Central limit theorems for fringe trees in patricia tries

Jasper Ischebeck

October 17, 2022

We give theorems about asymptotic normality of general additive functionals on patricia tries, derived from results on tries. These theorems are applied to show asymptotic normality of the distribution of random fringe trees in patricia tries. Formulas for asymptotic mean and variance are given. The proportion of fringe trees with k keys is asymptotically, ignoring oscillations, given by $(1-\rho(k))/(H+J)k(k-1)$ with the source entropy H, an entropy-like constant J, that is H in the binary case, and an exponentially decreasing function $\rho(k)$. Another application gives asymptotic normality of the independence number and the number of k-protected nodes.

1 Introduction

A trie is an abstract data structure for strings that directly uses the stored string as a path in a tree. It has various uses for sorting, searching and compressing. Because it directly uses the stored strings rather than a hash, it particularly excels at memory usage and prefix lookup. We will throughout the paper assume that the trie is built from infinite strings whose letters are i.i.d. The letters will be distributed by a fixed distribution on a finite alphabet $\mathcal{A}$. The most important case will be $\mathcal{A} = \{0, 1\}$. In this model of i.i.d. letters, the size and other important parameters of the trie are inversely proportional to the entropy of the letter distribution and have no deterministic bound.

The patricia trie (Practical Algorithm to Retrieve Information Coded in Alphanumeric), introduced 1968 by Morrison [Mor68] improves on the trie by eliminating nodes with only one child. It has deterministic size in the main $|\mathcal{A}| = 2$ case. Knuth [Knu73] gives a good explanation and some examples of usage. This structure is also called radix tree or compressed trie; more on terminology in Remark 2.1.

The number of strings will be either fixed or a Poisson variable. For certain letter probabilities, most notably the uniform case, (numerically small) oscillations occur in the asymptotics of both mean and variance for functionals of tries and patricia tries. Nonetheless, one has asymptotic normality for suitable normalizations. Janson [Jan22] gave central limit theorems for additive functionals on tries with rather weak conditions and a law of large numbers. His general results cover various parameters which had been studied one by one by various authors before, for example [JS15], [Szp90]. We show that these theorems can be extended to patricia tries by relating additive functionals on patricia tries to those on tries (Prop. 2.2). We get central limit theorems for the Poisson model and the fixed n model (Theorem 3.3) and as corollary a strong law of large numbers (Theorem 3.9).

We then go on and give applications of this theorem. This gives us e.g. the asymptotic distribution of random fringe trees of patricia tries (Theorem 4.5). Because the size of patricia tries is deterministic in the binary case, many statements are simpler in that case. We will usually begin with the general case and then show how the statements simplify in the binary case. Furthermore, we calculate the asymptotic mean and sometimes also the asymptotic variance of some other additive functionals, for example the amount of k-protected nodes in patricia tries.

2 Preliminaries

We use notation similar to Janson[Jan22], with some minor differences, such as writing p_α instead of $P(\alpha)$ for the probability of a string α.

In general, we use C to denote generic constants, which can change between different occurrences. log is the natural logarithm, even in the definition of the entropy.

2.1 Strings

Fix a finite alphabet $\mathcal{A}$ and define $\mathcal{A}^* = \bigcup_{n=0}^{\infty} \mathcal{A}^n$ as the set of *finite strings* on it. The items in the string are called *characters* or *chars* for short. Denote the empty string by ε. Let $|\alpha|$ be the *length* of string, defined as the amount of chars. We denote concatenation of two strings $\alpha, \beta \in \mathcal{A}^*$ by $\alpha\beta$ and extend this notation to sets of strings, such as in $\alpha B := \{\alpha\beta \mid \beta \in B\}$ for $B \subseteq \mathcal{A}^*$, and to letters $a \in \mathcal{A}$. Denote by $\mathcal{A}^{\mathbb{N}}$ the *infinite strings* (sequences) on $\mathcal{A}$.

We also fix a probability distribution p, called *source*, on the alphabet and sample infinite strings $\Xi = (\Xi(1), \Xi(2), \ldots)$, where $\Xi(1), \Xi(2), \ldots \in \mathcal{A}$ are independently distributed with p. For $a \in \mathcal{A}$ write $p_a := p(\{a\})$ for the point mass and for a finite string $\alpha = (a_1, \ldots, a_n) \in \mathcal{A}^*$ let $p_\alpha := \prod_{i=1}^{n} p_{a_i}$.

To avoid trivialities, we assume that $p_a > 0$ for every $a \in \mathcal{A}$ and $|\mathcal{A}| \geq 2$. Hence, for two independent infinite strings Ξ_1, Ξ_2 we have $\Xi_1 \neq \Xi_2$ almost surely. Let $S := \{-\log p_a \mid a \in \mathcal{A}\}$. If there is a real number $d > 0$ such that $S \subset d\mathbb{Z}$, we call the source p *periodic*. In this case define d_p as the biggest of such d; this d is called the greatest common divisor of S. If the source is not periodic, we define $d_p := 0$. The periodic case is where periodic oscillations typically occur. The uniform distribution on $\mathcal{A}$ on periodic; this is the *symmetric case*.

For complex $s \in \mathbb{C}$ let

$$\rho(s) := \sum_{a \in \mathcal{A}} p_a^s. \tag{2.1}$$

For natural $s \in \mathbb{N}$, we may interpret this as the probability of s strings to start with the same char. Note that for longer substrings of length $m \in \mathbb{N}$ we have

$$\sum_{\substack{\alpha \in \mathcal{A}^* \\ |\alpha|=m}} p_\alpha^s = \sum_{a_1 \dots a_m \in \mathcal{A}^m} \prod_{i=1}^{m} p_{a_i}^s = \rho(s)^m. \tag{2.2}$$

An important quantity of the source is the *entropy* H, given by

$$H = -\sum_{a \in \mathcal{A}} p_a \log p_a = -\rho'(1), \tag{2.3}$$

which is central in the analysis of tries and patricia tries. There is a similar quantity, which we call J, defined by

$$J = -\sum_{a \in \mathcal{A}} (1 - p_a) \log(1 - p_a), \tag{2.4}$$

which will show up in the size of patricia tries. In the most common case of $|\mathcal{A}| = 2$, this will be equal to the entropy H. Note that, however, J is bounded by

$$J \leq (|A| - 1) \log\left(1 - \frac{1}{|A| - 1}\right) < 1. \tag{2.5}$$

2.2 Trees

We consider finite, rooted trees where the edges have labels in $\mathcal{A}$ and every node has no two outgoing edges with the same label. These are subtrees of the infinite $|\mathcal{A}|$-ary tree T_∞ where the $|\mathcal{A}|$ outgoing edges of each node are each labelled with a unique letter $a \in \mathcal{A}$. We call these trees *($\mathcal{A}$-)labelled trees* and define the set of such trees as $\mathfrak{T}^+$ and $\mathfrak{T} := \mathfrak{T}^+ \cup \{\emptyset\}$.

The labels on the path from the root to a node $v \in T^\infty$ form a unique finite string, and we can associate the nodes in T^∞ with the finite strings $\mathcal{A}^*$. A node α is ancestor of a node β if and only if α is a prefix of β. The root is the empty string ε.

We identify a subtree with the set of its nodes, so for a subtree T we define $|T|$ as the amount of nodes, made up of the amount of *external nodes* (leaves) $|T|_e$ and the amount of *internal nodes* $|T|_i$. Let $\mathfrak{T}_n := \{T \in \mathfrak{T}_n \mid |T|_e = n\}$ be the set of subtrees with n leaves. Tries and patricia tries store their data only in leaves, so this is the most natural notion of size for them. Let $\bullet$ be the tree consisting of only the root, the only tree in $\mathfrak{T}_1$.

Given a subtree T and a string $v \in T$, let T^v be the subtree of T consisting of v and its descendants. Those subtrees are called *fringe trees* of T. For $v \in \mathcal{A}^* \setminus T$, we define $T^v = \emptyset$. When seen as subsets of $\mathcal{A}^*$, the descendants of v are $v\mathcal{A}^*$, the strings beginning with v.

For a node $v\alpha \in T^v$ the path to the new root v is only α, so the new tree is given as

$$T^v = \{\alpha \in \mathcal{A}^* \mid v\alpha \in T\}. \tag{2.6}$$

Let T^* be a (uniformly) random fringe subtree of a tree T, defined as T^v for a uniform random node $v \in T$. Note that v can also be a leaf. One might also consider a random fringe tree by taking only internal nodes. This is equivalent to conditioning T^* on $\{T^* \neq \bullet\}$. The results can then be easily transferred to this version.

Given a function $\varphi : \mathfrak{T} \to \mathbb{R}^n, n \geq 1$ with $\varphi(\emptyset) = 0$, we define the corresponding *additive functional* $\Phi : \mathfrak{T} \to \mathbb{R}^n$ as

$$\Phi(T) = \sum_{\alpha \in \mathcal{A}^*} \varphi(T^\alpha) = \sum_{v \in T} \varphi(T^v). \tag{2.7}$$

This can be written recursively as

$$\Phi(T) = \varphi(T) + \sum_{a \in \mathcal{A}} \Phi(T^a); \quad \Phi(\emptyset) = 0, \tag{2.8}$$

which also shows that every functional is an additive functional. The term "additive functional" for Φ is thus mainly defined by its relation to φ, which is called *toll function* of Φ.

2.3 Tries

A *trie* $T(\mathfrak{X})$ from strings $\mathfrak{X} = \{\Xi_1, \dots, \Xi_n\} \subset \mathcal{A}^{\mathbb{N}}$ is the minimal subtree of T_∞ such that every string Ξ_i is contained in a unique leaf whose path is a prefix of Ξ_i. The path of the leaf containing the string Ξ_i is thus the shortest prefix of Ξ_i that is not a prefix for all Ξ_j with $j \neq i$.

A more algorithmic, recursive definition is given as follows: For $n = 0$ the trie is empty. For $n = 1$ it only consists of the root, containing the only string. For $n \geq 2$ the strings are sorted by their first character $a \in \mathcal{A}$. The root is in the trie, and the a-subtree of the root is the trie from the strings $\{\Xi \in \mathcal{A}^* \mid (a, \Xi) \in \mathfrak{X}\}$ beginning with a.

From the construction we have that $T(\mathfrak{X})^v$ for $v \in \mathcal{A}^*$ contains the strings $v\mathcal{A}^{\mathbb{N}} \cap \mathfrak{X}$ starting with v and is itself a trie of the strings $\{\alpha \in \mathcal{A}^{\mathbb{N}} \mid v\alpha \in \mathfrak{X}\}$, except in the case where there is exactly one string Ξ starting with v, in which v could be not in $T(\mathfrak{X})$. This is the case if there is a shorter prefix of Ξ than v that is unique to Ξ. Then Ξ is farther up the tree. This exception is the reason we will have to special-case $\bullet$ in our theorems.

We now construct two models of random tries. Define $\mathcal{T}_n := T(\Xi_1, \dots, \Xi_n)$ for $n \in \mathbb{N}$ as the trie for n i.i.d. random strings as in Section 2.1. Call this the *fixed n model*. Define $\widetilde{\mathcal{T}}_\lambda := T(\Xi_1, \dots, \Xi_{N_\lambda})$ for $\lambda \in (0, \infty)$ as the trie for N_λ i.i.d. random strings, where N_λ is Poi(λ)-distributed and independent of the strings. This is the *Poisson model*.

We noted before that, for an $\alpha \in \mathcal{A}^*$, the strings in $\widetilde{\mathcal{T}}_\lambda^\alpha$ are the strings starting with α. In the Poisson model, let N_λ^α be the amount of these strings; N_λ^α is then Poi(λp_α)-distributed. Because of the recursive definition of tries, $\widetilde{\mathcal{T}}_\lambda^\alpha$ is a trie of N_λ^α independent strings and there-

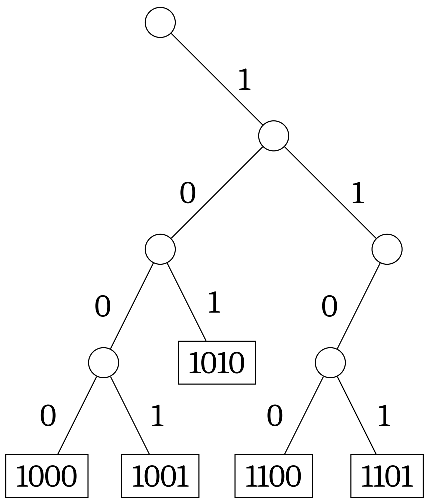

Figure 1: A trie made from the strings 1100, 1101, 1010, 1000, 1001, 1011

fore a copy of $\widetilde{\mathcal{T}}_{p_a\lambda}$. Additionally, $\widetilde{\mathcal{T}}_\lambda^a$ and $\widetilde{\mathcal{T}}_\lambda^b$ are independent for $a \neq b \in \mathcal{A}$. There is again an exception for the case $N_\lambda^a = 1$, where $\widetilde{\mathcal{T}}_\lambda^a$ could be either $\bullet$ or $\emptyset$.

2.4 Patricia trie

A *patricia trie* can be constructed from a trie by iteratively merging every node with only one child with its child. If this child also has just one child, it will also get merged with it, and so on. In this way, multiple nodes forming a chain get merged into a single node. The labels of the edges that connected the merged nodes are kept as a string $\alpha \in \mathcal{A}^*$ as an extra date in the child, we call them *contracted edges*. See Figure 2 for an example. So our patricia tries consist not only of a labelled tree T, but also of a map of contracted edges $T_i \to \mathcal{A}^*$ from the internal nodes T_i of T to $\mathcal{A}^*$.

We call this process pat, so $\text{pat}(T)$ is the patricia trie to a trie T. From $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_\lambda$ we define $\mathcal{P}_n := \text{pat}(\mathcal{T}_n)$ and $\widetilde{\mathcal{P}}_\lambda := \text{pat}(\widetilde{\mathcal{T}}_\lambda)$ as random models of Patricia tries.

The patricia trie can also be defined recursively and directly, similar to the trie. The patricia trie to a set $\mathfrak{X}$ of strings is given by $\emptyset$ if $|\mathfrak{X}| = 0$ and by $\bullet$ if $|\mathfrak{X}| = 1$. If $|\mathfrak{X}| \geq 2$, let v be the common prefix of all strings in $\mathfrak{X}$. The patricia trie T is then given as follows: It has v as contracted edges in the root, and T^a for $a \in \mathcal{A}$ is the patricia trie from the strings $\{\alpha \in \mathcal{A}^{\mathbb{N}} \mid va\alpha \in \mathfrak{X}\}$ starting with va.

We will for the most part ignore the contracted edges and consider a patricia trie equal to the tree, but we allow additive functionals to depend on them, with one important exception: We impose a restriction that the toll function must not depend on the contracted edges in the root. We therefore define an *additive functional on patricia tries* as an additive functional on trees whose toll function can additionally depend on the contracted edges besides those in the root.

