Check for updates

J. Appl. Probab. **63,** 334–347 (2026)
*doi:*10.1017/jpr.2025.10034

CENTRAL LIMIT THEOREMS FOR ADDITIVE FUNCTIONALS OF PATRICIA TRIES

JASPER ISCHEBECK ,* *Goethe University Frankfurt*

Abstract

General additive functionals of patricia tries are studied asymptotically in a probabilistic model with independent, identically distributed letters from a finite alphabet. Asymptotic normality is shown after normalization together with asymptotic expansions of the moments. There are two regimes depending on the algebraic structure of the letter probabilities, with and without oscillations in the expansion of moments. As applications firstly the proportion of fringe trees of patricia tries with k keys is studied, which is oscillating around $(1-\rho(k))/(2H)k(k-1)$, where H denotes the source entropy and $\rho(k)$ is exponentially decreasing. The oscillations are identified explicitly. Secondly, the independence number of patricia tries and of tries is considered. The general results for additive functions also apply, where a leading constant is numerically approximated. The results extend work of Janson on tries by relating additive functionals on patricia tries to additive functionals on tries.

Keywords: Fringe tree; additive functional; patricia trie; limit theorem; independence number

2010 Mathematics Subject Classification: Primary 60F05
Secondary 68P10; 68P05; 68W40

1. Introduction

Patricia tries are trees used in computer science as data structures to facilitate searching and sorting strings and in compression [11]. Many properties that relate to the performance of algorithms on patricia tries can be described by *additive functionals*, such as path length and size. These are functions on trees that can be written as sums over all nodes (see (2) below). These properties are typically studied in settings where the strings used to build the patricia tries arise from stochastic models. We fix a finite alphabet $\mathcal{A}$ and a distribution p on $\mathcal{A}$ and sample infinite strings with independent, identically distributed (i.i.d.) letters according to p. For $a \in \mathcal{A}$, we write p_a for the point mass of p at a. To avoid trivialities, we require $0 < p_a < 1$ for all $a \in \mathcal{A}$.

In this model, we show that when the number of strings in the patricia trie go to infinity, the values of additive functionals converge in distribution and with all moments against a normal distributed variable. The mean and variance are known to oscillate according to the algebraic structure of the letter probabilities p_a, and a first-order expansion including oscillations in terms of the contribution of each node is given. For this, we require that the contribution of each node to the additive functional is sublinear in the size of its subtree.

Received 28 June 2025; accepted 28 June 2025.
* Postal address: Robert-Mayer-Straße 10, 60325 Frankfurt am Main, Germany. Email address: ischebec@math.unifrankfurt.de

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

Some important properties of patricia tries and the related tries are included in Knuth [11], and many other properties have been studied since then. By studying those properties one by one, we can get precise asymptotic expansions, as in [2], [3], [7] and [13]. Methods and results applying to multiple properties at once have been developed by Hwang, Fuchs, and Zacharovas [5] using analytic methods, and recently by Janson [10] using more probabilistic methods. Asymptotic results for small additive functionals have been found for many other random tree models, e.g. [8] and [14] for simply generated trees/conditioned Galton–Watson trees, [1] and [14] for recursive trees and binary search trees, and [10] for tries, as mentioned above. See [4, Section 3.1.2] for an overview.

After introducing patricia tries and tries, we show that there is a pullback of additive functionals on patricia tries to additive functionals on tries. Using this pullback, we show our main theorem, Theorem 1, by applying Theorem 3.9 from Janson [10] to this pullback. Then we apply this theorem to the distribution of fringe trees, calculating the mean and variance of the number of fringe trees of a certain size. As a second application, we consider the independence number of the patricia trie and exemplify how to approximate mean and variance if they cannot be easily derived from the additive functional.

