(\chi + \frac{1}{H}\psi_{\mathsf{C}}(\log\lambda)\Big)\lambda. \tag{8.4}$$

Then Theorem 3.1 shows also that (5.11) holds, and thus Theorem 5.5 applies with any $r > 2$. Hence (5.15) holds, with all [absolute] moments, which, recalling (8.2) and (8.4), yields (3.30), with (3.32). In the special case $d_{\mathbf{p}} = 0$, this yields (3.26) with (3.28). This proves (i) and (ii).

(iii) follows by (5.12).

For (iv), suppose that $\liminf_{n\to\infty} \operatorname{Var}\Phi(\mathcal{T}_n)/n > 0$. By (5.13), this means that (5.16) holds. Hence, (5.18) holds, which is (3.35), with all [absolute] moments. Furthermore, (5.13), (8.2) and (8.4) show that,

$$0 < \liminf_{n\to\infty} \operatorname{Var}\Phi(\mathcal{T}_n)/n = \liminf_{n\to\infty}\bigl(\chi^2 + H^{-1}\psi_{\mathsf{V}}(\log n) - \bigl(\chi + H^{-1}\psi_{\mathsf{C}}(\log n)\bigr)^2\bigr)$$
$$= \inf_{x\in\mathbb{R}}\bigl(\chi^2 + H^{-1}\psi_{\mathsf{V}}(x) - (\chi + H^{-1}\psi_{\mathsf{C}}(x))^2\bigr), \quad (8.5)$$

where the final equality holds because this function of $x$ is continuous and periodic (and constant if $d_{\mathbf{p}} = 0$). Hence, also

$$\inf_{\lambda>0} b(\lambda)/\lambda = \inf_{x\in\mathbb{R}}\bigl(\chi^2 + H^{-1}\psi_{\mathsf{V}}(x)\bigr) > 0. \quad (8.6)$$

Consequently, (5.8) holds and (5.10) follows, which is (3.34), with all [absolute] moments.

For (v), (3.36) follows from (3.10), and then (3.37) follows by (5.12). □

*Proof of Theorem 3.12.* (i): It follows from Theorem 3.9, more precisely (3.29)–(3.30) together with (iii), that

$$\frac{\Phi(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} \overset{\mathrm{p}}{\longrightarrow} 0, \quad (8.7)$$

$$\frac{\Phi(\mathcal{T}_n) - \mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_n)}{n} \overset{\mathrm{p}}{\longrightarrow} 0. \quad (8.8)$$

The result (3.38)–(3.39) now follows from (3.10) in Theorem 3.1.

(ii): In this case, Theorem 3.1(iii) applies and shows that $\inf_t\bigl(H^{-1}\psi_{\mathsf{E}}(t) + \chi\bigr) > 0$, and thus that $\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda) > 2c\lambda$ for some $c > 0$ and all large $\lambda$. Hence, (8.8) implies (3.41). □

*Proof of Lemma 3.14.* Let $m \ge 1$ be such that $\operatorname{Var}\Phi(\mathcal{T}_m) > 0$, i.e., $\Phi(\mathcal{T}_m)$ is not deterministic. Let $b := \max\{n_0, m\}$.

We show first that $\operatorname{Var}\Phi(\mathcal{T}_b) > 0$. This is clear if $b = m$, so suppose that $b > m$. Let $\alpha$ and $\beta$ be two distinct letters in $\mathcal{A}$. Condition on the event $\mathcal{E}_{m,b}$ that the strings $\Xi^{(1)}, \dots, \Xi^{(m)}$ begin with $\alpha$, and $\Xi^{(m+1)}, \dots, \Xi^{(b)}$ begin with $\beta$. Then the root $\boldsymbol{\epsilon}$ of $\mathcal{T}_b$ has two children $\alpha$ and $\beta$, with $m$ and $b - m$ strings passed to them, respectively. By assumption, $\varphi(\mathcal{T}_b) = a_b$, and thus (2.18) yields

$$\Phi(\mathcal{T}_b) = a_b + \Phi(\mathcal{T}_b^{\alpha}) + \Phi(\mathcal{T}_b^{\beta}), \quad (8.9)$$

where (still conditioned on $\mathcal{E}_{m,b}$) $\Phi(\mathcal{T}_b^{\alpha})$ and $\Phi(\mathcal{T}_b^{\beta})$ are independent. Furthermore $\Phi(\mathcal{T}_b^{\alpha}) \overset{\mathrm{d}}{=} \Phi(\mathcal{T}_m)$, which is not deterministic; hence (8.9) shows that $\Phi(\mathcal{T}_b)$ conditioned on $\mathcal{E}_{m,b}$ is not deterministic. Thus $\Phi(\mathcal{T}_b)$ (unconditioned) is not deterministic, and $\operatorname{Var}\Phi(\mathcal{T}_b) > 0$ in this case too.

Consider the bucket trie $\mathcal{T}'_n$ with bucket size $b$ grown from the $n$ strings $\Xi^{(1)}, \dots, \Xi^{(n)}$. Then $\mathcal{T}'_n$ is a subtree of $\mathcal{T}_n$. Let $M_k$ be the number of buckets in $\mathcal{T}'_n$ that contain $k$ strings, $k = 1, \dots, b$. Recall that $\mathcal{T}_n$ may be obtained from the bucket trie $\mathcal{T}'_n$ by growing a small trie from every bucket; denote these small tries by $T_{ki}$, where $k = 1, \dots, b$ and $i = 1, \dots, M_k$, so $|T_{ki}|_{\mathsf{e}} = k$. Recall also that conditioned on $\mathcal{T}'_n$, all these small tries are independent, and that $T_{ki}$ is a copy of $\mathcal{T}_k$.

Let $I(\mathcal{T}'_n)$ denote the set of internal nodes of the bucket trie $\mathcal{T}'_n$. The, (2.16) implies the decomposition

$$\Phi(\mathcal{T}_n) = \sum_{v \in I(\mathcal{T}'_n)} \varphi(\mathcal{T}_n^v) + \sum_{k=1}^{b} \sum_{i=1}^{M_k} \Phi(T_{ki}). \tag{8.10}$$

By the construction of the bucket trie, $|\mathcal{T}_n^v|_{\mathsf{e}} > b$ for every $v \in I(\mathcal{T}'_n)$, and thus, by assumption, $\varphi(\mathcal{T}_n^v) = a_{|\mathcal{T}_n^v|_{\mathsf{e}}}$. Consequently, the first sum in (8.10) depends on the bucket trie $\mathcal{T}'_n$ but not on the small tries $T_{ki}$. Consequently, conditioning on the bucket trie,

$$\operatorname{Var}\bigl(\Phi(\mathcal{T}_n) \mid \mathcal{T}'_n\bigr) = \sum_{k=1}^{b} \sum_{i=1}^{M_k} \operatorname{Var}[\Phi(T_{ki})] = \sum_{k=1}^{b} M_k \operatorname{Var}[\Phi(\mathcal{T}_k)]. \tag{8.11}$$

Hence,

$$\begin{aligned}\operatorname{Var}(\Phi(\mathcal{T}_n)) \geq \mathbb{E}\operatorname{Var}\bigl(\Phi(\mathcal{T}_n) \mid \mathcal{T}'_n\bigr) &= \sum_{k=1}^{b} \mathbb{E}\, M_k \cdot \operatorname{Var}[\Phi(\mathcal{T}_k)] \\ &\geq \mathbb{E}\, M_b \cdot \operatorname{Var}[\Phi(\mathcal{T}_b)]. \end{aligned} \tag{8.12}$$

We have shown that $\operatorname{Var}[\Phi(\mathcal{T}_b)] > 0$, and it remains only to show that $\mathbb{E}\, M_b = \Omega(n)$, i.e., $\liminf_{n\to\infty} \mathbb{E}\, M_b/n > 0$.