Remark 2.1. This tree is more correctly referred to as a radix tree or compressed trie. For performance the original patricia tries merge internal with external nodes, storing the keys

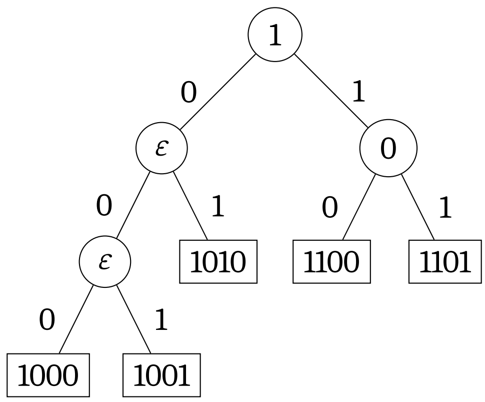

Figure 2: The patricia trie made from the same strings 1100, 1101, 1010, 1000, 1001, 1011 as in Figure 1. The labels in the nodes are the contracted edges.

in internal nodes, which makes the actual data structure more like an acyclic graph. But the edges pointing to external nodes are still distinguished from normal edges (they point "upwards"), and the algorithm always terminates after reaching such an edge, so in effect the tree is the same[Knu73]. The difference is thus only in implementation, and when mainly considering the tree, *radix tree* and *patricia trie* are used synonymously.

Another difference is that in patricia tries only the amount of contracted edges is kept per node in a so-called `SKIP` attribute because the actual contracted edges are irrelevant for the search algorithm.

We can link additive functionals on patricia tries with ones on tries:

Proposition 2.2. *An additive functional Φ on patricia tries defines an additive functional $\widetilde{\Phi} := \Phi \circ p$ on tries by push-forward with* pat*. The toll function $\widetilde{\varphi}(T)$ of $\widetilde{\Phi}$ is given by*

$$\widetilde{\varphi}(T) = \begin{cases} 0 & T \text{ has exactly 1 child} \\ \varphi(\operatorname{pat}(T)) & \text{else.} \end{cases} \tag{2.9}$$

Proof. Each node $\tilde{v}$ of the patricia trie $\operatorname{pat}(T)$ was created by compressing nodes in the trie. Associate $\tilde{v}$ with the youngest node v of the compressed nodes (the one that is a descendant of all the others). In this way, we have a bijection between the nodes of the patricia trie and the nodes in the trie which have not exactly one child.

The fringe tree of $\tilde{v}$ consists of the compressed nodes of the fringe tree of v, so $\operatorname{pat}(T)^{\tilde{v}} = \operatorname{pat}(T^v)$, except that $\operatorname{pat}(T^v)$ has no contracted edges in the root. Because the toll function φ is not allowed to depend on the contracted edges in the root, we nevertheless have $\varphi(\operatorname{pat}(T)^{\tilde{v}}) = \widetilde{\varphi}(T^v)$ and summing over all nodes yields the equation.

□

If we lift the restriction on toll functions, $\widetilde{\varphi}$ becomes a difference $\varphi(\operatorname{pat}(T)) - \varphi(\operatorname{pat}(T^a))$ if there is only one child $a \in \mathcal{A}$. Since we usually do not depend on the contracted edges this added complexity does not seem useful.

We can now extend results on additive functionals on tries to additive functionals on patricia tries by translating properties of $\widetilde{\Phi}$ into properties of Φ. If we define an additive functional Φ on patricia tries, φ is understood to be the corresponding toll function and vice versa. In similar fashion $\widetilde{\Phi}$, $\widetilde{\varphi}$ are then the pullbacks on tries. We might use subscripts to distinguish different additive functionals and their toll functions. From the definition we have $\Phi(\widetilde{\mathscr{P}}_\lambda) = \widetilde{\Phi}(\widetilde{\mathscr{T}}_\lambda)$ and $\Phi(\mathscr{P}_n) = \widetilde{\Phi}(\mathscr{T}_n)$. We mainly use $\Phi(\mathscr{P}_n)$ etc. in the results, but might use $\widetilde{\Phi}(\widetilde{\mathscr{T}}_\lambda)$ etc. in calculations, especially when together with $\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)$ etc.

Example 2.3. A simple toll function is given by

$$\varphi_\bullet(T) := \mathbf{1}\{T = \bullet\} = \mathbf{1}\{|T|_e = 1\}. \tag{2.10}$$

Because the only trie that is compressed to $\bullet$ is $\bullet$ itself, $\widetilde{\varphi}_\bullet = \varphi_\bullet$. The subtrees equal to $\bullet$ are the leaves, so $\Phi_\bullet(T) = \widetilde{\Phi}_\bullet(T)$ counts the leaves of T.

Example 2.4. We define $\varphi_p := \widetilde{\varphi}_p := \widetilde{\varphi}_1$ as the induced additive functional for the constant toll function $\varphi_1(T) := 1$. Then, $\varphi_p(T) = 1$ if and only if the root of T has not exactly one child and is therefore included in the patricia trie, and $\varphi_p(T) = 0$ else. With this, we can alternatively express the relation between a toll function φ on patricia tries to the toll function $\widetilde{\varphi}$ on tries as $\widetilde{\varphi} = (\varphi \circ p)\varphi_p$.

An additive functional Φ of patricia tries is called *increasing* if for two finite sets $\mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \mathscr{A}^*$ of strings $\widetilde{\Phi}(T(\mathfrak{X}_1)) \leq \widetilde{\Phi}(T(\mathfrak{X}_2))$ holds. This notion is central for the depoissonisation used by Janson [Jan22, (6.64) ff.].

2.5 Bucket trie

Bucket tries are a generalization of tries, where a leaf can pack up to a fixed number $b \geq 1$ of strings. In the recursive definition, this means that up to b strings, the bucket trie consists of only one node until it splits up like a normal trie. A leaf of the bucket trie is called *bucket* and can contain 1 to b strings. A normal trie is a bucket trie with bucket size 1.

Given a bucket size $b \geq 1$, we can construct the trie T from a set of strings by first constructing the bucket trie T' with bucket size b, and then replacing each bucket with a small trie out of the contents of the bucket. In our model of strings with i.i.d. chars, conditioned on the bucket trie, the small tries are independent of each other and the trie from a bucket of k strings is a copy of $\mathscr{T}_k$.

We can also construct a bucket patricia trie by compressing nodes in a bucket trie. Then we can construct the patricia trie out of the bucket patricia trie by again growing patricia tries out of the buckets.

By compressing all leaves with parents with less or equal b strings, we can get the bucket trie / patricia trie out of the trie / patricia trie. Call this operation buc_b.

We will use the bucket trie as a tool in proofs because it is clearer to “condition on the bucket trie” than “condition on nodes with more than b strings and the amount of strings in their children.”

2.6 Mellin transform

For a measurable function f on $(0, \infty)$ the *Mellin transform* is defined by

$$f^*(s) := \int_0^\infty f(x) x^{s-1} \mathrm{d}x \tag{2.11}$$

for all $s \in \mathbb{C}$ where the integral absolutely converges. Since the absolute value of x^{s-1} depends only on the real part of s, this is always a vertical strip in the complex plane.

The Mellin transform is obviously linear. An important example is given by $f(x) = x^k e^{-\lambda x}$ for $k \in \mathbb{R}$, $\lambda > 0$, where

$$f^*(s) = \frac{\Gamma(k+s)}{\lambda^{s+k}} \tag{2.12}$$

for $\operatorname{Re} s > -k$. Another example is $f(x) = e^{-\lambda x} - 1$ for $\lambda > 0$, where for $-1 < \operatorname{Re} s < 0$

$$f^*(s) = \frac{\Gamma(s)}{\lambda^s}. \tag{2.13}$$

Those two examples will come up a lot in calculations.

2.7 Convergence

We use standard o and O notation in both a global and asymptotic sense: $f(x) = O(g(x))$ for $x \in S$ means that $|f(x)| \leq |Cg(x)|$ for all $x \in S$, while $f(x) = O(g(x))$ for $x \to \infty$ means that $|f(x)| \leq |Cg(x)|$ for large x. Also, we use Ω and Θ: $f(x) = \Omega(g(x))$ means $g(x) = O(f(x))$; and $f(x) = \Theta(g(x))$ means that both $f(x) = O(g(x))$; $g(x) = O(f(x))$.

We use $\overset{\mathbb{P}}{\to}$ denote convergence in probability and $\overset{d}{\to}$ for convergence in distribution.

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

$$\mathbb{E}f(X_n) = \mathbb{E}f(Y_n) + o(1) \quad \text{as } n \to \infty. \tag{2.14}$$

If $S \subseteq \mathbb{R}$ and additionally

$$\mathbb{E}X_n^s = \mathbb{E}Y_n^s + o(1) \tag{2.15}$$

with both sides finite for $s \in \mathbb{N}$ holds, we say that $X_n \overset{d}{\approx} Y_n$ *with moments of order* s. Similarly, for $s > 0$ we say that $X_n \overset{d}{\approx} Y_n$ *with absolute moments of order* s if (2.14) holds and also

$$\mathbb{E}|X_n|^s = \mathbb{E}|Y_n|^s + o(1) \tag{2.16}$$

with both sides finite. We will write *with [absolute] moments* for approximation in distribution with normal moments and absolute moments. If $|Y_n|^s$ is uniformly integrable, then approximation with normal moments follows from approximation with absolute moments [Jan22, Lemma B.1].

3 Central limit theorems for tries

We restate some of Janson's results regarding additive functionals $\widetilde{\Phi}$ on random tries $\mathscr{T}_n$ or $\widetilde{\mathscr{T}}_\lambda$, replacing $\widetilde{\Phi}(\mathscr{T}_n)$ with $\Phi(\mathscr{P}_n)$ to change them into statements regarding patricia tries.

Theorem 3.1. *Let Φ be an additive functional on patricia tries with associated toll function φ and functions $\widetilde{\Phi}$, $\widetilde{\varphi}$ on tries as before. Suppose that, for some $\varepsilon > 0$, as $\lambda \to \infty$ both $\mathbb{E}[\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)] = O(\lambda^{1-\varepsilon})$ and $\operatorname{Var}(\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)) = O(\lambda^{1-\varepsilon})$ hold. Let*

$$\chi := \varphi(\bullet) \tag{3.1}$$

$$f_E(\lambda) := \mathbb{E}\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda) - \chi\lambda e^{-\lambda} \tag{3.2}$$

$$\begin{aligned} f_V(\lambda) := {} & 2\operatorname{Cov}\Big(\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda), \widetilde{\Phi}(\widetilde{\mathscr{T}}_\lambda)\Big) - \operatorname{Var}\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda) \\ & + 2\chi\lambda e^{-\lambda}(\mathbb{E}\widetilde{\Phi}(\widetilde{\mathscr{T}}_\lambda) - \mathbb{E}\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)) - \chi^2\lambda e^{-\lambda}(1-\lambda e^{-\lambda}) \end{aligned} \tag{3.3}$$

$$f_C(\lambda) := \operatorname{Cov}\Big(\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda), N_\lambda\Big) + \chi\lambda(\lambda-1)e^{-\lambda}. \tag{3.4}$$

Then the following hold. For $\lambda \to \infty$

$$\frac{\mathbb{E}\Phi(\widetilde{\mathscr{P}}_\lambda)}{\lambda} = \chi + \frac{1}{H}\psi_E(\log\lambda) + o(1), \tag{3.5}$$

$$\frac{\operatorname{Var}\Phi(\widetilde{\mathscr{P}}_\lambda)}{\lambda} = \chi^2 + \frac{1}{H}\psi_V(\log\lambda) + o(1), \tag{3.6}$$

$$\frac{\operatorname{Cov}(\Phi(\widetilde{\mathscr{P}}_\lambda), N_\lambda)}{\lambda} = \chi + \frac{1}{H}\psi_C(\log\lambda) + o(1), \tag{3.7}$$

where ψ_X for $X = E, V, C$ are bounded continuous functions defined as follows:

- *If $d_p = 0$, then ψ_X is constant and given for all t by*

$$\psi_X(t) := f_X^*(-1). \tag{3.8}$$

- *If $d := d_p > 0$, then ψ_X is a continuous d-periodic function having the Fourier series*

$$\psi_X(t) \sim \sum_{m=-\infty}^{\infty} f_X^*\Big(-1 - \frac{2\pi m}{d}i\Big)e^{2\pi imt/d}. \tag{3.9}$$

Moreover, if $X = E$ or if $f_X'(\lambda) = O(\lambda^{-\varepsilon_1})$ as $\lambda \to \infty$ for some $\varepsilon_1 > 0$, then the Fourier series (3.8) *converges absolutely, and thus $\sim$ may be replaced by $=$ in* (3.8).

Remark 3.2. Even if $d_p > 0$, we may regard the constant term $f_X^*(-1)$ as "average asymptotic value" because the oscillations are typically numerically small. In most of our examples, the f_X^* functions will consist of sums with the Gamma function Γ, for which is well known that $|\Gamma(x+iy)|$ converges to 0 swiftly for $|y| \to \infty$.

If $\chi = \varphi(\bullet)$ is 0, the formulas for the f_X simplify. If we now split up $\widetilde{\Phi}$ to a sum of $\widetilde{\varphi}$ in the definition of f_V, we get

$$f_V(\lambda) = 2 \sum_{\alpha \in \mathcal{A}^*} \mathrm{Cov}(\widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda), \widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda^\alpha)) - \mathrm{Cov}(\widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda), \widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda)), \tag{3.10}$$

which leads us to define the notation

$$\sum_\alpha^* := 2 \sum_{\alpha \in \mathcal{A}^*} - \sum_{\alpha = \varepsilon} = \sum_{\alpha \in \mathcal{A}^*} + \sum_{\substack{\alpha \in \mathcal{A}^* \\ \alpha \neq \varepsilon}} . \tag{3.11}$$

This lets us write (3.10) as

$$f_V(\lambda) = \sum_\alpha^* \mathrm{Cov}(\widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda), \widetilde{\varphi}(\widetilde{\mathcal{T}}_\lambda^\alpha)). \tag{3.12}$$

Through depoissonisation Janson derives the following theorem for the fixed n case, which is also our main result for patricia tries.

Theorem 3.3. *[Jan22, Theorem 3.9] (Central limit theorem in the fixed n case.) Let φ_+, φ_- be a bounded toll functions on patricia tries so that their additive functionals Φ_+, Φ_- are increasing. Let $\Phi := \Phi_+ - \Phi_-$ and $\widetilde{\varphi}$, $\widetilde{\Phi}$ be the corresponding toll function and additive functional on tries as in 2.2. Then, with notation from Theorem 3.1, especially f_X* (3.2)-(3.3) *and ψ_X* (3.9)*:*

i) If $d_p = 0$, then as $\lambda \to \infty$ respective $n \to \infty$,

$$\frac{\Phi(\widetilde{\mathcal{P}}_\lambda) - \mathbb{E}[\Phi(\widetilde{\mathcal{P}}_\lambda)]}{\sqrt{\lambda}} \overset{d}{\to} \mathcal{N}(0, \sigma^2) \tag{3.13}$$

$$\frac{\Phi(\mathcal{P}_n) - \mathbb{E}[\Phi(\mathcal{P}_n)]}{\sqrt{n}} \overset{d}{\to} \mathcal{N}(0, \widehat{\sigma}^2) \tag{3.14}$$

with all moments, where

$$\sigma^2 = \chi^2 + H^{-1} f_V^*(-1) \tag{3.15}$$

$$\widehat{\sigma}^2 = H^{-1} f_V^*(-1) - H^{-2} f_C^*(-1)^2 - 2\chi H^{-1} f_C^*(-1) \tag{3.16}$$

ii) For any d_p, as $\lambda \to \infty$ and $n \to \infty$,

$$\frac{\Phi(\widetilde{\mathcal{P}}_\lambda) - \mathbb{E}[\Phi(\widetilde{\mathcal{P}}_\lambda)]}{\sqrt{\lambda}} \overset{d}{\approx} \mathcal{N}(0, \sigma^2(\lambda)) \tag{3.17}$$

$$\frac{\Phi(\mathcal{P}_n) - \mathbb{E}[\Phi(\mathcal{P}_n)]}{\sqrt{n}} \overset{d}{\approx} \mathcal{N}(0, \widehat{\sigma}^2(n)) \tag{3.18}$$

with all moments, where

$$\sigma^2(\lambda) = \chi^2 + H^{-1}\psi_V(\log\lambda) \tag{3.19}$$
$$\widehat{\sigma}^2(n) = H^{-1}\psi_V(\log n) - H^{-2}\psi_C(\log n)^2 - 2\chi H^{-1}\psi_C(\log n) \tag{3.20}$$

iii) The expected values satisfy

$$\mathbb{E}[\widetilde{\Phi}(\mathcal{T}_n)] - \mathbb{E}[\widetilde{\Phi}(\widetilde{\mathcal{T}}_n)] \in o\left(\sqrt{n}\right), \tag{3.21}$$

and we may thus replace $\mathbb{E}[\widetilde{\Phi}(\mathcal{T}_n)]$ *with* $\mathbb{E}[\widetilde{\Phi}(\widetilde{\mathcal{T}}_n)]$ *in* (3.14) *and* (3.18).

iv) If $\operatorname{Var}\widetilde{\Phi}(\widetilde{\mathcal{T}}_n) \in \Omega(n)$, *then*

$$\frac{\widetilde{\Phi}(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}[\widetilde{\Phi}(\widetilde{\mathcal{T}}_\lambda)]}{\sqrt{\operatorname{Var}\Phi(\widetilde{\mathcal{T}}_\lambda)}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{3.22}$$
$$\frac{\widetilde{\Phi}(\mathcal{T}_n) - \mathbb{E}[\widetilde{\Phi}(\mathcal{T}_n)]}{\sqrt{\operatorname{Var}\widetilde{\Phi}(\widetilde{\mathcal{T}}_n)}} \xrightarrow{d} \mathcal{N}(0, \widehat{\sigma}^2) \tag{3.23}$$

with all moments for all d_p.

v) The expected values $\mathbb{E}[\Phi(\mathcal{T}_n)], \mathbb{E}[\Phi(\widetilde{\mathcal{T}}_\lambda)]$ *satisfy*

$$\frac{\mathbb{E}[\Phi(\widetilde{\mathcal{T}}_\lambda)]}{\lambda} = \frac{\psi_E(\log\lambda)}{H} + o(1) \tag{3.24}$$
$$\frac{\mathbb{E}[\Phi(\widetilde{\mathcal{T}}_n)]}{n} = \frac{\psi_E(\log n)}{H} + o(1). \tag{3.25}$$

To show the condition $\operatorname{Var}\widetilde{\Phi}(\mathcal{T}_n) = \Omega(n)$ in iv), Janson gives following lemma:

Lemma 3.4. *[Jan22, 3.14] Let* Φ *be an additive functional with bounded toll function* φ *and suppose that* $\widetilde{\varphi}(\mathcal{T}_n)$ *is deterministic for almost all* $n \in \mathbb{N}$ *and that* $\operatorname{Var}\Phi(\mathcal{P}_n) \neq 0$ *for some* $n \geq 1$. *Then* $\operatorname{Var}\Phi(\mathcal{P}_n) = \Omega(n)$.

We will often have toll functions that are 0 if the tree exceeds a certain size, then we can derive asymptotic normality using this lemma. An example where the condition $\operatorname{Var}\widetilde{\Phi}(\mathcal{T}_n) = \Omega(n)$ does not hold is the size of a binary patricia trie, as it is deterministic. It can be extended to following lemma:

Lemma 3.5. *Let* Φ *be an additive functional with toll function* φ *and let* φ' *be another toll function. Suppose that there is a* $b \geq 2$, *such that*

1. $\varphi(T)$ *for* $|T|_e > b$ *only depends on the bucket trie* $\operatorname{buc}_b(T)$ *of* T *with bucket size* b *and on* $\varphi'(T^\alpha)$ *for the* b*-buckets* $\alpha \in (\operatorname{buc}_b(T))_e$.

2. *The additive functional $\Phi(\mathscr{T}_b)$ for the buckets is not solely dependent on $\varphi'(\mathscr{T}_b)$, that is, we have*

$$\mathbb{E}[\mathrm{Var}(\Phi(\mathscr{T}_b) \mid \varphi'(\mathscr{T}_b))] > 0 \tag{3.26}$$

Then $\mathrm{Var}\,\Phi(\mathscr{T}_n) = \Omega(n)$.

Remark 3.6. Using Cramér-Wold, a multivariate version of this theorem can be shown. The results are linear in Φ, so we can use polarization to get results for the covariance between two additive functionals Φ_1, Φ_2. The full formula for the f_X function is given in [Jan22, (3.19)]; in the important special case of $\varphi_1(\bullet) = \varphi_2(\bullet) = 0$, it is

$$\begin{aligned} f_{V,12}(\lambda) := \mathrm{Cov}(\widetilde{\varphi}_1(\widetilde{\mathscr{T}}_\lambda), \widetilde{\Phi}_2(\widetilde{\mathscr{T}}_\lambda)) + \mathrm{Cov}(\widetilde{\Phi}_1(\widetilde{\mathscr{T}}_\lambda), \widetilde{\varphi}_2(\widetilde{\mathscr{T}}_\lambda)) \\ - \mathrm{Cov}(\widetilde{\varphi}_1(\widetilde{\mathscr{T}}_\lambda), \widetilde{\varphi}_2(\widetilde{\mathscr{T}}_\lambda)). \end{aligned} \tag{3.27}$$

This gives us as usual a function $\psi_{V,12}(\log n)$. Then, the covariance satisfies

$$n^{-1}\,\mathrm{Cov}(\Phi_1(\mathscr{P}_n), \Phi_2(\mathscr{P}_n)) \to \widehat{\sigma}_{12}(n) \tag{3.28}$$

where

$$\widehat{\sigma}_{12}(n) = H^{-1}\psi_{V,12}(\log n) - H^{-2}\psi_{C,1}(\log n)\psi_{C,2}(\log n) \tag{3.29}$$

Remark 3.7. The condition of boundedness on $\widetilde{\varphi}$ can be relaxed to the conditions sublinearity conditions $\mathbb{E}[\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)], \mathrm{Var}(\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)) = O(\lambda^{1-\varepsilon})$ from Theorem 3.1 and the condition that for all $r > 2$ we have

$$\mathbb{E}\left|\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda) - \mathbb{E}[\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda)]\right|^r = O(\lambda^{r/2}). \tag{3.30}$$

The function ψ_C can be calculated from ψ_E as follows:

Lemma 3.8. *Let $\widetilde{\varphi}$ be a toll function on tries as in Theorem 3.1. Then, for all λ, t and at least for $s \in \mathbb{C}$ with $\mathrm{Re}\, s \in (-2, -1 + \varepsilon/2)$*

$$f_C^*(s) = -s f_E^*(s) \tag{3.31}$$

$$f_C^*(-1) = f_E^*(s) \tag{3.32}$$

$$\psi_C(t) = \psi_E(t) + \psi_E'(t). \tag{3.33}$$

From Theorem 3.3 Janson derives a weak law of large numbers, while leaving as open question if one also had a.s. convergence. Using a proof technique for the strong law of large numbers for sums of i.i.d. random variables with a finite fourth absolute moment, we can show a.s. convergence. We therefore have following strong law of large numbers for patricia tries.

Theorem 3.9. *[Jan22, Theorem 3.12] Let φ be a bounded toll function on patricia tries as in*

Theorem 3.3. Let $f_E^*(s)$, $\psi_E(t)$ and χ be as in Theorem 3.1. Then, as $\lambda \to \infty$ and $n \to \infty$,

$$\frac{\Phi(\widetilde{\mathscr{P}}_\lambda)}{\lambda} - H^{-1}\psi_E(\log(\lambda)) - \chi \to 0 \text{ a.s.,} \tag{3.34}$$

$$\frac{\Phi(\mathscr{P}_n)}{n} - H^{-1}\psi_E(\log(n)) - \chi \to 0 \text{ a.s.} \tag{3.35}$$

In particular, if $d_p = 0$, then, as $n \to \infty$,

$$\frac{\Phi(\mathscr{P}_n)}{n} \to H^{-1}f_E^*(-1) + \chi \text{ a.s.} \tag{3.36}$$

Proof. We proof the fixed n case (3.35), the Poisson case (3.34) is analogous. Note that $\mathbb{E}[\Phi(\mathscr{P}_n)]$ is asymptotically $H^{-1}\psi_E(\log\lambda) + \chi$ according to (3.25). So (3.35) is equivalent to showing that $n^{-1}Z_n := n^{-1}(\Phi(\mathscr{P}_n) - \mathbb{E}\Phi(\mathscr{P}_n)) \to 0$ a.s. Writing the quantifiers of "not converging" out, we have

$$\begin{aligned}\left\{\frac{Z_n}{n} \nrightarrow 0\right\} &= \bigcup_{m\in\mathbb{N}}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right\} \\ &= \bigcup_{m\in\mathbb{N}}\limsup_{n\to\infty}\left\{\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right\}\end{aligned} \tag{3.37}$$

Since we have convergence of all moments of $Z_n/\sqrt{n}$ in (3.18), we have $\mathbb{E}|Z_n/n|^4 \le C/n^2$ for a constant $C > 0$ and all $n \ge 1$. Markov's inequality implies

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right) = \sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{Z_n}{n}\right|^4 > \frac{1}{m^4}\right) \le \sum_{n=1}^{\infty}\frac{m^4C}{n^2} < \infty,$$

hence the Lemma of Borel-Cantelli yields

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right\}\right) = 0$$

for all $m \ge 1$. In view of (3.37) subadditivity implies the assertion. □

4 Fringe trees of patricia tries

We now use this theorem for the distribution of random fringe patricia tries.

4.1 Size of patricia tries

To calculate the distribution of fringe trees, we want to first count the number of nodes that could be the root of the fringe tree. This section is irrelevant for the most common $|\mathscr{A}| = 2$

case, as in this case, the size of the patricia trie is deterministic and given by $|\mathscr{P}_n| = 2n - 1$. This is because every internal node has outdegree 2.

In this section, we define the size of the patricia trie as the amount of internal nodes. That is an additive functional Φ_i with toll function $\varphi_i(T) := \mathbf{1}\{\text{the root of } T \text{ is an internal node}\}$. Thus, the induced toll function on tries is given as

$$\widetilde{\varphi}_i(T) = \mathbf{1}\{\text{the root of } T \text{ has more than one child}\}.$$

Adding strings only increases the number of internal nodes, so Φ_i is increasing.

Theorem 4.1. *Theorem 3.3 applies to the size $\Phi_i(T) =: |T|_i$ of patricia tries T. If $d_p = 0$, the limit for the expected value is given by*

$$\frac{|\mathscr{P}_n|_i}{n} \to H^{-1} f_E^*(-1) = \frac{J}{H} = H^{-1} \sum_{a\in\mathscr{A}} (1-p_a)|\log(1-p_a)|. \tag{4.1}$$

This result can be found in [Bou01, Prop 6]. To show the equality in (4.1), remember the notation $\Xi_j(i)$ for the i-th char in the j-th string. Because $\chi = \varphi_i(\bullet) = 0$, the formulas simplify somewhat:

$$\begin{aligned} f_{E,i}(\lambda) &= \mathbb{E}\widetilde{\varphi}(\widetilde{\mathscr{T}}_\lambda) = 1 - \mathbb{P}\left(\Xi_1(1) = \cdots = \Xi_{N_\lambda}(1)\right) \\ &= 1 - \mathbb{P}(N_\lambda = 0) - \sum_{a\in\mathscr{A}} \mathbb{P}\left(N_\lambda > 0, a = \Xi_1(1) = \cdots = \Xi_{N_\lambda}(1)\right) \\ &= 1 - e^{-\lambda} - \sum_{a\in\mathscr{A}} e^{\lambda(p_a-1)} - e^{-\lambda} \qquad (4.2) \\ &= 1 + (|\mathscr{A}|-1)e^{-\lambda} - \sum_{a\in\mathscr{A}} e^{\lambda(p_a-1)} \qquad (4.3) \end{aligned}$$

Applying the Mellin transform to (4.2) and integrating by parts yield

$$\begin{aligned} f_{E,i}^*(s) &= \int_0^\infty \left(1 - e^{-\lambda} - \sum_{a\in\mathscr{A}} e^{\lambda(p_a-1)} - e^{-\lambda}\right)\lambda^{s-1}\mathrm{d}\lambda \\ &= -\int_0^\infty \left(e^{-\lambda} - \sum_{a\in\mathscr{A}} -(1-p_a)e^{\lambda(p_a-1)} + e^{-\lambda}\right)s^{-1}\lambda^s\mathrm{d}\lambda \\ &= \frac{\Gamma(s+1)}{s}\left(-1 + \sum_{a\in\mathscr{A}} -(1-p_a)^{-s} + 1\right) \qquad (4.4) \end{aligned}$$

for $-1 < \operatorname{Re} s < 0$ and this function can be extended to $s = -1$ by

$$
\begin{aligned}
f_{E,i}^*(-1) &= \lim_{s\to -1} \frac{\Gamma(s+2)}{s} \cdot \frac{1}{s-(-1)} \left(-1 + \sum_{a\in\mathcal{A}} 1-(1-p_a)^{-s} \right) \\
&= \frac{\Gamma(1)}{-1} \frac{\mathrm{d}}{\mathrm{d}s} \sum_{a\in\mathcal{A}} 1-(1-p_a)^{-s}\Big|_{s=-1} \\
&= \sum_{a\in\mathcal{A}} (1-p_a)|\log(1-p_a)| = J \qquad (4.5)
\end{aligned}
$$

As an important constant of the size of patricia tries, we will use this J quite often from now on. We can see that in the case of only two symbols, this is the entropy H and so $\mathbb{E}[|\mathcal{T}_n|_i]n^{-1} \to 1$. That is as expected, as $\Phi_i(T) = |T|_e - 1$. For comparison, tries have $f_{E,i}^*(-1) = 1$.