It is easily seen that $M_b$ a.s. equals the number of nodes in $\mathcal{T}_n$ that have exactly $b$ strings passed to them and have more than one child. Hence, $M_b = \Phi_{b*}(\mathcal{T}_n)$ where the additive functional $\Phi_{b*}$ has toll function $\varphi_{b*}(T)$, defined as the indicator that $|T|_{\mathsf{e}} = b$ and that the root of $T$ has more than one child. If we add a new string to a trie $T$, then $\Phi_{b*}(T)$ may decrease by at most 1, since $\varphi(T^v)$ can be affected only for $v$ in the path from the root to the new leaf or pair of leaves, and in this path there is at most one node with $\varphi(T^v) \neq 0$. It follows that if $\varphi_\bullet(T) := \mathbf{1}\{T = \bullet\}$ as in Example 2.1, so $\Phi_\bullet(T) = |T|_{\mathsf{e}}$, then $\Phi_{b*} + \Phi_\bullet$ is an increasing functional. Hence, Theorem 3.12 applies to $\varphi_{b*}$, with $\varphi_+ := \varphi_{b*} + \varphi_\bullet$ and $\varphi_- := \varphi_\bullet$. Thus, (3.41) holds, which implies $\mathbb{E}\, M_b = \mathbb{E}\, \Phi_{b*}(\mathcal{T}_n) \geq \frac{1}{2}cn$ for large $n$; this completes the proof. □

*Proof of Lemma 3.15.* Let again $X_{\boldsymbol{\alpha}} := \varphi(\widetilde{\mathcal{T}}_\lambda^{\boldsymbol{\alpha}})$, see (7.11). Since $|\varphi(T)| \leq C$, we have, using (7.3),

$$\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda), X_{\boldsymbol{\alpha}}\bigr)\bigr| \leq \mathbb{E}\,|C X_{\boldsymbol{\alpha}}| + C\, \mathbb{E}\,|X_{\boldsymbol{\alpha}}| \leq C_\lambda P(\boldsymbol{\alpha})^2. \tag{8.13}$$

Hence, (2.7) implies that the sum in (3.43) converges absolutely for every $\lambda$. Furthermore, again by (7.3), $\Phi(\widetilde{\mathcal{T}}_\lambda) = \sum_{\boldsymbol{\alpha}} X_{\boldsymbol{\alpha}}$ with convergence in $L^1$, and thus, since $\varphi(\widetilde{\mathcal{T}}_\lambda)$ is bounded,

$$\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),\Phi(\widetilde{\mathcal{T}}_\lambda)\bigr)=\sum_{\boldsymbol{\alpha}}\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr), \tag{8.14}$$

which shows the equality of the expressions in (3.17) and (3.43).

By Lemma 6.1(ii) (with $m=2$), $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}_\lambda)|\le C\lambda^2$ for $\lambda\le 1$. Hence, if $\lambda\le 1$, then for every $\alpha$, recalling (7.11), $\mathbb{E}\,|\varphi(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)|\le C\lambda^2P(\boldsymbol{\alpha})^2$ for every $\boldsymbol{\alpha}\in\mathcal{A}^*$, and thus, arguing as in (8.13),

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\bigr|\le\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}C\,\mathbb{E}\,|X_{\boldsymbol{\alpha}}|\le\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}C\lambda^2P(\boldsymbol{\alpha})^2=C\lambda^2. \tag{8.15}$$

For $\lambda\ge 1$, we use instead the decomposition and notation in the proof of Lemma 7.4. Let $\varepsilon_{\boldsymbol{\alpha}}:=\operatorname{sign}\bigl(\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\bigr)\in\{\pm1\}$. Then, for each $k\ge 0$,

$$\begin{aligned}\sum_{P(\boldsymbol{\alpha})\in J_k}\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\bigr|&=\sum_{P(\boldsymbol{\alpha})\in J_k}\varepsilon_{\boldsymbol{\alpha}}\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\\&=\operatorname{Cov}\Bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),\sum_{P(\boldsymbol{\alpha})\in J_k}\varepsilon_{\boldsymbol{\alpha}}X_{\boldsymbol{\alpha}}\Bigr).\end{aligned} \tag{8.16}$$

Furthermore, the variables $X_{\boldsymbol{\alpha}}=\varphi(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}+}_\lambda)$ are independent for $\boldsymbol{\alpha}\in J_k$, and thus, see (7.31), (7.33) and (7.35)–(7.36),

$$\begin{aligned}\operatorname{Var}\Bigl(\sum_{P(\boldsymbol{\alpha})\in J_k}\varepsilon_{\boldsymbol{\alpha}}X_{\boldsymbol{\alpha}}\Bigr)&=\sum_{P(\boldsymbol{\alpha})\in J_k}\operatorname{Var}X_{\boldsymbol{\alpha}}=\operatorname{Var}Z_{\lambda,k}\\&\le C\lambda\min\bigl((\lambda p_{\max}^k)^{-\varepsilon},\lambda p_{\max}^k\bigr).\end{aligned} \tag{8.17}$$

Hence, by (8.16) and the Cauchy–Schwarz inequality, recalling that $\varphi$ is bounded,

$$\sum_{P(\boldsymbol{\alpha})\in J_k}\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\bigr|\le C\lambda^{1/2}\min\bigl((\lambda p_{\max}^k)^{-\varepsilon/2},(\lambda p_{\max}^k)^{1/2}\bigr). \tag{8.18}$$

We may now sum over $k\ge 0$ and obtain, since $p_{\max}<1$,

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}^*}\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),X_{\boldsymbol{\alpha}}\bigr)\bigr|\le C\lambda^{1/2}. \tag{8.19}$$