Example 4.2. In the symmetric case with $|\mathcal{A}| = 3$, the mean term of the size is

$$f_{E,i,3}^*(-1) = J = 2\log\left(\frac{3}{2}\right) \approx 0.81093. \qquad (4.6)$$

That means that $\frac{\mathbb{E}[|\mathcal{P}_n|_i]}{n}$ is asymptotically oscillating around $J/H \approx 0.73814$. The second-biggest term in the Fourier series is, for $s = -1 + 2\pi/\log 3$,

$$f_{E,i,3}^*(s) = \frac{\Gamma(s)}{s+1}\left(2 - \left(\frac{2}{3}\right)^{-s}\right) \approx 1.80305\cdot 10^{-5} + 1.57181\cdot 10^{-5} i. \qquad (4.7)$$

So there are indeed oscillations, but they are numerically small.

By (4.4) and Lemma 3.8 the covariance function is

$$f_{C,i}^*(s) = \Gamma(s+1)\left(1 + \sum_{a\in\mathcal{A}} (1-p_a)^{-s} - 1\right) \qquad (4.8)$$

and $f_{C,i}^*(-1) = f_{E,i}^*(-1)$.

For the variance function (3.3) we have to consider $\widetilde{\varphi}_i(\mathcal{T}_n)\widetilde{\varphi}_i(\mathcal{T}_n^\alpha)$ for $\varepsilon \neq \alpha \in \mathcal{A}^*$. It is 1 if and only if $\widetilde{\varphi}_i(\mathcal{T}_n^\alpha)$ is 1 and there is at least one string that has not the same first char as α. So we have

$$f_{V,i}(\lambda) = \operatorname{Var}(\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda)) + 2\operatorname{Cov}(\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda), \widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda) - \widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda)) \qquad (4.9)$$

$$
\begin{aligned}
f_{V,i}(\lambda) &= f_{E,i}(\lambda) - f_{E,i}^2(\lambda) \\
&\quad + 2\sum_{b\in\mathcal{A}} (1 - e^{-(1-p_b)\lambda} - f_{E,i}(\lambda)) \sum_{\alpha\in\mathcal{A}^*} f_{E,i}(p_b p_\alpha \lambda). \qquad (4.10)
\end{aligned}
$$

Multiplying (4.10) out, we get a sum of lots of terms of the form $\pm e^{-q\lambda}, q \geq 0$, which can be treated like in f_E. The rest is left to the reader.

Using (4.3) we see that $f_{E,i}(0) = f'_{E,i}(0) = 0$, so we also have $f_{V,i}(0) = f'_{V,i}(0) = 0$, thus

$$f^*_{V,i}(s) = \frac{1}{s(s+1)} \int_0^\infty f''_{V,i}(\lambda)\lambda^{s+1} \mathrm{d}\lambda \tag{4.11}$$

for $\operatorname{Re} s < 0$ using partial integration. Multiplying (4.10) out, we get a sum of lots of terms of the form $\pm e^{-q\lambda}, q \geq 0$. To avoid having to spell each of them out every time, we show how a single one transforms and write down the final result in (4.14). Those with $q > 0$ show up in $f^*_{V,i}$ as

$$\pm \frac{1}{s(s+1)} \int_0^\infty \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} e^{-q\lambda} \lambda^{s+1} \mathrm{d}\lambda = \pm \frac{\Gamma(s+2)}{s(s+1)q^s}. \tag{4.12}$$

This has a singularity on $s = -1$. This singularity is removable because the sum of $\pm q$, which is $f'_{V,i}(0)$, is 0. Hence, the value on $s = -1$ is given by the sum of

$$\frac{\Gamma(1)}{-1} \frac{\mathrm{d}}{\mathrm{d}s} q^{-s}\Big|_{s=-1} = q \log q. \tag{4.13}$$

In the end, we have, with $\delta(q) := q \log q$ and α_1 being the first char of a string $\alpha \in \mathcal{A}$,

$$\begin{aligned} f^*_{V,i}(-1) = &\sum_{a \in \mathcal{A}} \delta(1 - p_a) - (|A| - 1)^2\, 2 \log 2 + 2(|A| - 1)\delta(2 - p_a) \\ &+ \sum_{a,b \in \mathcal{A}} -\delta(2 - p_a - p_b) \\ &+ 2 \sum_{\substack{\alpha \in \mathcal{A}^* \setminus \{\varepsilon\} \\ a := \alpha_1}} \Big[-\delta(1 - p_\alpha) + \sum_{b \in \mathcal{A}} \delta(1 - p_b) \qquad \text{(4.14a)} \\ &- (|A| - 1)^2 \delta(1 + p_\alpha) + \sum_{b \in \mathcal{A}} (|A| - 1)\delta(1 + p_\alpha(1 - p_b)) \\ &- (|A| - 1)\delta(1 - p_a + p_\alpha) + \sum_{b \in \mathcal{A}} \delta(1 - p_a + p_\alpha(1 - p_b)) \\ &+ (|A| - 1) \sum_{c \in \mathcal{A}} \delta(1 - p_c + p_\alpha) - \sum_{b,c \in \mathcal{A}} \delta(1 - p_c + p_\alpha(1 - p_b)) \Big]. \end{aligned} \tag{4.14}$$

Note that while the entire sum over α converges, this is not the case if summed up over each term. For example if one sums all α with $|\alpha| \leq k$ for $k \in \mathbb{N}$, the two terms in the row (4.14a) form a telescoping series in following fashion:

$$\begin{aligned}
&\sum_{\substack{1\le|\alpha|\le k\\ a:=\alpha_1}} \left(-\delta(1-p_a) + \sum_{b\in\mathcal{A}} \delta(1-b)\right)\\
&= - \sum_{0\le|\alpha|\le k-1} \sum_{a\in\mathcal{A}} \delta(1-p_a) + \sum_{1\le|\alpha|\le k} \sum_{b\in\mathcal{A}} \delta(1-p_b)\\
&= \left(|\mathcal{A}|^k - 1\right)J. && (4.15)
\end{aligned}$$

In the case of binary ($\mathcal{A} = \{0,1\}$) patricia tries, we already know that $\mathrm{Var}(\widetilde{\mathcal{P}}_\lambda) = \mathrm{Var}(N_\lambda - 1 \vee 0) \to \lambda$ and that thus $\psi_{V,i}(\log\lambda) = H$ from (3.17). In the binary case, summing over $1-p_b$ for $b \in \mathcal{A}$ is the same as summing over p_b, so the rows after (4.14a) also become telescoping series as in (4.15) and one can show after some calculations that the entire term then indeed is H.

4.2 Size of fringe trees

We now study the amount of strings in a random fringe subtree $\mathcal{P}_n^*$ of a patricia trie $\mathcal{P}_n$, as defined in Section 2.2. This is the size measured in the amount of leaves, noted as $|T|_e$, and is equal for trie and patricia trie. We count the number of subtrees of size $k \geq 1$ with the functional

$$\varphi_k(T) := \mathbf{1}\{|T|_e = k\}. \tag{4.16}$$

Note that subtrees of size 1 are the leaves, so $\varphi_1 = \varphi_\bullet = \widetilde{\varphi}_\bullet$. This case behaves different from the others, so we will first consider only $k \geq 2$.

While Φ_k is not increasing, it can be written as $\Phi_{\geq k} - \Phi_{\geq k+1}$, where $\varphi_{\geq k} := \mathbf{1}\{|T|_e \geq k\}$ is bounded and $\Phi_{\geq k}$ is increasing, so we can apply Theorem 3.3.

Because $\widetilde{\varphi}_k(\mathcal{T}_n) = 0$ a.s. for $n > k$ and e.g. $\mathrm{Var}\,\widetilde{\Phi}_k(\mathcal{T}_{k+2}) > 0$, we can apply Lemma 3.4 to show (3.23).

Theorem 4.3. *Theorem 3.3 applies to the amount $\Phi_k(\mathcal{P}_n)$ of fringe trees of size k in a random patricia trie. Especially, we have*

$$\frac{\Phi_k(\mathcal{P}_n) - \mathbb{E}\Phi_k(\mathcal{P}_n)}{\sqrt{\mathrm{Var}\,\Phi_k(\mathcal{P}_n)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{4.17}$$

with convergence of all [absolute] moments. In the asymmetric case ($d_p = 0$) for $k \geq 2$ we have convergence

$$\frac{\Phi_k(\mathcal{P}_n)}{n} \longrightarrow \frac{1-\rho(k)}{Hk(k-1)} \tag{4.18}$$

almost surely.

Remark 4.4. Note that the results depend on the source through $\rho(k)$ and H. This is different to fringe trees of tries, whose size only depends on the entropy, see [Jan22, Theorem 4.4].

Tries have $\frac{\Phi_k(\mathcal{T}_n)}{n} \to \frac{1}{Hk(k-1)}$, without the $1-\rho(k)$ term, that converges to 1 exponentially fast. We will later see that $(1-\rho(k))^{-1}$ is the expected amount of contracted edges in a node whose fringe tree has size k.

Proof. We have $\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda) = \mathbf{1}\{N_\lambda = k\}\varphi_p(\widetilde{\mathcal{T}}_\lambda)$, so the mean function is given by

$$\begin{aligned} f_{E,k}(\lambda) = \mathbb{E}\left[\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\right] &= \mathbb{P}(N_\lambda = k)\mathbb{E}[\varphi_p(\widetilde{\mathcal{T}}_\lambda) \mid N_\lambda = k] \\ &= \frac{\lambda^k}{k!}e^{-\lambda}\left(1 - \sum_{\alpha\in\mathcal{A}} p_\alpha^k\right) \\ &= \frac{\lambda^k}{k!}e^{-\lambda}\big(1-\rho(k)\big) \end{aligned} \tag{4.19}$$

with $\rho(k)$ as defined in (2.1). With (2.12) the Mellin transform is given by

$$f^*_{E,k}(s) = (1-\rho(k))\frac{\Gamma(k+s)}{k!} \tag{4.20}$$

$$f^*_{E,k}(-1) = \frac{1-\rho(k)}{k(k-1)}. \tag{4.21}$$

This shows (4.18) using (3.36). □

To calculate the asymptotic variance, we note that for $\varepsilon \neq \alpha \in \mathcal{A}^*$, $\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)$ and $\widetilde{\varphi}_k(\widetilde{\mathcal{T}}^\alpha_\lambda)$ cannot be 1 at the same time. If both subtrees had the same amount of strings, all strings would go to α and the root would not be in the patricia trie. Thus, we have from (3.12)

$$\begin{aligned} f_{V,k}(\lambda) &= \sum_\alpha{}^* \operatorname{Cov}(\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda), \widetilde{\varphi}_k(\widetilde{\mathcal{T}}^\alpha_\lambda)) \\ &= f_{E,k}(\lambda) - \sum_\alpha{}^* f_{E,k}(\lambda) f_{E,k}(p_\alpha\lambda) \\ &= \frac{1-\rho(k)}{k!}\lambda^k e^{-\lambda} - \sum_\alpha{}^* \left(\frac{1-\rho(k)}{k!}\right)^2 \lambda^{2k} p_\alpha^k e^{-\lambda(1+p_\alpha)} \end{aligned} \tag{4.22}$$

$$f^*_{V,k}(s) = \frac{1-\rho(k)}{k!}\Gamma(k+s) - \sum_\alpha{}^* \left(\frac{1-\rho(k)}{k!}\right)^2 p_\alpha^k \Gamma(s+2k)(1+p_\alpha)^{-s-2k} \tag{4.23}$$

$$f^*_{V,k}(-1) = \frac{1-\rho(k)}{k(k-1)} - (2k-2)!\left(\frac{1-\rho(k)}{k!}\right)^2 \sum_\alpha{}^* \frac{p_\alpha^k}{(1+p_\alpha)^{2k-1}}. \tag{4.24}$$

Now we turn to the covariances. We first calculate the asymptotic covariance between Φ_k and Φ_i. Remember the formula (3.27) for the asymptotic covariance:

$$f_{V,ki}(\lambda) = \operatorname{Cov}(\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda), \widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda)) + \operatorname{Cov}(\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda), \widetilde{\Phi}_k(\widetilde{\mathcal{T}}_\lambda) - \widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)) \tag{4.25}$$

For the first covariance we note that

$$\begin{aligned}\mathbb{E}[\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda)] &= \mathbb{E}[\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)]\mathbb{E}[\widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda) \mid \widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda) = 1] \\ &= f_{E,k}(\lambda)\mathbb{E}[\widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda)]\end{aligned} \tag{4.26}$$

because conditioning on $\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda) = 1$ means that $|\widetilde{\mathcal{T}}_\lambda|_e = k$ and $\widetilde{\mathcal{P}}_\lambda$ has no contracted edges in the root. The condition $|\widetilde{\mathcal{T}}_\lambda|_e = k$ makes the distribution equal to that of $\mathcal{T}_k$. For a fixed amount k of strings, the amount of contracted edges in the root is independent of the tree structure, so the second condition does not change the value. This is clear from the recursive definition of tries or alternatively follows from the later Lemma 4.9. The value $\mathbb{E}[\widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda)] = \mathbb{E}[|\mathcal{P}_k|_i]$ is constant. From (4.26) we then have

$$\mathrm{Cov}(\widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda), \widetilde{\Phi}_i(\widetilde{\mathcal{T}}_\lambda)) = f_{E,k}(\lambda)(\mathbb{E}[|\mathcal{P}_k|_i] - \mathbb{E}[|\widetilde{\mathcal{P}}_\lambda|_i]). \tag{4.27}$$

For the second covariance we look at the subtrees $\widetilde{\mathcal{T}}^a_\lambda$ for $a \in \mathcal{A}$. Let N^a_λ be the amount of strings starting with a. If $\widetilde{\Phi}_k(\widetilde{\mathcal{T}}^a_\lambda)$ is not zero, then $\widetilde{\mathcal{T}}^a_\lambda$ is not empty. In this case, $\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda)$ is 1 if and only if $N^a_\lambda \neq N_\lambda$ (not all strings start with a). Hence,

$$\begin{aligned}\mathbb{E}\left[\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda)\left(\widetilde{\Phi}_k(\widetilde{\mathcal{T}}_\lambda) - \widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\right)\right] &= \sum_{a\in\mathcal{A}} \mathbb{E}\left[\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda)\widetilde{\Phi}_k(\widetilde{\mathcal{T}}^a_\lambda)\right] \\ &= \sum_{a\in\mathcal{A}} \mathbb{E}\left[\mathbf{1}\{N_\lambda \neq N^a_\lambda\}\widetilde{\Phi}_k(\widetilde{\mathcal{T}}^a_\lambda)\right] \\ &= \sum_{a\in\mathcal{A}} \left(1 - e^{-(1-p_a)\lambda}\right)\mathbb{E}\widetilde{\Phi}_k(\widetilde{\mathcal{T}}^a_\lambda), \\ \mathrm{Cov}\left(\widetilde{\varphi}_i(\widetilde{\mathcal{T}}_\lambda), \widetilde{\Phi}_k(\widetilde{\mathcal{T}}_\lambda) - \widetilde{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\right) &= \sum_{a\in\mathcal{A}} \left(1 - e^{-(1-p_a)\lambda} - f_{E,i}(\lambda)\right)\mathbb{E}\widetilde{\Phi}_k(\widetilde{\mathcal{T}}^a_\lambda).\end{aligned} \tag{4.28}$$