The two estimates (8.15) for $\lambda\le 1$ and (8.19) for $\lambda\ge 1$ imply the same estimates (with a different $C$) for $\sum^*_{\boldsymbol{\alpha}}$, and it follows that, for $\sigma:=\operatorname{Re}s\in(-2,-\frac12)$,

$$\int_0^\infty\sum_{\boldsymbol{\alpha}}^*\bigl|\operatorname{Cov}\bigl(\varphi(\widetilde{\mathcal{T}}_\lambda),\varphi(\widetilde{\mathcal{T}}^{\boldsymbol{\alpha}}_\lambda)\bigr)\lambda^{s-1}\bigr|\,\mathrm{d}\lambda\le\int_0^\infty C\min\{\lambda^{\sigma+1},\lambda^{\sigma-1/2}\}\,\mathrm{d}\lambda<\infty. \tag{8.20}$$

Hence, Fubini's theorem shows that we may interchange the sum and integral in (3.44). Thus (3.44) follows by (3.43). □

*Proof of Lemma 3.16.* By replacing $\varphi$ by $\varphi_*$ as in the proof of Theorem 3.1, we may again assume $\chi = 0$; recall (7.45) and (7.51). Furthermore, by considering the positive and negative parts of $\varphi$ separately, we may also assume $\varphi \geq 0$. Then, by (3.16) and (6.1), we have, with $a_n := \mathbb{E}\,\varphi(\mathcal{T}_n)$,

$$f_{\mathsf{E}}(\lambda) = e^{-\lambda} \sum_{n=2}^{\infty} \frac{a_n}{n!} \lambda^n. \tag{8.21}$$

Hence, at least for $-2 < \operatorname{Re} s < -1 + \varepsilon$, first for real $s$ and then generally,

$$f_{\mathsf{E}}^*(s) = \int_0^\infty f_{\mathsf{E}}(\lambda) \lambda^{s-1} \,\mathrm{d}\lambda = \sum_{n=2}^{\infty} \frac{a_n}{n!} \int_0^\infty \lambda^{n+s-1} e^{-\lambda} \,\mathrm{d}\lambda = \sum_{n=2}^{\infty} \frac{a_n}{n!} \Gamma(n+s), \tag{8.22}$$

yielding (3.45). Taking $s = -1$ yields (3.46), recalling $f_{\mathsf{C}}^*(-1) = f_{\mathsf{E}}^*(-1)$ from Lemma 3.6. □

**Acknowledgement.** I thank Pawel Hitczenko for help with references on Rosenthal's inequality.

## Appendix A. Asymptotics of certain sums

The following theorem is essentially [18, Theorem 5.1], with some extensions as discussed in the proof. Recall the definition of the entropy $H$ in (2.2), the greatest common divisor $d_{\mathbf{p}} := \gcd\{-\log p_\alpha : \alpha \in \mathcal{A}\}$ in Section 2.10, and the Mellin transform $f^*$ in (2.27).

**Theorem A.1.** *Suppose that $f$ is a real-valued function on $(0, \infty)$, and that*

$$F(\lambda) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^*} f(\lambda P(\boldsymbol{\alpha})), \tag{A.1}$$

*for $\lambda > 0$, with $P(\boldsymbol{\alpha})$ given by* (2.3)*. Assume further that $f$ is a.e. continuous and satisfies the estimates*

$$f(x) = O(x^2), \qquad 0 < x < 1, \tag{A.2}$$

$$f(x) = O(x^{1-\varepsilon}), \qquad 1 < x < \infty, \tag{A.3}$$

*for some $\varepsilon > 0$.*

(i) *If $d_{\mathbf{p}} = 0$, then, as $\lambda \to \infty$,*

$$\frac{F(\lambda)}{\lambda} \to \frac{1}{H} f^*(-1) = \frac{1}{H} \int_0^\infty f(x) x^{-2} \,\mathrm{d}x. \tag{A.4}$$

(ii) *More generally, for any $d_{\mathbf{p}}$, as $\lambda \to \infty$,*

$$\frac{F(\lambda)}{\lambda} = \frac{1}{H} \psi(\log \lambda) + o(1), \tag{A.5}$$

*where $\psi$ is a bounded function defined as follows:*

(a) *If $d_{\mathbf{p}} = 0$ then $\psi$ is constant: for all $t$,*

$$\psi(t) := f^*(-1). \tag{A.6}$$

(b) *If $d = d_{\mathbf{p}} > 0$, then $\psi$ is a bounded $d$-periodic function having the Fourier series*

$$\psi(t) \sim \sum_{m=-\infty}^{\infty} \widehat{\psi}(m) e^{2\pi \mathrm{i} mt/d} \tag{A.7}$$

*with*

$$\widehat{\psi}(m) = f^*\Bigl(-1 - \frac{2\pi m}{d}\mathrm{i}\Bigr) = \int_0^\infty f(x) x^{-2-2\pi \mathrm{i} m/d}\,\mathrm{d}x. \tag{A.8}$$

*Furthermore,*

$$\psi(t) = d \sum_{k=-\infty}^{\infty} e^{kd-t} f\bigl(e^{t-kd}\bigr). \tag{A.9}$$

(iii) *If $f$ is continuous, then $\psi$ is too.*

(iv) *If $f$ is continuous and $f > 0$ on $(0,\infty)$, then $\inf_t \psi(t) > 0$. Hence, $\psi(t) = \Theta(1)$ and, as $\lambda \to \infty$, $F(\lambda) = \Theta(\lambda)$.*

(v) *Suppose that $d_{\mathbf{p}} > 0$. If $f$ is continuous and the Fourier series* (A.7) *converges absolutely, then its sum equals $\psi(t)$ everywhere, so we may replace $\sim$ in* (A.7) *by $=$. In particular, this holds if $f$ is continuously differentiable on $(0,\infty)$ and $f'(x) = O(x^\varepsilon)$ as $x \to 0$ and $f'(x) = O(x^{-\varepsilon})$ as $x \to \infty$ for some $\varepsilon > 0$.*

*Proof.* (i),(ii),(iii): This is, as said above, essentially [18, Theorem 5.1]. There are three technical differences:

(a) [18] considers for simplicity only $\mathcal{A} = \{0,1\}$. However, the same proof holds for arbitrary $\mathcal{A}$.

(b) [18, Theorem 5.1] assumes that $f \geq 0$. This is technically convenient in the proof (e.g., all sums and integrals are defined), but the result extends immediately to real-valued $f$ by considering its positive and negative parts.

(c) [18, Theorem 5.1] as stated there assumes (A.3) with $\varepsilon = 1$, but as said in [18, Remark 5.2], the proof holds for any $\varepsilon > 0$. (We may similarly relax (A.2) to $f(x) = O(x^{1+\varepsilon})$ for $0 < x < 1$, but we have no use for this.)

With these extensions, [18, Theorem 5.1] yields (i)–(iii). (Note that, with $g(t) = e^f f(e^{-t})$ as in [18], $\widehat{g}(u) = f^*(-1+u\mathrm{i})$. Also, (iii) is trivial if $d_{\mathbf{p}} = 0$.)

(iv): This too is trivial if $d_{\mathbf{p}} = 0$ by (A.6) and (A.4), since the integral in (A.4) now is positive.