Combining (4.27) and (4.28) and plugging the values for $f_{E,i}$ from (4.4) and $f_{E,k}$ from (4.19) in, we get,

$$\begin{aligned}f_{V,ki}(\lambda) =& \frac{1-\rho(k)}{k!}\lambda^k e^{-\lambda}\Bigg(\mathbb{E}|\mathcal{P}_k|_i \\ &- \sum_{\alpha\in\mathcal{A}^*}\Big(1 + (|\mathcal{A}|-1)e^{-p_\alpha\lambda} - \sum_{b\in\mathcal{A}} e^{-(1-p_b)p_\alpha\lambda}\Big)\Bigg) \\ &+ \sum_{a\in\mathcal{A}}\Bigg((|\mathcal{A}|-1)e^{-\lambda} - e^{-(1-p_a)\lambda} + \sum_{b\in\mathcal{A}} e^{-(1-p_b)\lambda}\Bigg) \\ &\sum_{\alpha\in\mathcal{A}^*}\frac{1-\rho(k)}{k!}(p_a p_\alpha\lambda)^k e^{-p_a p_\alpha\lambda}.\end{aligned} \tag{4.29}$$

The Mellin transform can be calculated with (2.12). This yields, summing up over $a\alpha$ instead of α and denoting the first char of α with α_1,

$$\begin{aligned}
f^*_{V,ki}(-1) =& \frac{1-\rho(k)}{k(k-1)}\Bigg[\mathbb{E}|\mathcal{P}_k|_i \\
&- \sum_{\alpha\in\mathcal{A}^*}\Bigg(1+\frac{|A|-1}{(1+p_\alpha)^{k-1}} - \sum_{b\in\mathcal{A}}\frac{1}{(1+(1-p_b)p_\alpha)^{k-1}}\Bigg) \\
&+ \sum_{\substack{\alpha\in\mathcal{A}^*\\ \alpha\neq\varepsilon}} p_\alpha^k\Bigg(\frac{|A|-1}{(1+p_\alpha)^{k-1}} - \frac{1}{(1-p_{\alpha_1}+p_\alpha)^{k-1}} \\
&\quad + \sum_{b\in\mathcal{A}}\frac{1}{(1-p_b+p_\alpha)^{k-1}}\Bigg)\Bigg]
\end{aligned} \tag{4.30}$$

Note, that for $|\mathcal{A}| = 2$ we directly have

$$f^*_{V,ki}(-1) = f^*_{C,k}(-1) = f^*_{E,k}(-1) = \frac{1-\rho(k)}{k(k-1)} \tag{4.31}$$

because of the deterministic size and Lemma 3.8. This was the $k \geq 2$ case.

For the special case of $\varphi_1 = \varphi_\bullet = \widetilde{\varphi}_\bullet$, the number of leaves, we refer to Janson, as the amount of leaves is the same in the patricia trie and the trie. The basic idea is to replace $\widetilde{\varphi}_\bullet$ with a new toll function

$$\widetilde{\varphi}_*(T) = \sum_{a\in\mathcal{A}} \varphi_\bullet(T^a), \tag{4.32}$$

on tries which instead counts the children that are leaves. For all tries $T \neq \bullet$ the functionals $\widetilde{\Phi}_*(T)$ and $\widetilde{\Phi}_\bullet(T)$ are equal, so they have the same asymptotic properties. Because $\widetilde{\varphi}_*$ has $\widetilde{\varphi}_*(\bullet) = \chi = 0$, it is easier to handle.

We use the functions for $\widetilde{\varphi}_*$ instead of $\widetilde{\varphi}_1$, which are

$$\psi_{E,1}(t) := \psi_{C,1}(t) := \psi_{V,1}(t) := H, \qquad \psi_{V,1i}(t) := \psi_{C,i}(t), \tag{4.33}$$

see Example 3.17, especially (3.60)-(3.62) in [Jan22].

Theorem 4.5. *The conditional fringe tree size distribution of $\mathcal{P}_n$ given $\mathcal{P}_n$ has asymptotically normal fluctuations. For $k \geq 1$ let a_{kn} be either $\mathbb{P}\left(|\mathcal{P}_n^*|_e = k\right) = \mathbb{E}\frac{\widetilde{\Phi}_k(\mathcal{T}_n)}{|\mathcal{P}_n|}$ or $\frac{\mathbb{E}\widetilde{\Phi}_k(\mathcal{T}_n)}{\mathbb{E}|\mathcal{P}_n|}$.*

i) Then, with all [absolute] moments, as $n \to \infty$,

$$\sqrt{n}\Big(\mathbb{P}\left(|\mathcal{P}_n^*|_e = k \mid \mathcal{P}_n\right) - a_{kn}\Big) = \sqrt{n}\left(\frac{\widetilde{\Phi}_k(\mathcal{T}_n)}{|\mathcal{P}_n|} - a_{kn}\right) \overset{d}{\approx} \mathcal{N}(0, \widehat{\sigma}_k^2(n)), \tag{4.34}$$

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

$$\widehat{\sigma}_k^2(n) := \frac{H}{\psi_{E,+}(t)^2}\Big(\psi_{V,k}(t) - 2\frac{\psi_{E,k}(t)}{\psi_{E,+}(t)}\psi_{V,ki}(t) + \frac{\psi_{E,k}(t)^2}{\psi_{E,+}(t)^2}\psi_{V,i}(t)\Big) - \frac{1}{\psi_{E,+}(t)^4}\big(\psi_{E,+}(t)\psi_{C,k}(t) - \psi_{E,k}(t)\psi_{C,i}(t)\big)^2 \tag{4.35}$$

ii) *In particular, if* $d_p = 0$*, then* $\widehat{\sigma}_k^2(n)$ *is constant and given by*

$$\widehat{\sigma}_k^2(n) = \frac{H}{(H+J)^4}\Big((H+J)^2 f^*_{V,k}(-1) - 2(H+J)\frac{1-\rho(k)}{k(k-1)}f^*_{V,ki}(-1) + \frac{(1-\rho(k))^2}{k^2(k-1)^2}(f^*_{V,i}(-1) - H)\Big); \qquad k \geq 2 \tag{4.36}$$

$$\widehat{\sigma}_1^2(n) = (H+J)^{-4}(H^3 f^*_{V,i}(-1) - H^2(J)^2). \tag{4.37}$$

iii) *Moreover, the approximation in distribution* (4.34) *holds jointly for any finite number of* k*, with a multivariate normal distribution* $\mathcal{N}(0, (\widehat{\sigma}_{kl}^2(n))_{k,l})$*, where* $\widehat{\sigma}_{kl}^2(n)$ *for* $k \neq l$ *can be expressed similar to* (4.35) *using polarization.*

iv) *In the aperiodic* $d_p = 0$ *case, the expected values satisfy, as* $n \to \infty$*,*

$$a_{kn} \to \begin{cases} \frac{1-\rho(k)}{k(k-1)(H+J)} & k \geq 2 \\ \frac{H}{H+J} & k = 1. \end{cases} \tag{4.38}$$

Remark 4.6. For the binary case $|\mathcal{T}_n| = 2n - 1$ is not random, so it suffices to directly use Theorem 3.3 with (4.21) and (4.24), which gives nicer formulas than (4.35) and (4.36).

Remark 4.7. We already know that the limits in (4.38) form a distribution because $f^*_{E,i} = \sum_{k=2}^{\infty} f^*_{E,k}$, but summing the limits up might make clearer where the constant J comes from. The relevant sum is

$$\begin{aligned}\sum_{k=2}^{\infty}\frac{\rho(k)}{k(k-1)} &= \sum_{a\in\mathcal{A}}\sum_{k=2}^{\infty}\frac{p_a^k}{k-1} - \frac{p_a^k}{k} \\ &= \sum_{a\in\mathcal{A}}\sum_{k=2}^{\infty}\frac{p_a^{k-1}}{k-1} - \frac{p_a^k}{k} - \frac{(1-p_a)p_a^{k-1}}{k-1} \\ &= \sum_{a\in\mathcal{A}} p_a - (1-p_a)\log(1-p_a) = 1 - J,\end{aligned} \tag{4.39}$$

using the series expansion of $\log(1-x)$.

We already have convergence of $\Phi_k(\mathscr{P}_n)$ and $|\mathscr{P}_n|$, the convergence of the quotient follows with standard methods. We repeat a lemma[Jan22, 4.8] from Janson, which states:

Lemma 4.8. *Let (X_n, Y_n) be a sequence of random vectors, such that*

$$n^{-0,5}(X_n - \mathbb{E}X_n, Y_n - \mathbb{E}Y_n) \overset{d}{\approx} \mathcal{N}\left(0, \begin{pmatrix} \sigma^2_{XX}(n) & \sigma^2_{XY}(n) \\ \sigma^2_{XY}(n) & \sigma^2_{YY}(n) \end{pmatrix}\right) \tag{4.40}$$

and $\mathbb{E}X_n = O(n), \mathbb{E}Y_n = \Theta(n)$ *and* $\sigma^2_{XX}(n), \sigma^2_{XY}(n), \sigma^2_{YY}(n) \in O(1)$.

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

$$\sqrt{n}\left(\frac{X_n}{Y_n} - \frac{x_n}{y_n}\right) \overset{d}{\approx} \mathcal{N}\left(0, \frac{n^2}{y_n^2}\left(\sigma^2_{XX}(n) - 2\frac{x_n}{y_n}\sigma^2_{XY}(n) + \frac{x_n^2}{y_n^2}\sigma^2_{Y,Y}(n)\right)\right) \tag{4.41}$$

(ii) *If, moreover,* (4.40) *holds with all moments, and $Y_n \geq cn$ a.s. for some $c > 0$ and all n, then* (4.41) *holds with all moments. Furthermore, we may then replace x_n/y_n by $\mathbb{E}[X_n/Y_n]$ in the left side of* (4.41).

Proof of Theorem 4.5. We apply Lemma 4.8 to $X_n := \Phi_k(\mathscr{P}_n), Y_n := |\mathscr{P}_n| \geq n$. A multivariate variant of Theorem 3.3 yields the condition (4.40). The variances $\sigma^2_{12}(n)$ are given by $H^{-1}\psi_{V,12}(t) - H^{-2}\psi_{C,1}(t)\psi_{C,2}(t)$ for $t = \log(n)$ according to (3.27). Putting this into (4.41) gives (4.35).

For ii), we replace the $\psi(t)$ functions with $f^*(-1)$, using also that $f^*_C(-1) = f^*_E(-1)$ according to Lemma 3.8. We then use that $f^*_{E,i}(-1) = J$ to get

$$\begin{aligned}\widehat{\sigma}^2_k(n) = \frac{H}{(H+J)^2}\Bigg(&f^*_{V,k}(-1) - 2(H+J)^{-1}\frac{1-\rho(k)}{k(k-1)}f^*_{V,ki}(-1) \\ &+ (H+J)^{-2}\left(\frac{1-\rho(k)}{k(k-1)}\right)^2 f^*_{V,i}(-1)\Bigg) \\ &- (H+J)^{-4}\left(\frac{1-\rho(k)}{k(k-1)}\right)^2 H \end{aligned} \tag{4.42}$$

for $k \geq 2$ and expanding by $(H+J)^2$ gives (4.36).

For $k = 1$, we use the functions defined in (4.33) and get

$$\begin{aligned}\widehat{\sigma}^2_k(n) &= \frac{H}{(H+J)^2}\left(H - 2(H+J)^{-1}HJ + (H+J)^{-2}H^2 f^*_{V,i}(-1)\right) \\ &\quad - (H+J)^{-4}(2H^2 - H^2)^2 \\ &= \frac{1}{(H+J)^4}\left(H^2(H+J)^2 - 2(H+J)H^2J - H^4 + H^3 f^*_{V,i}(-1)\right) \\ &= \frac{1}{(H+J)^4}\left(H^3 f^*_{V,i}(-1) - H^2(J)^2\right) \end{aligned} \tag{4.43}$$

□

4.3 Distribution of fringe patricia tries

We now turn to the tree structure of the fringe patricia tries. We first ignore the contracted edges. Let $T \in \mathfrak{T}$ be a fixed tree. We count the fringe trees equal to T with the toll function

$$\varphi(T') = \mathbf{1}\{T = T' \text{ as trees}\}. \tag{4.44}$$

Again, we exclude the case $T = \bullet$, which is φ_1 from the last section.

Let $k := |T|_e$. Similarly to Φ_k, we can write Φ_T as the difference of the additive functionals $\Phi_T + \Phi_{>k}$ and $\Phi_{>k}$, each with bounded toll functions. Again, $\varphi_T(\mathscr{P}_n) = 0$ a.s. for $n > k+1$, so Lemma 3.4 applies, and we have Var $\Phi_k(\mathscr{P}_n) \in \Omega(n)$. So all statements in Theorem 3.3 also apply to $\Phi_T(\mathscr{P}_n)$. We start by a general lemma about patricia tries.

Lemma 4.9. *Let T be a fixed labelled tree where no node has outdegree 1 and $k := |T|_e$. Let $T_e \subset \mathscr{A}^*$ be its leaves and T_i be its internal nodes.*

i) The probability $p_T := \mathbb{P}(\mathscr{P}_k = T)$ of a random patricia trie of size k to be T is given by

$$p_T = k! \prod_{v \in T_e} p_v \prod_{w \in T_i} \frac{1}{1 - \rho(|T^w|_e)}. \tag{4.45}$$

ii) Conditioned on $\mathscr{P}_k = T$, the contracted edges per node are independent. In an internal node $v \in T_i$, their distribution is given by $q_{|T^v|_e}$, with q_i, $i \geq 2$ defined as

$$q_i(\{\alpha\}) := p_\alpha^i (1 - \rho(i)); \; \alpha \in \mathscr{A}^*. \tag{4.46}$$

Since the distribution of contracted edges in a node v is only dependent on the fringe subtree T^v, this lemma also applies to their distribution in a random fringe tree $\mathscr{P}_n^*$.

Because $\mathbb{E}\widetilde{\varphi}_T(\mathscr{T}_k)$ is the probability of pat$(\mathscr{T}_k)$ to be T as a tree and have no contracted edges in the root, $\mathbb{E}\widetilde{\varphi}_T(\mathscr{T}_k) = p_T(1 - \rho(k))$. This shows that

$$\begin{aligned} f_{E,T}(\lambda) &= \mathbb{P}\Big(\widetilde{\varphi}_T(\widetilde{\mathscr{T}}_\lambda) = 1 \Big| |\widetilde{\mathscr{T}}_\lambda|_e = k, \varphi_p(\widetilde{\mathscr{T}}_\lambda) = 1\Big)\mathbb{P}\Big(\widetilde{\varphi}_k(\widetilde{\mathscr{T}}_\lambda) = 1\Big) \\ &= \frac{\mathbb{P}(\widetilde{\varphi}_T(\mathscr{T}_k) = 1)}{\mathbb{P}(\varphi_p(\mathscr{T}_k))} f_{E,k}(\lambda) \\ &= p_T f_{E,k}(\lambda) \end{aligned} \tag{4.47}$$

holds, and therefore we have $f_{E,T}^*(s) = p_T f_{E,k}^*(s)$. We can similarly derive $f_{V,T}$ from $f_{E,T}$ as in (4.22), and we get

$$f_{V,T}^*(-1) = \frac{p_T(1-\rho(k))}{k(k-1)} - (2k-2)!\left(\frac{p_T(1-\rho(k))}{k!}\right)^2 \sum_\alpha{}^* \frac{p_\alpha^k}{(1+p_\alpha)^{2k-1}}. \tag{4.48}$$

Because of Lemma 4.9, the same holds if we fix a patricia trie T with contracted edges and define $\varphi_T(T') = \mathbf{1}\{T = T'\}$ as equality of T and T' with tree structure as well as contracted edges. We then also define $p_T := \mathbb{P}(\mathscr{P}_k = T)$ with equality also in contracted edges.