Thus, suppose $d_{\mathbf{p}} > 0$. Then $\psi(t) > 0$ for every real $t$ by (A.9). Since $\psi$ is periodic by (ii) and continuous by (iii), it follows that $\inf_t \psi(t) > 0$. Hence, $F(\lambda) = \Theta(\lambda)$ as $\lambda \to \infty$ by (A.5).

(v): $\psi$ is continuous by (iii) and has by assumption a Fourier series that converges everywhere, which implies that the Fourier series converges to $\psi$, see e.g. [34, III.(3.4) and applications after it].

If $f$ is continuously differentiable on $(0, \infty)$ and $f'(x) = O(x^{\varepsilon})$ as $x \to 0$ and $f'(x) = O(x^{-\varepsilon})$ as $x \to \infty$, then (A.9) can be differentiated termwise and the resulting sum converges uniformly on compact sets. Hence $\psi$ has a continuous derivative, and is, in particular, Lipschitz on $[0, d]$, and thus its Fourier series converges absolutely by a theorem by Bernstein [34, Theorem VI.(3.1)]. ☐

**Remark A.2.** It follows from (A.2)–(A.3) that the Mellin transform $f^*(s)$ exists at least when $-2 < \operatorname{Re} s < -1 + \varepsilon$, and thus in particular when $\operatorname{Re} s = -1$ as in (A.4) and (A.8). ☐

## Appendix B. Approximation in distribution and moments

We prove here the following lemma, which includes Lemma 2.6 together with a converse. It extends the standard result that if $X_n \xrightarrow{\mathrm{d}} Y$, then convergence of absolute moments of order $s$ is equivalent to uniform integrability of $|X_n|^s$, and implies convergence of moments, see e.g. [11, Theorem 5.5.9]. Recall the definitions in Section 2.12.

**Lemma B.1.** *Let $(X_n)_1^{\infty}$ and $(Y_n)_1^{\infty}$ be random vectors in $\mathbb{R}^d$. Let further $s > 0$ be a real number, and suppose that the sequence $(|Y_n|^s)$ is uniformly integrable. Then the following are equivalent:*

(i) *$X_n \overset{\mathrm{d}}{\approx} Y_n$ with absolute moments of order $s$.*

(ii) *$X_n \overset{\mathrm{d}}{\approx} Y_n$ and the sequence $(|X_n|^s)$ is uniformly integrable.*

*Furthermore, if* (i) *or* (ii) *holds (and thus both hold), and $s$ is an integer, then also*

(iii) *$X_n \overset{\mathrm{d}}{\approx} Y_n$ with moments of order $s$.*

*Moreover, if* (i) *or* (ii) *holds, then they also hold with $s$ replaced by any positive $s' < s$, and* (iii) *holds with $s$ replaced by any smaller positive integer.*

*Proof.* (ii) $\Longrightarrow$ (i),(iii). The assumptions that $|Y_n|^s$ and $|X_n|^s$ are uniformly integrable imply that the sequences $(X_n)$ and $(Y_n)$ are tight. Hence, for any subsequence $(n_j)$, there exists a subsubsequence along which $X_n \xrightarrow{\mathrm{d}} X$ and $Y_n \xrightarrow{\mathrm{d}} Y$ for some random variables $X$ and $Y$. The assumption $X_n \overset{\mathrm{d}}{\approx} Y_n$ implies $X \overset{\mathrm{d}}{=} Y$. Furthermore, the uniform integrability and convergence in distribution imply that, still along the subsubsequence, $\mathbb{E}\,|X_n|^s \to \mathbb{E}\,|X|^s$ and $\mathbb{E}\,|Y_n|^s \to \mathbb{E}\,|Y|^s$, see [11, Theorem 5.5.9]. Since $\mathbb{E}\,|X|^s = \mathbb{E}\,|Y|^s$, it follows that (2.31) holds along the subsubsequence. By the subsequence principle, (2.31) holds for the full sequence, and thus (i) holds.

If $s$ is an integer, then (2.30) follows by the same argument.

(i) $\Longrightarrow$ (ii). This is similar. The uniform integrability of $|Y_n|^s$ implies that the sequence $Y_n$ is tight. Hence, for any subsequence $(n_j)$, there exists a subsubsequence along which $Y_n \xrightarrow{\mathrm{d}} Y$ for some random variable $Y$. The assumption $X_n \overset{\mathrm{d}}{\approx} Y_n$ implies that also $X_n \xrightarrow{\mathrm{d}} Y$ along the subsubsequence.

Furthermore, still along the subsubsequence, the assumption on uniform integrability implies $\mathbb{E}|Y_n|^s \to \mathbb{E}|Y|^s$, and since we assume (i) and thus (2.31), we have $\mathbb{E}|X_n|^s \to \mathbb{E}|Y|^s$. Hence, by [11, Theorem 5.5.9] again, $|X_n|^s$ is uniformly integrable along the subsubsequence. The subsequence principle holds also for uniform integrability, and thus (ii) holds. (To see this, assume that the full sequence $|X_n|^s$ is not uniformly integrable; then there exists an $\varepsilon > 0$ and a subsequence $n_j$ such that $\mathbb{E}\big[|X_{n_j}|^s \mathbf{1}\{|X_{n_j}|^s > j\}\big] > \varepsilon$, but then no subsubsequence is uniformly integrable, a contradiction.)

Finally, it is well known that if (ii) holds for some $s$, it holds for all smaller $s'$ as well. ☐

**Remark B.2.** The standard case with $X_n \xrightarrow{\mathrm{d}} Y$ is the case when all $Y_n$ are equal; in that case uniformly integrable of $|Y_n|^s$ is redundant. In general, however, it is needed. Here are some counterexamples without uniform integrability.

(i) Let $X_n = Y_n = n$ (non-random); then (i) holds trivially but not (ii).
(ii) Let $\mathbb{P}(X_n = -n) = 1 - \mathbb{P}(X_n = n) = 1/n$ and $Y_n = n$. Then $|X_n| = |Y_n|$ so (2.31) is trivial and (i) holds for any $s$, but (ii) fails for all $s$ and (iii) for odd integers $s$.
(iii) Let $\mathbb{P}(X_n = n^2) = 1 - \mathbb{P}(X_n = 0) = 1/n$ and $\mathbb{P}(Y_n = n^3) = 1 - \mathbb{P}(Y_n = 0) = 1/n^3$. Then $X_n \xrightarrow{\mathrm{d}} 0$ and $Y_n \xrightarrow{\mathrm{d}} 0$ and thus $X_n \overset{\mathrm{d}}{\approx} Y_n$. Furthermore, (i) holds for $s = 2$ but not for $s = 1$.