Theorem 4.10. *Theorem 3.3 applies to the amount $\Phi_T(\mathscr{P}_n)$ of fringe trees equal to a fixed labelled tree T in a random patricia trie n, so we have central limit theorems* (3.18) *for $\Phi_T(\mathscr{P}_n)$ with all moments and with periodic fluctuations. Asymptotics of the mean are given in* (3.25) *with $f_{E,T}$ as in* (4.47).

Theorem 4.11. *Let T be a fixed labelled tree, or a fixed patricia trie with contracted edges. The fringe tree distribution of $\mathscr{P}_n$ satisfies*

$$\mathbb{P}(\mathscr{P}_n^* = T \mid \mathscr{P}_n) - \frac{\psi_{E,T}(\log n)}{\psi_{E,i}(\log n) + H} \xrightarrow{\mathbb{P}} 0. \tag{4.49}$$

In the asymmetric case $d_p = 0$, this limit is given by

$$\mathbb{P}(\mathscr{P}_n^* = T \mid \mathscr{P}_n) \xrightarrow{\mathbb{P}} \begin{cases} \frac{p_T(1-\rho(k))}{(H+J)k(k-1)}, & |T|_e \geq 2, \\ \frac{H}{J+H}, & T = \bullet. \end{cases} \tag{4.50}$$

Before we state the equivalent of Theorem 4.5, we note $f_{V,Ti}$ is almost the same as $p_T f_{V,ki}$ (see (4.30)), but with the constant $\mathbb{E}[|\mathscr{P}_k|_i]$ replaced by $|T|_i$.

Theorem 4.12. *The fringe tree distribution of $\mathscr{P}_n$ has asymptotically normal fluctuations, in the following sense. Let T be a fixed tree and let either $a_{T,n} = \mathbb{P}(\mathscr{P}_n^* = T)$ or $a_{T,n} = \frac{\mathbb{E}\Phi_T(\mathscr{P}_n)}{\mathbb{E}|\mathscr{P}_n|}$. Then, with all moments, as $n \to \infty$,*

$$\sqrt{n}\Big(\mathbb{P}(\mathscr{P}_n^* \mid \mathscr{P}_n) - a_{T,n}\Big) = \sqrt{n}\left(\frac{\Phi_T(\mathscr{P}_n)}{|\mathscr{P}_n|} - a_{T,n}\right) \overset{d}{\approx} \mathscr{N}(0, \widehat{\sigma}_T^2(n)), \tag{4.51}$$

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

$$\begin{aligned} \widehat{\sigma}_T^2(n) :=& \frac{H}{\psi_{E,+}(t)^2}\left(\psi_{V,T}(t) - 2\frac{\psi_{E,T}(t)}{\psi_{E,+}(t)}\psi_{V,Ti}(t) + \frac{\psi_{E,T}(t)^2}{\psi_{E,+}(t)^2}\psi_{V,i}(t)\right) \\ &+ \frac{1}{\psi_{E,+}(t)^4}\big(\psi_{E,+}(t)\psi_{C,T}(t) - \psi_{E,T}(t)\psi_{C,i}(t)\big)^2 \end{aligned} \tag{4.52}$$

Moreover, the approximation in distribution (4.51) *holds jointly for any finite number of T, with a multivariate normal distribution $\mathscr{N}(0, (\widehat{\sigma}_{TT'}^2(n))_{T,T'})$, where $\widehat{\sigma}_{TT'}^2(n)$ for $T \neq T'$ can be expressed similar to* (4.52) *using polarization.*

Proof of Lemma 4.9. The defining property of a trie is that there is exactly one string that has a leaf as a prefix. So for a valid trie T', the probability of a random trie of size k to be T' is $p_{T'} := k! \prod_{v \in T_e} p_v$.

Having no node of outdegree 1 makes T a valid patricia trie. The probability p_T is the sum of the $p_{T'}$ of the tries $T' \in p^{-1}(T)$ that are given by assigning contracted edges to the inner nodes.

Fixing contracted edges $c_v \in \mathscr{A}^*$ for every inner node $v \in T_i$ gives us a unique trie T'. Let $w \in T$ be a leaf in the patricia trie. Then the path of the corresponding leaf $\widetilde{w} \in T'$ in the trie consists of the path of w and all contracted edges c_v in the nodes $v \in T$ on the path of w to the root. Thus, the probabilities $p_{\widetilde{w}}$ for the leaves $\widetilde{w}$ in T' are given by p_w for the corresponding leaf w in T times p_{c_v} for the contracted edges on each inner node v on the way to the root. Each term p_{c_v} gets multiplied as often as there are leaves in T^v. So the probability $p_{T'}$ is given by

$$\begin{aligned} p_{T'} &= k! \prod_{w \in T_e} p_w \prod_{v \in T_i} p_{\alpha_v}^{|T^v|_e} \\ &= k! \prod_{w \in T_e} p_w \prod_{v \in T_i} \frac{1}{1-\rho(|T^v|_e)} \prod_{v \in T_i} q_{|T^v|_e}(\{\alpha_v\}) \end{aligned}$$

If we show that the q_i for $i \geq 2$ really are distributions and sum up to 1, this already shows the Lemma. Indeed,

$$\sum_{\alpha \in \mathscr{A}^*} q_i(\{\alpha\}) = (1-\rho(i)) \sum_{\alpha \in \mathscr{A}^*} p_\alpha^i = (1-\rho(i)) \sum_{n=0}^{\infty} \rho(i)^n = 1,$$

with the last sum converging because $\rho(i) < 1$ for $i > 1$. □

4.4 Bucket patricia tries

A bucket patricia trie (see Section 2.5) is a patricia trie where leaves have up to $b \geq 1$ strings. There is a straightforward connection between the amount of buckets and fringe tree sizes.

For bucket size $k \geq 1$, let $\Phi_{b,k}$ be the amount of buckets for bucket size k. For $b = 1$, the buckets are just the leaves in the patricia trie, so $\Phi_{b,1} = \Phi_\bullet$. If one now increases the bucket size to b, the nodes with b strings become buckets, and their children (which must have already been buckets in $b-1$) are not included anymore. So

$$\Phi_{b,b}(T) = \Phi_{b,b-1}(T) - (a_b(T) - 1)\Phi_k(T), \tag{4.53}$$

where $a_b(T)$ is the average amount of children a node with k strings in T has. This is always 2 in the binary case.

But even in the general case, $a_b(\widetilde{\mathscr{P}}_\lambda)$ is independent of $\Phi_k(\widetilde{\mathscr{P}}_\lambda)$ and has expected value independent of λ. We define

$$E_b := \mathbb{E}[a_b(\widetilde{\mathscr{P}}_\lambda) \mid \Phi_k(\widetilde{\mathscr{P}}_\lambda) > 0] - 1 = \mathbb{E}[|\mathscr{P}_k \cap \mathscr{A}|] - 1. \tag{4.54}$$

With this, we have for the expected value

$$f_{E,b,b}(\lambda) = f_{E,*}(\lambda) - \sum_{k=2}^{b} E_k f_{E,k}(\lambda) \tag{4.55}$$

using the usual substitution of φ_* for $\varphi_\bullet$. Buckets have no buckets as descendants, so we note that the variance is given by

$$f_{V,b,b}(\lambda) = f_{E,b,b}(\lambda) - \sum_{\alpha}^{*} f_{E,b,b}(\lambda) f_{E,b,b}(p_\alpha \lambda). \tag{4.56}$$

The functional $\Phi_{b,b}$ is increasing and suffices Lemma 3.4, so we can apply Theorem 3.3 on it. The amount of buckets and their size in a patricia trie and trie are the same, so this is basically the same analysis as for tries, see Janson's Section 4.7[Jan22] for an analysis that discerns buckets based on how many strings they contain.

The internal nodes of the bucket patricia trie are the nodes of the patricia trie with more than b strings, so the amount is given by $\Phi_{>b} = \sum_{k=b+1}^{\infty} \Phi_k$. This quickly gives

$$f_{E,>b}(\lambda) = \sum_{k=b+1}^{\infty} f_{E,k}(\lambda) \tag{4.57}$$

$$f_{V,>b}(\lambda) = \sum_{\alpha}^{*} (1 - f_{E,>b}(\lambda)) f_{E,>b}(p_\alpha \lambda), \tag{4.58}$$

and the exact terms can be calculated using (4.19) for $f_{E,k}(\lambda)$.

5 Other additive functionals

The general theorem can also be applied on other additive functionals not directly related to fringe trees, for example independence number or the number of k-protected nodes.

5.1 k-protected nodes

A node in a tree is called k-protected if the minimum distance to a descendant that is a leaf is at least k. The 1-protected nodes are therefore the internal nodes.

For $k \geq 2$ let $\Phi_{k\text{-prot.}}$ be the additive functional counting k-protected nodes. Adding a leaf can make its up to k nearest ancestors lose their protection, so $\Phi_{k\text{-prot.}}$ is not increasing, but $\Phi_{k\text{-prot.}} + k\Phi_\bullet$ is. Thus, we can apply Theorem 3.3.

Theorem 5.1. *The central limit theorem Theorem 3.3 applies to the amount $\Phi_{k\text{-prot.}}(\mathscr{P}_n)$ of k-protected nodes in the random patricia trie.*

For $k \geq 2$ a node is k-protected if and only if all of its children are $k-1$-protected. This leads us to following recursive equation:

$$\begin{aligned} f_{E,k\text{-prot.}}(\lambda) &= \prod_{a\in\mathcal{A}} \left(\mathbb{E}\varphi_{(k-1)\text{-prot.}}(\widetilde{\mathscr{P}}_{p_a}\lambda) + e^{-p_a\lambda}\right) \\ &\quad - e^{\lambda} - \sum_{a\in\mathcal{A}} \mathbb{E}\varphi_{(k-1)\text{-prot.}}(\widetilde{\mathscr{P}}_{p_a}\lambda) e^{-(1-p_a)\lambda} \qquad (5.1) \\ &= \sum_{\substack{S\subseteq\mathcal{A}\\|S|\geq 2}} e^{-\sum_{b\notin S} p_b\lambda} \prod_{a\in S} \mathbb{E}\varphi_{(k-1)\text{-prot.}}(\widetilde{\mathscr{P}}_{p_a}\lambda) \qquad (5.2) \end{aligned}$$

For $k = 2$, the 2-protected nodes are the internal nodes that have no leaves as children. We can thus use the inclusion-exclusion-principle to count nodes that have children that are leaves at $C \subseteq \mathcal{A}$ places. Having multiple children makes the node automatically internal in patricia tries. Let $\varphi_* = \widetilde{\varphi}_*$ be the toll function counting children that are leaves as in (4.32). Then

$$f_{E,2\text{-prot.}}(\lambda) = f_{E,i}(\lambda) - f_{E,*}(\lambda) + \sum_{\substack{C\subseteq\mathcal{A}\\|C|\geq 2}} \prod_{a\in C} (-p_a\lambda) e^{-p_a\lambda}. \qquad (5.3)$$

This gives us the following Mellin transform:

$$f^*_{E,2\text{-prot.}}(s) = f^*_{E,i}(s) - f^*_{E,*}(s) + \sum_{\substack{C\subseteq\mathcal{A}\\|C|>1}} \frac{\Gamma(s+|C|)}{(\sum_{a\in C} p_a)^{s+|C|}} \prod_{a\in C} (-p_a), \qquad (5.4)$$

with the value on $s = -1$ given by

$$f^*_{E,2\text{-prot.}}(-1) = J - H + \sum_{\substack{C\subseteq\mathcal{A}\\|C|>1}} (|C|-2)! \Big(\sum_{a\in C} p_a\Big)^{1-|C|} \prod_{a\in C} (-p_a) \qquad (5.5)$$

according to (4.5), (4.33) and (2.13).

Example 5.2. If $\mathcal{A} = \{0, 1\}$, then

$$f^*_{E,2\text{-prot.}}(-1) = H - H + 1 \cdot p_0 p_1 = p_0 p_1. \qquad (5.6)$$

In the symmetric case, this is $\frac{1}{4}$. By analogue of Theorem 4.5, the proportion of 2-protected nodes is oscillating around

$$\frac{f^*_{E,2\text{-prot.}}(-1)}{J+H} = \frac{1}{8\log 2} \approx 0.18034. \qquad (5.7)$$

This is less than in tries (as every merged node is 2-protected), where it is $(1.25 - \log 2)/(1 + \log 2) \approx 0.32888$[Jan22]. In binary search trees the proportion converges to $\frac{11}{30} = 0.3\overline{6}$

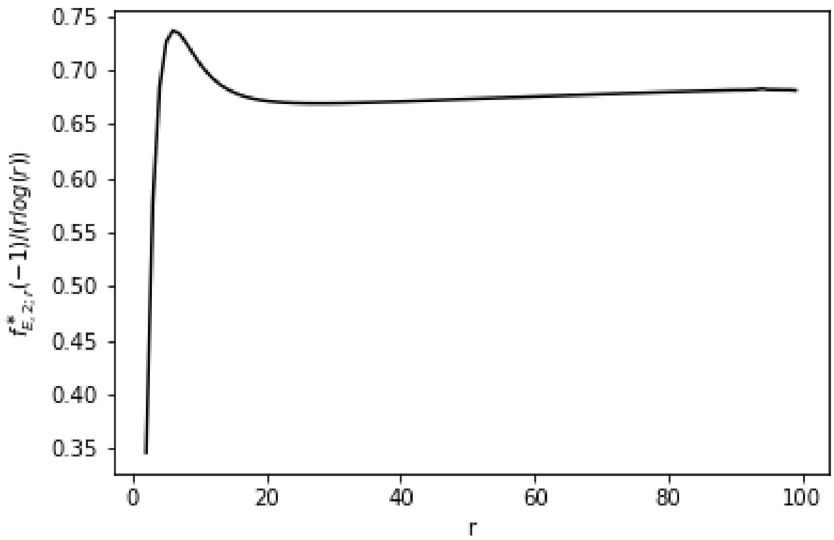

Figure 3: Values of $\frac{f^*_{E,2\text{-prot.}}(-1)}{r\log(r)}$ in the symmetric case with $r := |\mathscr{A}|$.

[MW12], and in uniformly random binary trees to $33/64 = 0.515625$ [DJ14].