Such examples indicate that moment approximation as in (2.29)–(2.31) is not of much interest unless $|Y_n|^s$ are uniformly integrable. ☐

**Remark B.3.** If $s$ is an even integer, then $|X_n|^s = X_n^s$ when $d = 1$, and in general $|X_n|^s$ is a linear combination of moments $\mathbb{E}\, X_n^{\mathbf{m}}$ with $|\mathbf{m}| = s$. Hence, (2.30) implies (2.31), which together with Lemma B.1 shows that, assuming that $|Y_n|^s$ are uniformly integrable, $X_n \overset{\mathrm{d}}{\approx} Y_n$ with absolute moments of order $s$ is equivalent to $X_n \overset{\mathrm{d}}{\approx} Y_n$ with moments of order $s$. Again, the condition of uniform integrability is needed here, as shown by Remark B.2(ii). ☐

## Appendix C. A lower bound on the approximation error

The rate of convergence of the asymptotic results (3.7) and (3.10) has been studied in detail by Flajolet, Roux and Vallée [8]. They focussed on the aperiodic case $d_{\mathbf{p}} = 0$ and gave upper bounds for the error, with a very slow rate of convergence. It is implicit in their arguments that their bounds are essentially the best possible, and we state one version of that as Theorem C.2 below. For simplicity we consider there, and in most part of this appendix, only the special case of the size, as in Section 4.1, although the results are more general and the method applies also to other examples in Section 4, cf. [8, Definition 4 and Lemma 6].

However, we first review the much simpler periodic case. This case is well-known, see e.g. [16, Section 7.2], and included her for completeness and comparison.

C.1. **The symmetric case.** Consider the case $p_\alpha = 1/r$ for every $\alpha \in \mathcal{A}$, where $r := |\mathcal{A}|$. Note that $d_{\mathbf{p}} = H = \log r$. Let $\varphi$ and $f_{\mathsf{E}}$ be as in Theorem 3.1 and assume for simplicity $\chi = 0$. Then, by (7.38),

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \sum_{m=0}^{\infty} \lambda^{-1} r^m f_{\mathsf{E}}\bigl(r^{-m}\lambda\bigr) = \sum_{m=-\infty}^{\infty} \frac{f_{\mathsf{E}}\bigl(r^{-m}\lambda\bigr)}{r^{-m}\lambda} + R(\lambda), \tag{C.1}$$

where

$$R(\lambda) = -\sum_{k=1}^{\infty} \frac{f_{\mathsf{E}}\bigl(r^{k}\lambda\bigr)}{r^{k}\lambda}. \tag{C.2}$$

The assumption (3.1) implies that the sum (C.2) converges, and

$$R(\lambda) = O\bigl(\lambda^{-\varepsilon}\bigr). \tag{C.3}$$

Furthermore, the last sum in (C.1) equals $H^{-1}\psi_{\mathsf{E}}(\log\lambda)$, see (3.15). Hence, the elementary calculation (C.1) yields (3.10) with the error term $R(\lambda) = O(\lambda^{-\varepsilon})$. In fact, in several examples in Section 4, $f_{\mathsf{E}}(\lambda)$ decreases exponentially as $\lambda \to \infty$, and then the same holds for $R(\lambda)$.

In the special case of the size, treated in Section 4.1, (4.3) yields

$$R(\lambda) = -\sum_{k=1}^{\infty} \frac{1-(1+r^k\lambda)e^{-r^k\lambda}}{r^k\lambda} = -\frac{1}{r-1}\lambda^{-1} + O\bigl(e^{-r\lambda}\bigr) \tag{C.4}$$

as $\lambda \to \infty$. Thus the error rate (C.3) (with $\varepsilon = 1$) is exact in this case.

C.2. **The asymmetric periodic case.** Suppose $d = d_{\mathbf{p}} > 0$; then there exist positive integers $\kappa_\alpha$ such that $-\log p_\alpha = \kappa_\alpha d$ and thus $p_\alpha = e^{-\kappa_\alpha d}$, $\alpha \in \mathcal{A}$. Hence,

$$\rho(s) = \sum_{\alpha\in\mathcal{A}} p_\alpha^s = \sum_{\alpha\in\mathcal{A}} e^{-\kappa_\alpha d s} = Q\bigl(e^{-ds}\bigr), \tag{C.5}$$

where $Q$ is the polynomial $Q(z) = \sum_{\alpha\in\mathcal{A}} z^{\kappa_\alpha}$ of degree $\kappa := \max\kappa_\alpha$. We exclude the symmetric case in Section C.1; then $\kappa \geq 2$. Denote the roots of $Q(w) = 1$ by $w_1, \dots, w_\kappa$, with $w_1 = e^{-d} > 0$. Consider again $\Phi(T) := |T|_{\mathsf{i}}$, the size. A standard inversion of the Mellin transform yields, see e.g. [8, Lemma 6] (although there stated for the aperiodic case), or [16, Section 7.2],

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \sum_{z_j} a_j \lambda^{z_j-1} + O\bigl(\lambda^{-M}\bigr) \tag{C.6}$$

for any $M < \infty$, where $a_j$ are some complex numbers and $z_j$ ranges over the roots of $\rho(z_j) = 1$. By (C.5), these roots are (changing the notation)

$$z_{k,\ell} := \frac{-\log w_k}{d} + \frac{2\pi \mathrm{i}}{d}\ell, \qquad k \in \{1, \dots, \kappa\},\ \ell \in \mathbb{Z}. \tag{C.7}$$

The roots $z_{1,\ell}$ have $\operatorname{Re} z_{1,\ell} = 1$, and the corresponding terms in (C.6) yield the periodic function $H^{-1}\psi_{\mathsf{E}}(\log \lambda)$ in (3.10). Similarly, the terms for $z_{k,\ell}$ for a fixed $k > 1$ sum to $\lambda^{-\varepsilon_k} g_k(\log \lambda)$, where $g_k$ is a $d$-periodic function and

$$\varepsilon_k := 1 - \operatorname{Re} z_{k,0} = 1 + \log |w_k|/d, \tag{C.8}$$

which necessarily satisfies $\varepsilon_k > 0$. Hence, if $\varepsilon := \min_{2 \le k \le \kappa} \varepsilon_k > 0$, then (C.6) yields (3.10) with an error term

$$R(\lambda) = O\bigl(\lambda^{-\varepsilon}\bigr). \tag{C.9}$$

Furthermore, this is the exact order of the error term (for typical $\lambda$). Here, $\varepsilon > 0$ depends on the probabilities $(p_\alpha)_\alpha$ and may be arbitrarily small, even in the binary case. This too is certainly known, but we do not know a reference and give an example for completeness.

**Example C.1.** Consider the binary case, with $\mathbf{p} = (p, 1-p)$. In the periodic case $d_{\mathbf{p}} > 0$, denote $\varepsilon$ above by $\varepsilon(p)$; in the aperiodic case let $\varepsilon(p) = 0$. Let $\delta > 0$.

Let $p_0$ be any number with $\log(1-p_0)/\log p_0$ irrational. Then there exist roots $s$ of $\rho(s) = p_0^s + (1-p_0)^s = 1$ with $1 - \delta < \operatorname{Re} s < 1$, see e.g. [8], or the proof of Theorem C.2 below. Let $s_0$ be one such root. It follows from the implicit function theorem that for every $p$ sufficiently close to $p_0$, there exists $s$ with $p^s + (1-p)^s = 1$ and $s$ so close to $s_0$ that $\operatorname{Re} s \in (1-\delta, 1)$. Hence, $\varepsilon(p) \le 1 - \operatorname{Re} s < \delta$. We may here choose $p$ such that $\log(1-p)/\log p$ is rational. Consequently, the set of $p$ such that $d_{\mathbf{p}} > 0$ and $\varepsilon(p) < \delta$ is dense in $(0,1)$ for every $\delta > 0$.

For a concrete example, let $m \ge 1$ and let $p = p_m$ be the unique positive root of $p + p^{(m+1)/m} = 1$. It is not difficult to show that as $m \to \infty$, $p_m \to \frac{1}{2}$ and $\varepsilon(p_m) \to 0$. We omit the details. $\square$

C.3. **The aperiodic case.** In the aperiodic case $d_{\mathbf{p}} = 0$ of Theorem 3.1, (3.7) (or (3.10)) says that

$$\frac{\mathbb{E}\,\Phi(\widetilde{\mathcal{T}}_\lambda)}{\lambda} = \chi + H^{-1} f_{\mathsf{E}}^*(-1) + R(\lambda), \tag{C.10}$$

where $R(\lambda) = o(1)$ as $\lambda \to \infty$. As said above, it follows from Flajolet, Roux and Vallée [8] that $R(\lambda)$ typically tends to 0 very slowly. More precisely, we have the following, for simplicity considering only the size.

**Theorem C.2** (implicit in [8])**.** *Assume $d_{\mathbf{p}} = 0$ and let $\Phi(T) = |T|_{\mathsf{i}}$, the size of $T$. Then* (C.10) *holds (with $\chi = 0$ and $f_{\mathsf{E}}^*(-1) = 1$) and there exist $C < \infty$ and arbitrarily large $\lambda$ such that*

$$|R(\lambda)| > \exp\bigl(-C(\log \lambda)^{(|\mathcal{A}|-1)/(|\mathcal{A}|+1)}\bigr). \tag{C.11}$$

Note that the lower bound in (C.11) is larger than $\lambda^{-\varepsilon}$ for any $\varepsilon > 0$ (and large $\lambda$). Cf. the results in the periodic cases above.

Flajolet, Roux and Vallée [8] prove corresponding upper bounds, for most but not all probability vectors $(p_\alpha)$; see in particular [8, Theorem 4 and Corollaries 1 and 2]. As said above, the lower bound in Theorem C.2 is only implicit in [8]; a detailed proof seems to require some work, and since we do not know any published proof, we give one below for completeness.

**Remark C.3.** It follows further from [8] that for special vectors $(p_\alpha)_{\alpha\in\mathcal{A}}$, even larger lower bounds hold; in fact, by considering the binary case with $\log p_0/\log p_1$ an irrational number that can be approximated extremely well by rationals (a suitable Liouville number), we can make $R(\lambda)$ converge arbitrarily slowly to 0. □

We first prove a lemma.

**Lemma C.4.** *Suppose that*

$$h(x) = \sum_{j=-\infty}^{\infty} a_j e^{\zeta_j x}, \tag{C.12}$$

*where the complex numbers* $a_j$ *and* $\zeta_j = -s_j + \mathrm{i}t_j$ *satisfy the following, for some* $c > 0$ *and all* $j \in \mathbb{Z}$*,*

$$s_j > 0, \tag{C.13}$$

$$t_{j+1} - t_j \geq c, \tag{C.14}$$

$$|a_j| = e^{-\Theta(|t_j|)+O(1)}. \tag{C.15}$$

*Then there exists* $\delta > 0$ *and* $C$ *such that for every* $j$ *with* $s_j < \delta$ *and* $t_j > 0$*, there exists* $x$ *with* $t_j/(2s_j) < x < 2t_j/s_j$ *and* $|h(x)| > e^{-Ct_j}$*.*

*The same result holds for a one-sided sum* $h(x) = \sum_{j=0}^{\infty} a_j e^{\zeta_j x}$*.*

*Proof.* In this proof $C_i$ and $c_i$ denote positive constants that depend only on $c$ in (C.14) and the implicit constants in (C.15). Note first that (C.15) and (C.14) imply

$$\sum_k |a_k| \leq C_1. \tag{C.16}$$

In particular, the sum (C.12) converges for every $x \geq 0$.

We assume first that $s_j \leq \delta$ for every $j$, and treat then the general case.

*Case 1:* $\max_j s_j \leq \delta \leq 1$. Consider a $j$ with $t_j \geq 1$. Let $Z \sim N(0,1)$ be a standard normal random variable and define

$$\mu_j := t_j/s_j \geq t_j/\delta \geq 1/\delta, \tag{C.17}$$

$$h_j(x) := e^{-\zeta_j x} h(x) = \sum_k a_k e^{(\zeta_k - \zeta_j)x}, \tag{C.18}$$

$$Z_j := \mu_j + 2\mu_j^{1/2} Z \sim N\bigl(\mu_j, 4\mu_j\bigr). \tag{C.19}$$

Note that the assumption $s_k \in [0, \delta]$ and (C.16) imply that the sums (C.12) and (C.18) converge for every real $x$, with

$$|h_j(x)| \le \sum_k |a_k| e^{(s_j - s_k)x} \le C_1 e^{\delta |x|}. \tag{C.20}$$

In particular, $h_j(x)$ is defined for all real $x$. Furthermore, by (C.18),

$$\begin{aligned}\mathbb{E}\, h_j(Z_j) &= \sum_k a_k \,\mathbb{E}\, e^{(\zeta_k - \zeta_j) Z_j} = \sum_k a_k \,\mathbb{E}\, e^{(\zeta_k - \zeta_j)\mu_j + 2\mu_j (\zeta_k - \zeta_j)^2} \\ &=: \sum_k A_k, \end{aligned} \tag{C.21}$$

where we thus denote the terms in the sum by $A_k$. Note that $A_j = a_j$, so, by (C.15),

$$|A_j| = |a_j| \ge e^{-C_2 t_j}. \tag{C.22}$$

For $k \ne j$, we note that (C.15) implies $|a_k| \le C_3$, and thus, by (C.21) and (C.17),

$$\begin{aligned}|A_k| &= |a_k| e^{(s_j - s_k)\mu_j + 2\mu_j (s_j - s_k)^2 - 2\mu_j (t_j - t_k)^2} \le C_3 e^{\delta\mu_j + 2\mu_j \delta^2 - 2\mu_j |t_j - t_k|^2} \\ &\le e^{C_4 + 3\delta\mu_j - 2\mu_j |t_j - t_k|^2} \le e^{C_5 \delta \mu_j - 2\mu_j |t_j - t_k|^2}. \end{aligned} \tag{C.23}$$