By the inclusion-exclusion-principle, the amount of 2-protected nodes in a binary patricia trie T is $\Phi_i(T) - \Phi_*(T) + \Phi_2(T)$, with the first two being exactly 1 apart for $|T|_e > 1$. The variance is thus asymptotically the variance of Φ_2, given in (4.24). In the symmetric case, we hence have

$$\begin{aligned} f^*_{V,2\text{-prot.}}(-1) &= \frac{1}{4} - 2\Big(\frac{1}{4}\Big)^2 \sum_{\alpha}^{*} \frac{4^{-|\alpha|}}{(1+2^{-|\alpha|})^3} \\ &= \frac{1}{4} - \frac{1}{4}\sum_{k=0}^{\infty} \frac{4^{-k}}{(1+2^{-k})^3} + \frac{1}{32} \\ &\approx 0.28038. \end{aligned} \tag{5.8}$$

Example 5.3. In the symmetric case with $r := |\mathscr{A}|$, (5.5) becomes

$$f^*_{E,r}(-1) = (r-1)\log\Big(\frac{r}{r-1}\Big) - \log(r) + \sum_{k=2}^{r} (-1)^k \binom{r}{k} \frac{(k-2)!}{rk^{k-1}}. \tag{5.9}$$

We do not know the asymptotics of that last sum, but we can compare it to the trie. The same principle as in (5.3) can be applied to tries, with $f_{E,i}(\lambda)$ replaced by $1 - e^{\lambda} - \lambda e^{\lambda}$, the function counting internal nodes in the trie. The Mellin transform of this function at $s = -1$ is 1, see [Jan22, (4.5)]. Let $\tilde{f}_{E,r}$ be the f_E-function for 2-protected nodes in the trie. Then we have

$$f^*_{E,r}(-1) = \tilde{f}^*_{E,r}(-1) - 1 + J. \tag{5.10}$$

Janson shows that $\tilde{f}^*_{E,r}(-1) \sim \frac{1}{2r}$[Jan22, 4.15], but the second term has $1 - J \sim \frac{1}{2r}$, too, what leaves us with

$$f^*_{E,r}(-1) = o\Big(\frac{1}{r}\Big). \tag{5.11}$$

This shows that the 2-protected nodes in tries are asymptotically dominated by those which have only one child and therefore do not show up in the patricia trie. Figure 3 suggests that $f^*_{E,r}(-1)$ could be $\Theta\big((r \log r)^{-1}\big)$.

For $k > 2$, we need to calculate $\mathbb{E}\varphi_{(k-1)\text{-prot.}}(\widetilde{\mathscr{P}}_\lambda)$ in (5.1). By conditioning on the contracted edges α in the root (the common prefix of all strings), we get

$$\mathbb{E}\varphi_{(k-1)\text{-prot.}}(\widetilde{\mathscr{P}}_\lambda) = \sum_{\alpha \in \mathscr{A}^*} e^{-(1-p_\alpha)\lambda} f_{E,(k-1)\text{-prot.}}(p_\alpha \lambda). \tag{5.12}$$

Putting this into (5.1) gives us multiple sums over $\alpha \in \mathscr{A}^*$ even for $k = 3$, so we do not pursue this further.

5.2 Independence and matching number

Matching number and independence number are two similar measures of graphs, which as additive functionals have toll functions defined by similar recursion equations.

Let $G = (V, E)$ be a graph. A *matching* is a set $M \subset E$ such that every node $v \in V$ has at most one edge $e \in M$ with $v \in e$. An *independent set* is a set $S \subset V$ such that no vertices in S are neighbors. *Matching number* $\nu(G)$ and *(node) independence number* $\alpha(G)$ are the maximum cardinalities of matchings respectively independent sets.

On a tree T, these functionals can be calculated recursively. Let $\Phi(T)$ be either $\alpha(T)$ or $\nu(T)$. Both independence number and matching number only grow by at most 1 after adding a node. For every subtree $T^\beta, \beta \in \mathscr{A}^*$ we can check if $\varphi(T^\beta) := \Phi(T^\beta) - \Phi(T^\beta \setminus \{\beta\}) \in \{0, 1\}$ is one, that is if the root is necessary in a maximum matching / independent set. Such nodes are called *essential*.

In the case of independent set, the root can only be added if all of its children are not in the set, that is if all children are not essential. For a matching, the root can only be matched if some child is not essential. These conditions are just the inverse of each other, so we can see that $\varphi_\alpha + \varphi_\nu = 1$ and $\alpha(T) + \nu(T) = |T|$. Therefore, we can now mainly consider independence number. For a more detailed description see e.g. [Jan20].

The function φ is already the toll function for Φ, and it is bounded. Furthermore, Φ is increasing, so we can apply Theorem 3.3:

Theorem 5.4. *The central limit theorem Theorem 3.3 applies to the independence number α and the matching number ν of random tries $\mathscr{T}_n$ and of random patricia tries $\mathscr{P}_n$. We have for*

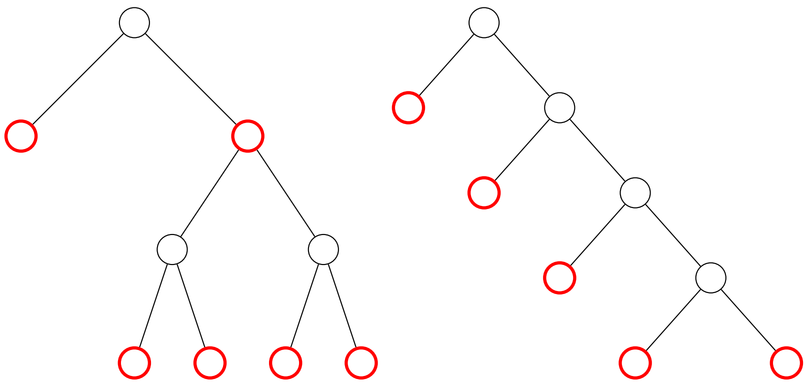

Figure 4: Two patricia tries of size 5, with different independence numbers. (independent set marked in red) φ_α is zero for both, thus showing that $\alpha(\mathscr{P}_5)$ is random even when conditioned on $\varphi_\alpha(\mathscr{P}_5) = 0$.

example

$$\frac{\alpha(\mathscr{P}_n) - \mathbb{E}[\alpha(\mathscr{P}_n)]}{\sqrt{n}} \overset{d}{\approx} \mathscr{N}(0, \widehat{\sigma}^2_\alpha(n)) \tag{5.13}$$

$$\frac{\nu(\mathscr{P}_n) - \mathbb{E}[\nu(\mathscr{P}_n)]}{\sqrt{n}} \overset{d}{\approx} \mathscr{N}(0, \widehat{\sigma}^2_\nu(n)) \tag{5.14}$$

with constant ($d_p = 0$) or at least bounded functions $\widehat{\sigma}^2_\alpha(n), \widehat{\sigma}^2_\nu(n)$.

For Theorem 3.3iv) we have to show that Var $\alpha(\mathscr{P}_n) = \Omega(n)$, but we cannot use Lemma 3.4 since $\widetilde{\varphi}_\alpha(\mathscr{T}_n)$ is not deterministic for big n. We use Lemma 3.5 with $\varphi' = \widetilde{\varphi}_\alpha$. The essentiality φ_α of a node only depends on the essentiality of its children, so if we know the essentiality of the buckets in the bucket trie, we know the essentiality of its internal children. What is left is to find a $b \geq 2$, such that Var$(\alpha(\widetilde{\mathscr{P}}_b) \mid \varphi_\alpha(\widetilde{\mathscr{P}}_b)$ is not always zero. This is the case for $b = 5$: See Figure 4 for the two [patricia] tries of size 5 with different independence numbers and same φ_α. Thus, Var $\alpha(\mathscr{P}_n) = \Omega(n)$, and we also have

$$\frac{\alpha(\mathscr{P}_n) - \mathbb{E}[\alpha(\mathscr{P}_n)]}{\sqrt{\text{Var}(\alpha(\mathscr{P}_n))}} \overset{d}{\to} \mathscr{N}(0, 1), \tag{5.15}$$

$$\frac{\nu(\mathscr{P}_n) - \mathbb{E}[\nu(\mathscr{P}_n)]}{\sqrt{\text{Var}(\nu(\mathscr{P}_n))}} \overset{d}{\to} \mathscr{N}(0, 1). \tag{5.16}$$

Note that the same argument also holds for tries, so we can replace $\mathscr{P}_n$ with $\mathscr{T}_n$ in (5.15) and (5.16).

We already know that $\frac{\alpha(\mathscr{P}_n)}{n}$ converges to $H^{-1} f^*_{E,\alpha}(-1) + 1$ a.s. in the asymmetric case according to Theorem 3.9, but we cannot calculate $f_{E,\alpha}$ like before. This is because φ_α basically depends on the entire trie structure and the recursion only relates $f_{E,\alpha}(\lambda)$ to $f_{E,\alpha}(p_\alpha \lambda)$ for

$a \in \mathscr{A}$. E.g. for the symmetric, binary case, we have for tries

$$\begin{aligned}\tilde{f}_{E,a}(\lambda) &= \mathbb{E}[\varphi_a(\widetilde{\mathscr{T}}_\lambda)] - \mathbb{P}(N_\lambda = 1)\\ &= \mathbb{E}[(1-\varphi_a(\widetilde{\mathscr{T}}^0_\lambda))(1-\varphi_a(\widetilde{\mathscr{T}}^1_\lambda))] + \mathbb{P}(N_\lambda = 1) - \mathbb{P}(N_\lambda = 1)\\ &= \left(1 - \tilde{f}_{E,a}\left(\frac{\lambda}{2}\right) - \frac{\lambda}{2}e^{-\lambda/2}\right)^2. \end{aligned} \tag{5.17}$$

We guess that $f_{E,a}(\lambda)$ itself already has $\log(\lambda)$-periodic oscillations.

We can nevertheless approximate $f_{E,a}$ by conditioning on the size of fringe trees and calculate $a_n := \mathbb{E}[\varphi_a(\mathscr{P}_n)]$ for $0 \le n \le N$ until some big integer $N \in \mathbb{N}$. Then $\mathbb{E}\left[a(\mathscr{P}_n)\right]$ can be bounded by

$$0 \le \mathbb{E}\left[a(\mathscr{P}_n)\right] - \left(n + \sum_{k=2}^{N} \mathbb{E}[\varphi_a(\mathscr{P}_k)]\mathbb{E}\left[\Phi_k(\mathscr{P}_n)\right]\right) \le \mathbb{E}\left[\Phi_{>N}(\mathscr{P}_n)\right]. \tag{5.18}$$

Using the results from Section 4.2, we can use the simultaneous convergence of Φ_i and $\Phi_k; 2 \le k \le N$ in the asymmetric case. For $\Phi_{>N}$ we use $\Phi_{>N} = \Phi_i - \sum_{k=2}^{N} \Phi_k$ and the fact that the limits of $\mathbb{E}[\Phi_k]; k \ge 2$ sum up to the limit of $\mathbb{E}[\Phi_i]$ and get:

$$\begin{aligned} 0 \le \lim_{n\to\infty} \frac{\mathbb{E}\left[a(\mathscr{P}_n)\right]}{n} - \left(1 + \sum_{k=2}^{N} \frac{(1-\rho(k))a_k}{k(k-1)H}\right) &\le \sum_{k=N+1}^{\infty} \frac{1-\rho(k)}{k(k-1)H}\\ &\le \frac{1}{NH}. \end{aligned} \tag{5.19}$$

We could also divide by $|\mathscr{P}_n|$ and get statements about the proportion of essential nodes, as in Theorem 4.5. This approach to approximate the mean can also be used for other bounded additive functionals. For the symmetric, binary case, the a_n are given by

$$a_n = \prod_{k=1}^{n-1} \binom{n}{k} \frac{(1-a_k)(1-a_{n-k})}{2^n - 2} \tag{5.20}$$

for $n \ge 2$ and $a_1 = 1, a_0 = 0$. See Figure 5 for a plot of the first values of a. Calculating these values up to $N = 800$ (higher values gave overflows in double-precision floats) gives the bounds

$$0.60225 \le \frac{f^*_{E,a}(-1)}{2H} \le 0.60316 \tag{5.21}$$

for the asymptotic mean of the proportion of essential nodes.

5.3 Amount of children

For an alphabet $\mathscr{A}$ bigger than 2, one might be interested in how many children the nodes have and count how many nodes have exactly k children, for $2 \le k \le |\mathscr{A}|$. Call this additive

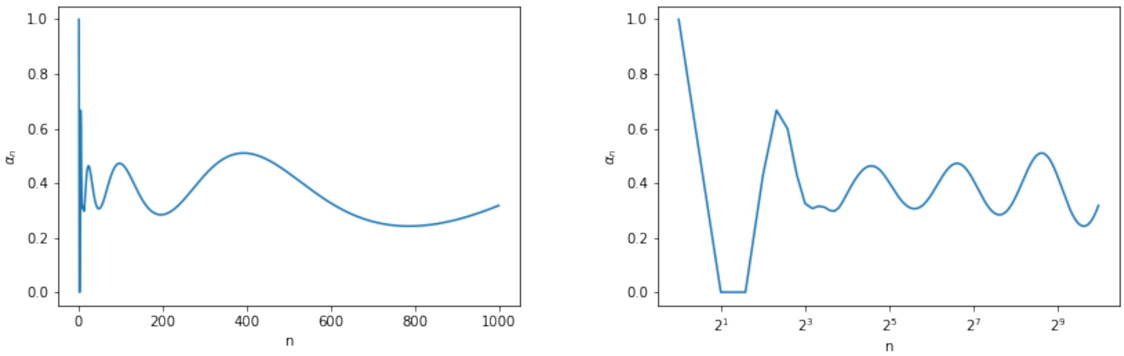

Figure 5: The first values of $\alpha_n := \mathbb{E}\varphi_\alpha(\mathscr{P}_n)$, on a normal and a logarithmic x scale, showing the oscillations.

functional Φ_k for this section. As every node with more than one child is also in the patricia trie, this is the same amount as in the trie; we have $\Phi_k = \widetilde{\Phi}_k$ and $\varphi_k = \widetilde{\varphi}_k$.

The mean function is given by

$$f_{E,k}(\lambda) = \sum_{\substack{I\subseteq\mathcal{A}\\|I|=k}} \prod_{a\notin I} e^{-p_a\lambda} \prod_{a\in I}(1-e^{-p_a\lambda}) \\ = \sum_{\substack{I\subseteq\mathcal{A}\\|I|\le k}} (-1)^{k+|I|}\binom{n-|I|}{k-|I|}\prod_{a\notin I} e^{-p_a\lambda} \tag{5.22}$$

and its Mellin transform thus by

$$f^*_{E,k}(s) = \sum_{\substack{I\subseteq\mathcal{A}\\|I|\le k}} (-1)^{k+|I|}\binom{n-|I|}{k-|I|}\Gamma(s)\prod_{a\notin I} p_a^{-s}. \tag{5.23}$$

The value at $s=-1$ can be calculated using the usual methods, and is

$$f^*_{E,k}(-1) = \sum_{\substack{I\subseteq\mathcal{A}\\|I|\le k}} (-1)^{k+|I|}\binom{n-|I|}{k-|I|}\prod_{a\notin I} p_a \log\Bigl(\prod_{a\notin I} p_a\Bigr). \tag{5.24}$$

We now explore the asymptotics of this term in the symmetric case for $n := |\mathcal{A}| \to \infty$. The term then becomes

$$f^*_{E,k}(-1) = \binom{n}{k}\sum_{j=1}^{k}(-1)^{k+j}\binom{k}{j}\Bigl(1-\frac{j}{n}\Bigr)\log\Bigl(1-\frac{j}{n}\Bigr) \tag{5.25}$$

These kinds of alternating sums with binomial coefficients are studied for example in [FS95]. The terms of the sums can be expressed as residues of a meromorphic function, which gives asymptotics by calculating or estimating terms contributed by other singularities.