It follows from (C.14) that whenever $j \ne k$, we have $|t_j - t_k| \ge c|j-k| \ge c$. Hence, if $C_6 := c^2/2$ and $\delta$ is small enough, then (C.23) implies

$$|A_k| \le e^{\mu_j (C_5 \delta - 2c^2 |j-k|^2)} \le e^{\mu_j (C_5\delta - c^2 - c^2 |j-k|^2)} \le e^{-C_6 \mu_j - c^2 \mu_j |j-k|^2}. \tag{C.24}$$

Recalling (C.17), we thus find that for $\delta$ small enough,

$$\sum_{k \ne j} |A_k| \le C_7 e^{-C_6 \mu_j} \le C_7 e^{-(C_6/\delta) t_j} \le e^{-2C_2 t_j}. \tag{C.25}$$

Combining (C.25) with (C.22) and (C.21), we find that for $t_j$ sufficiently large,

$$\bigl|\mathbb{E}\, h_j(Z_j)\bigr| \ge |A_j| - \sum_{k \ne j} |A_k| \ge \tfrac12 |a_j| \ge e^{-(C_2+1) t_j}. \tag{C.26}$$

Next, let $I_j := (\frac12 \mu_j, \frac32 \mu_j)$. By the Cauchy–Schwarz inequality, (C.20) and a standard tail estimate for the normal distribution, if $\delta$ is small enough,

$$\begin{aligned}\bigl|\mathbb{E}\bigl[h_j(Z_j) \mathbf{1}\{Z_j \notin I_j\}\bigr]\bigr|^2 &\le \mathbb{E}\bigl[h_j(Z_j)^2\bigr]\, \mathbb{P}\bigl(Z_j \notin I_j\bigr) \\ &\le C_1^2 \bigl(\mathbb{E}\, e^{2\delta Z_j} + \mathbb{E}\, e^{-2\delta Z_j}\bigr)\, \mathbb{P}\bigl(|Z| \ge \tfrac14 \mu_j^{1/2}\bigr) \\ &\le C_8 e^{2\delta\mu_j + 8\delta^2 \mu_j} e^{-\frac{1}{32}\mu_j} \le C_8 e^{-\frac{1}{64}\mu_j} \le C_8 e^{-\frac{1}{64\delta} t_j}. \end{aligned} \tag{C.27}$$

Hence, if $\delta$ is small enough, using $t_j \ge 1$ and (C.26),

$$\bigl|\mathbb{E}\bigl[h_j(Z_j)\mathbf{1}\{Z_j \notin I_j\}\bigr]\bigr| \le e^{-(C_2+2)t_j} \le e^{-1} |\,\mathbb{E}\, h_j(Z_j)|, \tag{C.28}$$

and thus, using (C.26) again,

$$\bigl|\mathbb{E}\bigl[h_j(Z_j)\mathbf{1}\{Z_j \in I_j\}\bigr]\bigr| \ge \tfrac12 |\,\mathbb{E}\, h_j(Z_j)| \ge e^{-C_9 t_j} \tag{C.29}$$

Consequently, there exists $x_j \in I_j$ such that $|h_j(x_j)| \geq e^{-C_9 t_j}$, and hence,

$$|h(x_j)| = e^{-s_j x_j}|h_j(x_j)| \geq e^{-(3/2)t_j}|h_j(x_j)| \geq e^{-C_{10}t_j}. \tag{C.30}$$

We have shown that there exist $\delta_1 > 0$ and $T \geq 1$ such that if $s_k \leq \delta_1$ for every $k$, then the result (C.30) holds for every $j$ with $t_j \geq T$. We ignore temporarily the finite number of $j$ with $0 < t_j < T$.

*Case 2: The general case.* Let $\delta_1$ and $T \geq 1$ be as just said in Case 1. Let $\Lambda := \{\zeta_j\}$ and consider the subsets $\Lambda_< := \{\zeta_j : s_j \leq \delta_1\}$ and $\Lambda_> := \{\zeta_j : s_j > \delta_1\}$. Define the corresponding sums

$$h_<(x) := \sum_{\zeta_j \in \Lambda_<} a_j e^{\zeta_j x}, \qquad h_>(x) := \sum_{\zeta_j \in \Lambda_>} a_j e^{\zeta_j x}. \tag{C.31}$$

We may assume that $\{\zeta_j \in \Lambda_< : t_j > 0\}$ is infinite, since otherwise we may choose $\delta > 0$ such that $s_j > \delta$ for all $j$ with $t_j > 0$, and the result is trivial.

Then Case 1 applies to $h_<$ (after relabelling $\zeta_j$). Hence, if $\delta \leq \delta_1$, for every $j$ with $s_j < \delta$ and $t_j \geq T$, there exists $x_j$ such that $t_j/(2s_j) < x_j < 2t_j/s_j$ and $|h_<(x_j)| > e^{-C_{10}t_j}$. Furthermore, recalling (C.16), if $\delta$ is small enough,

$$\begin{aligned} |h_>(x_j)| &\leq \sum_{\zeta_j \in \Lambda_>} |a_j| e^{-s_j x_j} \leq C_1 e^{-\delta_1 x_j} \\ &\leq C_1 e^{-(\delta_1/2s_j)t_j} \leq C_1 e^{-(\delta_1/2\delta)t_j} \leq \tfrac{1}{2} e^{-C_{10}t_j}, \end{aligned} \tag{C.32}$$

and consequently,

$$|h(x_j)| \geq |h_<(x_j)| - |h_>(x_j)| \geq \tfrac{1}{2} e^{-C_{10}t_j} \geq e^{-(C_{10}+1)t_j}. \tag{C.33}$$

This shows the result for any $j$ with $s_j < \delta$ and $t_j \geq T$. By decreasing $\delta$, we may further assume that $s_j \geq \delta$ for each of the finitely many $j$ with $0 < t_j < T$, and the result then holds for every $j$ with $s_j < \delta$ and $t_j > 0$. □

*Proof of Theorem C.2.* Let $r := |\mathcal{A}|$, the number of letters in the alphabet, and assume without loss of generality that $\mathcal{A} = \{1, \dots, r\}$.

By [8, Lemma 6], $R(\lambda)$ in (C.10) can be written $R(\lambda) = R_1(\lambda) + R_2(\lambda)$ where for some small $\varepsilon > 0$, $R_2(\lambda) = O(\lambda^{-\varepsilon})$ and

$$R_1(\lambda) = \sum_k \frac{(1 - z_k)\Gamma(-z_k)}{\rho'(z_k)} \lambda^{z_k - 1}, \tag{C.34}$$

summing over the set $\{z_k\}$ of roots of $\rho(z) = 1$ satisfying $1 - \varepsilon \leq \operatorname{Re} z_k < 1$. Thus $h(x) := R_1(e^x)$ is a function of the type in (C.12), with $s_j = 1 - \operatorname{Re} z_j$ and $t_j = \operatorname{Im} z_j$, and (C.13)–(C.15) hold by results in [8].