If we had polynomials in above sum instead of the logarithm, following would happen: For $m \geq 0$, we have

$$\sum_{j=0}^{k} (-1)^{k+j} \binom{k}{j} (j)_m = (k)_m \sum_{j=m}^{k} (-1)^{k+j} \binom{k-m}{j-m} = (k)_m \delta_{km}, \tag{5.26}$$

so to find the sum for a polynomial, we have to find the coefficient of $(j)_k$ in the basis $((j)_k)_{k\geq 0}$. Also, this term is always zero for polynomials of degree lower than k. For j^k it is 1, and for j^{k+1}, it is $\binom{k}{2}$, since $(j)_{k+1} = j^{k+1} - \binom{k}{2} j^k + O(j^{k-1})$. The coefficients of this polynome are the Stirling numbers of first kind.

For fixed k, we can use the Taylor series of $(1-x)\log(1-x) = -x + \sum_{m\geq 1} \frac{x^m}{m(m-1)}$, which converges absolutely since $\frac{j}{n} \leq \frac{k}{n} < 1$.

$$\begin{aligned}
f^*_{E,k}(-1) &= \binom{n}{k} \sum_{m\geq k} \frac{1}{m(m-1)} \sum_{j=1}^{k} (-1)^{k+j} \binom{k}{j} \left(\frac{j}{n}\right)^m \\
&= \binom{n}{k} \frac{k! n^{-k}}{k(k-1)} + \binom{n}{k} \frac{k! k(k+1)}{k(k+1)2} n^{-k-1} + O(n^{-2}) \\
&= \frac{1}{k(k-1)} - \frac{k(k-1)}{2k(k-1)} n^{-1} + \frac{1}{2} n^{-1} + O(n^{-2}) \\
&= \frac{1}{k(k-1)} + O(n^{-2})
\end{aligned} \tag{5.27}$$

Not unsurprisingly, we have asymptotically as many nodes with k children as nodes with k strings, if the amount of letters goes to infinity.

If we set $k = n$ and change the order of summation, we get

$$\begin{aligned}
f^*_{E,n}(-1) &= \frac{1}{n} \sum_{k=1}^{n} \binom{n}{k} (-1)^k k \log \frac{k}{n} \\
&= \frac{1}{n} \sum_{k=1}^{n} \binom{n}{k} (-1)^k k \log \frac{k}{\log n},
\end{aligned}$$

where we could replace n with $\log n$ because the sum over linear functions is zero for $n \geq 2$, as seen before. This sum can be estimated similarly to [FS95, Example 6]. We leave out the details and only roughly sketch the method used. The terms of the sum are the residues of

$\omega_n(s)\log(s\log n)$, with

$$\omega_n(s) = \prod_{k=1}^{n} \frac{k}{k-s}. \tag{5.28}$$

With the residue theorem, this sum can be calculated by integrating along a path that encircles $1, \dots n$ in a big circle, extending left, but avoiding the singularity at $\mathbb{R}_{\leq 0}$. Because $\omega_n(s)$ tends to zero, only a small part $\mathscr{C}_0$ around 0 is relevant. The rest is smaller than any negative power of $\log n$. This part is the path between $-\log^{-\frac{1}{2}} n \pm (\log^{-1} n)i$, at a constant distance of $\log^{-1} n$ to $\mathbb{R}_{\leq 0}$.

By Stirling's formula one has the approximation

$$\omega_n(s) = n^s\Gamma(1-s)\left(1 + O\left(\frac{\log n}{n}\right)\right) \tag{5.29}$$

which is valid uniformly over $\mathscr{C}_0$ and the integral over the error term is again small. With a change of variables $\zeta = s\log n$, with $\mathscr{D}_0 = \log n\mathscr{C}_0$, we have

$$\begin{aligned} f^*_{E,n}(-1) &\approx \frac{1}{2\pi i n}\int_{\mathscr{C}_0} n^s\Gamma(1-s)\log(s\log n)\mathrm{d}s \\ &= \frac{1}{2\pi i n\log n}\int_{\mathscr{D}_0} e^{\zeta}\Gamma\left(1-\frac{\zeta}{\log n}\right)\log(\zeta)\mathrm{d}\zeta \\ &= \frac{1}{n\log n}\sum_{m\geq 0}(-1)^m\frac{\Gamma^{(m)}(1)}{m!\log^m n}\frac{1}{2\pi i}\int_{\mathscr{D}_0} e^{\zeta}\zeta^m\log\zeta\mathrm{d}\zeta \end{aligned} \tag{5.30}$$

by expanding $\Gamma\left(1-\frac{\zeta}{\log n}\right)$ and interchanging the order of integration and summation. The path $\mathscr{D}_0$ can be extended back towards $-\infty$ with a small error term, forming the path $\mathscr{L}$. The integrals of $e^{\zeta}\zeta^m$ along $\mathscr{L}$ are known, see [WW21, 12.22].

$$\begin{aligned} \frac{1}{2\pi i}\int_{\mathscr{L}} e^{\zeta}\zeta^m\log\zeta\mathrm{d}\zeta &= \frac{d}{dm}\frac{1}{2\pi i}\int_{\mathscr{L}} e^{\zeta}\zeta^m\mathrm{d}\zeta \\ &= \frac{d}{dm}\frac{1}{\Gamma(-m)} \\ &= (-1)^m m! \end{aligned} \tag{5.31}$$

Combining (5.30) and (5.31), we get

$$f^*_{E,n}(-1) = \frac{1}{n\log n}\sum_{m\geq 0}\frac{\Gamma^{(m)}(1)}{\log^m n} + O\left(\frac{1}{n\log^R n}\right) \tag{5.32}$$

for R an arbitrary large integer.

5.4 Shape parameter

The *shape parameter* of a tree is the logarithm of the product of all fringe tree sizes, it is thus the additive functional to the toll function $\varphi(T) := \log|T|$. The shape parameter was studied by Fill [Fil96] for binary search trees, where it is the probability mass. For a fixed binary search tree T with n nodes and B_n a random binary search tree with n nodes, $\mathbb{P}(B_n = T)^{-1}$ is the shape of T. It can be shown that this parameter is smallest for balanced trees, this is why this parameter is regarded a crude measure for "shape".

The shape functional is increasing, but the toll function is not bounded. However, the proof of Theorem 3.3 shows that the assumption of boundedness can be relaxed to the moment conditions in Theorem 3.1 for $\widetilde{\varphi}$, $\widetilde{\varphi}_+$ and $\widetilde{\varphi}_-$ [Jan22, Remark 3.8] and the condition (3.30) for higher centralized moments. We can use Lemma 3.5 (with $\varphi' = \phi$ and $b = 4$) to show that $\mathrm{Var}(\Phi(\mathscr{P}_n)) = \Omega(n)$. For patricia tries $\widetilde{\mathscr{P}}_\lambda$, we have $\log|\widetilde{\mathscr{P}}_\lambda| \leq \log(2N_\lambda)$, so the moment condition is fulfilled. We still have to prove condition

Theorem 5.5. *Theorem 3.3 holds for the shape parameter on patricia tries. We have, for example,*

$$\frac{\Phi(\mathscr{P}_n) - \mathbb{E}\Phi(\mathscr{P}_n)}{\sqrt{n}} \overset{d}{\approx} N(0, \widehat{\sigma}(n)) \tag{5.33}$$

with all moments. In the binary case with $d_p = 0$, the mean converges:

$$\frac{\mathbb{E}\Phi(\mathscr{P}_n)}{\lambda} \to \frac{1}{H} \sum_{k\geq 2} \frac{(1-\rho(k))\log(2k-1)}{k(k-1)}. \tag{5.34}$$

Proof. In the binary case, the size of a tree with k leaves is $2k - 1$ deterministically. By conditioning on the amount of leaves, we have

$$f_{E,\mathrm{sh}}(\lambda) = \sum_{k\geq 2} \log(2k-1) f_{E,k}(\lambda). \tag{5.35}$$

Since the Mellin transform is basically an integral, monotone convergence gives

$$f^*_{E,\mathrm{sh}}(s) = \sum_{k\geq 2} \log(2k-1) f^*_{E,k}(s). \tag{5.36}$$

The term at $s = -1$ is then, using (4.21):

$$\sum_{k\geq 2} \frac{(1-\rho(k))\log(2k-1)}{k(k-1)}. \tag{5.37}$$

□

6 Proofs of Lemmas

Proof of Lemma 3.5. This uses the basic idea from the proof of Lemma 3.4. We condition on the bucket trie $\mathcal{T}_n' := \mathrm{buc}_b(\mathcal{T}_n)$ for bucket size b and additionally on $\varphi_\alpha'(\mathcal{T}_n^v)$ for each bucket v. Call this condition A. Using $()_i$ for the internal and $()_e$ for the external nodes, the additive functional then splits into

$$\Phi(\mathcal{T}_n) = \sum_{v\in(\mathcal{T}_n')_i} \varphi(\mathcal{T}_n^v) + \sum_{v\in(\mathcal{T}_n')_e} \Phi(\mathcal{T}_n^v), \tag{6.1}$$

where the first part is dependent on our condition A and the second consists of small patricia tries, independent of $\mathcal{T}_n'$. Now we can bound the variance by only looking at buckets with b strings and have

$$\begin{aligned}
\mathbb{E}\mathrm{Var}\,\Phi(\mathcal{T}_n) &\geq \mathbb{E}\mathrm{Var}(\Phi(\mathcal{T}_n) \mid A) \\
&= \mathbb{E} \sum_{v\in(\mathcal{P}_n')_e} \mathrm{Var}(\Phi(\mathcal{T}_n^v) \mid \varphi'(\mathcal{P}_n^v)) \\
&\geq \mathbb{E}\left[\widetilde{\Phi}_b(\mathcal{T}_n)\right]\mathbb{E}\left[\mathrm{Var}(\Phi(\mathcal{T}_b) \mid \varphi'(\mathcal{T}_b))\right],
\end{aligned} \tag{6.2}$$

using the independence of the small tries conditioned on $\mathcal{T}_n'$ and that the $\widetilde{\Phi}_b$ counts the amount of buckets of size b. We have seen in section 4.2 in (4.21) that $\mathbb{E}[\widetilde{\Phi}_b(\mathcal{T}_n)] = \Theta(n)$ and because $\mathbb{E}\left[\mathrm{Var}(\Phi(\mathcal{T}_b) \mid \varphi'(\mathcal{T}_b))\right] > 0$, we have also $\mathbb{E}[\mathrm{Var}\,\Phi(\mathcal{T}_n)] = \Omega(n)$. □

7 Conclusion and open questions

The central limit theorem on tries could be extended to patricia tries rather easily in Proposition 2.2. For most applications, including the size of the patricia trie and the size of fringe trees, the mean functions f_E were not much more complicated than in the trie case, while the variance functions including the size Φ_i often had lengthy terms. There were often sums over all strings $\alpha \in \mathcal{A}^*$, so it would make sense to study the characteristics of these sums, especially to make long terms like (4.14) more approachable.

The recursive nature of k-protected nodes and the independence number made studying the mean harder, especially the question for the asymptotic proportion of k-protected nodes for $k \to \infty$ stays open.

Another open question would be if there are other tree models where one can use this approach of comparing with the trie. Digital search trees are a lot like tries and can also be constructed from tries, but they have no correspondence of fringe trees.

References

[Bou01] Jérémie Bourdon. "Size and path length of Patricia tries: Dynamical sources context". In: *Random Structures & Algorithms* 19.3-4 (2001), pp. 289–315. DOI: https://doi.org/10.1002/rsa.10015. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.10015.

[DJ14] Luc Devroye and Svante Janson. "Protected nodes and fringe subtrees in some random trees". In: *Electronic Communications in Probability* 19.none (2014), pp. 1–10. DOI: 10.1214/ECP.v19-3048.

[Fil96] James Allen Fill. "On the distribution of binary search trees under the random permutation model". In: *Random Structures & Algorithms* 8.1 (1996), pp. 1–25. DOI: https://doi.org/10.1002/(SICI)1098-2418(199601)8:1<1::AID-RSA1>3.0.CO;2-1. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291098-2418%28199601%298%3A1%3C1%3A%3AAID-RSA1%3E3.0.CO%3B2-1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291098-2418%28199601%298%3A1%3C1%3A%3AAID-RSA1%3E3.0.CO%3B2-1.

[FS95] Philippe Flajolet and Robert Sedgewick. "Mellin transforms and asymptotics: Finite differences and Rice's integrals". In: *Theoretical Computer Science* 144.1 (1995), pp. 101–124. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(94)00281-M. URL: https://www.sciencedirect.com/science/article/pii/030439759400281M.

[Jan20] Svante Janson. "On the independence number of some random trees". In: *Electronic Communications in Probability* 25.none (2020), pp. 1–14. DOI: 10.1214/20-ECP345. URL: https://doi.org/10.1214/20-ECP345.

[Jan22] Svante Janson. "Central limit theorems for additive functionals and fringe trees in tries". In: *Electronic Journal of Probability* 27.none (2022), pp. 1–63. DOI: 10.1214/22-EJP776. URL: https://doi.org/10.1214/22-EJP776.

[JS15] Philippe Jacquet and Wojciech Szpankowski. "Algorithms and Data Structures". In: *Analytic Pattern Matching: From DNA to Twitter*. Cambridge University Press, 2015, pp. 135–154. DOI: 10.1017/CBO9780511843204.009.

[Knu73] Donald Knuth. *The Art Of Computer Programming, vol. 3: Sorting And Searching*. Addison-Wesley, 1973, pp. 391–392.

[Mor68] Donald R. Morrison. "PATRICIA—Practical Algorithm To Retrieve Information Coded in Alphanumeric". In: *J. ACM* 15.4 (Oct. 1968), pp. 514–534. ISSN: 0004-5411. DOI: 10.1145/321479.321481.

[MW12] Hosam M. Mahmoud and Mark Daniel Ward. "Asymptotic distribution of two-protected nodes in random binary search trees". In: *Applied Mathematics Letters* 25.12 (2012), pp. 2218–2222. ISSN: 0893-9659. DOI: https://doi.org/10.1016/j.aml.2012.06.005. URL: https://www.sciencedirect.com/science/article/pii/S0893965912002789.

[Szp90] Wojciech Szpankowski. “Patricia Tries Again Revisited”. In: *J. ACM* 37 (Oct. 1990), pp. 691–711. DOI: 10.1145/96559.214080.

[WW21] E. T. Whittaker and G. N. Watson. “The Gamma-Function”. In: *A Course of Modern Analysis*. Ed. by Victor H.Editor Moll. 5th ed. Cambridge University Press, 2021, pp. 243–275. DOI: 10.1017/9781009004091.021.