We assume that $d_{\mathbf{p}} = 0$; thus, for any fixed $k \in \mathcal{A}$ at least one ratio $\log p_\ell / \log p_k$ is irrational. By [12, Theorem 200], there exist infinitely many positive integers $q$ such that for some integers $\kappa_\ell$

$$\left| q \frac{\log p_\ell}{\log p_k} - \kappa_\ell \right| < q^{-1/(r-1)}, \qquad \ell = 1, \dots, r. \tag{C.35}$$

(Note that the case $\ell = k$ is trivial, so we really consider a vector of $r - 1$ elements.) In the terminology of [8] the approximation function $f_k(q)$ of the vector $(\log p_\ell/\log p_k)_\ell$ satisfies $f_k(q) \geq q^{1/(r-1)}$ for infinitely many $q$. Let $(q_j)_1^\infty$ be an increasing sequence of positive integers such that (C.35) holds for each $q = q_j$. Then the proof of [8, Theorem 2(ii)] shows that for each sufficiently large $q_j$, there exists a root $z_j = 1 - s_j + \mathrm{i}t_j$ of $\rho(z_j) = 1$ with

$$0 < s_j \leq C_1 q_j^{-2/(r-1)}, \tag{C.36}$$

$$C_2 q_j \leq t_j \leq C_3 q_j. \tag{C.37}$$

We apply Lemma C.4 to the function $h(x)$ and find, for sufficiently large $j$, $x_j$ such that, using (C.36) and (C.37),

$$x_j \geq \frac{t_j}{2s_j} \geq C_4 \frac{t_j}{q_j^{-2/(r-1)}} \geq C_5 \frac{t_j}{t_j^{-2/(r-1)}} = C_5 t_j^{(r+1)/(r-1)} \tag{C.38}$$

and

$$|h(x_j)| \geq e^{-C_6 t_j} \geq e^{-C_7 x_j^{(r-1)/(r+1)}}. \tag{C.39}$$

Let $\lambda_j := e^{x_j}$. Since (C.36) implies $s_j \to 0$ and thus $x_j \to \infty$, we have $\lambda_j \to \infty$. Furthermore, by (C.39),

$$|R_1(\lambda_j)| = |h(x_j)| \geq e^{-C_7 x_j^{(r-1)/(r+1)}} = e^{-C_7 (\log \lambda_j)^{(r-1)/(r+1)}}. \tag{C.40}$$

□

## References


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

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

[3] Miklós Bóna: $k$-protected vertices in binary search trees. *Adv. in Appl. Math.* **53** (2014), 1-11.

[4] Miklós Bóna & Boris Pittel: On a random search tree: asymptotic enumeration of vertices by distance from leaves. *Adv. in Appl. Probab.* **49** (2017), no. 3, 850–876.

[5] Luc Devroye & Svante Janson: Protected nodes and fringe subtrees in some random trees. *Electron. Commun. Probab.* **19** (2014), no. 6, 10 pp.

[6] Michael Drmota: *Random Trees. An interplay between Combinatorics and Probability.* SpringerWienNewYork, 2009.

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

[8] Philippe Flajolet, Mathieu Roux & Brigitte Vallée: Digital trees and memoryless sources: from arithmetics to analysis. *Discrete Math. Theor. Comput. Sci. Proc.*, **AM** (2010), 233–260.

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

[10] Michael Fuchs, Hsien-Kuei Hwang & Vytas Zacharovas: An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoret. Comput. Sci.* **527** (2014), 1–36.

[11] Allan Gut: *Probability: A Graduate Course.* 2nd ed., Springer, New York, 2013.

[12] G. H. Hardy & E. M. Wright: *An Introduction to the Theory of Numbers.* 4th ed., Oxford Univ. Press, Oxford, 1960.

[13] Cecilia Holmgren & Svante Janson: Limit laws for functions of fringe trees for binary search trees and recursive trees. *Electronic J. Probability* **20** (2015), no. 4, 1–51.

[14] Cecilia Holmgren & Svante Janson: Fringe trees, Crump–Mode–Jagers branching processes and $m$-ary search trees. *Probability Surveys* **14** (2017), 53–154.

[15] Philippe Jacquet & Mireille Régnier: Normal limiting distribution for the size and the external path length of tries. Report RR-0827, INRIA, 1988. `https://hal.inria.fr/inria-00075724/PDF/RR-0827.pdf`

[16] Philippe Jacquet & Wojciech Szpankowski: *Analytic Pattern Matching.* Cambridge Univ. Press, Cambridge, UK, 2015

[17] Svante Janson: Monotonicity, asymptotic normality and vertex degrees in random graphs. *Bernoulli* **13**:4 (2007), 952–965.

[18] Svante Janson: Renewal theory in analysis of tries and strings. *Theoretical Computer Science* **416** (2012), 33–54.

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

[20] Svante Janson: Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Structures Algorithms* **48** (2016), no. 1, 57–101.

[21] Olav Kallenberg: *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.

[22] Donald E. Knuth: *The Art of Computer Programming. Vol. 3: Sorting and Searching.* 2nd ed., Addison-Wesley, Reading, Mass., 1998.

[23] Hosam Mahmoud: *Evolution of Random Search Trees*, Wiley, New York, 1992.

[24] Hosam H. Mahmoud and Mark Daniel Ward: Asymptotic distribution of two-protected nodes in random binary search trees. *Applied Mathematics Letters* **25** (2012), no. 12, 2218–2222.

[25] A. Meir & J. W. Moon: On the log-product of the subtree-sizes of random trees. (English summary) Random Structures Algorithms 12 (1998), no. 2, 197212.

[26] Hanène Mohamed & Philippe Robert: A probabilistic analysis of some tree algorithms. *Ann. Appl. Probab.* **15** (2005), no. 4, 2445–2471.

[27] Olle Nerman: Stochastic monotonicity and conditioning in the limit. *Scand. J. Statist.* **25** (1998), no. 3, 569–572.

[28] I. F. Pinelis: Estimates of moments of infinite-dimensional martingales. (Russian) *Mat. Zametki* **27** (1980), no. 6, 953–958, 990. English translation: *Math. Notes* **27** (1980), no. 5–6, 459–462.

[29] Iosif Pinelis: Exact Rosenthal-type bounds. *Ann. Probab.* **43** (2015), no. 5, 2511–2544.

[30] Mireille Régnier & Philippe Jacquet: New results on the size of tries. *IEEE Trans. Inform. Theory* **35** (1989), no. 1, 203–205.

[31] Wojciech Szpankowski: *Average Case Analysis of Algorithms on Sequences.* Wiley-Interscience, New York, 2001.

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

[33] Stephan Wagner: Central limit theorems for additive tree parameters with small toll functions. *Combin. Probab. Comput.* **24** (2015), no. 1, 329–353.

[34] Antoni Zygmund: *Trigonometric Series.* 2nd ed., Cambridge Univ. Press, Cambridge, UK, 1959.

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

*Email address*: `svante.janson@math.uu.se`

*URL*: `http://www.math.uu.se/~svante/`