2. Preliminaries

The preliminaries are largely the same as for [10] so we will try to keep this short, and refer readers to [10] for a more detailed description. For our fixed, finite alphabet $\mathcal{A}$ we define $\mathcal{A}^* := \bigcup_{n=0}^{\infty} \mathcal{A}^n$ as the set of finite strings over $\mathcal{A}$. The elements of strings are called *characters*. We write ε for the empty string and $\alpha\beta$ for the concatenation of strings $\alpha, \beta \in \mathcal{A}^*$. We extend this notation to sets of strings in the usual manner, with $aT := \{at \mid t \in T\}$ for $T \subseteq \mathcal{A}^*$, $a \in \mathcal{A}^*$. As noted before, we have a probability measure p on $\mathcal{A}$, with point masses $p_a \in (0, 1)$, $a \in \mathcal{A}$. For a string $\alpha = (a_1, a_2, \dots, a_n) \in \mathcal{A}^*$, we define by extension $p_\alpha := p_{a_1} \dots p_{a_n}$.

This set $\mathcal{A}^*$ can naturally be identified with the nodes of the infinite $|\mathcal{A}|$-ary tree, with a node $a \in \mathcal{A}^*$ being an ancestor of a node $b \in \mathcal{A}^*$ if and only if a is a prefix of b. We call subtrees of $\mathcal{A}^*$ including the root *$\mathcal{A}$-ary trees* and associate them with their node set $T \subset \mathcal{A}^*$. For technical reasons, we also consider the empty graph $\varnothing$ as a tree. Associating the trees with their node sets forms a bijection between $\mathcal{A}$-ary trees and subsets $T \subset \mathcal{A}^*$ with the property that for all $a \in T$ all prefixes of a are also in T. We write $\mathfrak{T}$ for the set of $\mathcal{A}$-ary trees and let $\bullet := \{\varepsilon\}$ be the tree consisting only of the root.

For a set $T \subset \mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$ of finite or infinite strings and a finite string $v \in \mathcal{A}^*$, we define T^v as the strings starting with v with v removed. In exact terms, that means $T^v := \{\alpha \mid v\alpha \in T\}$. If T is a tree and v a node of T, we call T^v the *fringe tree*. It is then the subtree of T consisting of v and its descendants.

Given a finite set $\mathfrak{X} \subseteq \mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$ of (finite or infinite) strings where no string is a prefix of another, the *trie* $T(\mathfrak{X})$ is a $\mathcal{A}$-ary tree defined as follows; see [7, Section 6.1] and [11, Chapter 6.3]. The trie for the empty set is empty. The trie for a single string consists of only one leaf, the root. We say that this leaf *stores* the string. The interesting case is the last: the trie for multiple strings consists of the root, and for every character $a \in \mathcal{A}$ this root has a child with the subtree being a trie built from $\mathfrak{X}^a$, the strings starting with a. We are therefore splitting the set of strings into subtrees by their first character. Formally, we have $T(\varnothing) := \varnothing$, for $|\mathfrak{X}| = 1$ we have $T(\mathfrak{X}) := \bullet$, and for $|\mathfrak{X}| \geq 2$

$$T(\mathfrak{X}) := \{\varepsilon\} \cup \bigcup_{a \in \mathcal{A}} aT(\mathfrak{X}^a). \tag{1}$$

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

In the *patricia trie*, introduced by Morrison [12] in 1968, a node v can additionally store a string I_v as the *common prefix* of all strings that are stored in leaves under it. We write $\operatorname{lcp}(T)$ for the longest common prefix of a set $T \subset \mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$ of strings. The patricia trie for one or zero strings is the same as the trie. For multiple strings, the longest common prefix $\operatorname{lcp}(\mathfrak{X})$ is saved in an attribute I of the root, and the splitting is done on the first character that not all strings have in common. Thus, for every character $a \in A$, the root has as subtree the patricia trie built from $\mathfrak{X}^{\operatorname{lcp}(\mathfrak{X})a}$, the strings starting with $\operatorname{lcp}(\mathfrak{X})a$. Formally, we define $\operatorname{pat} T(\varnothing)$ as $\varnothing$ and $\operatorname{pat} T(\mathfrak{X}) = \bullet$ for $|\mathfrak{X}| = 1$ as for tries. For $|\mathfrak{X}| \geq 2$ we define

$$\operatorname{pat} T(\mathfrak{X}) := \{\varepsilon\} \cup \bigcup_{a \in \mathcal{A}} a \operatorname{pat} T(\mathfrak{X}^{\operatorname{lcp}(\mathfrak{X})a}).$$

Then the common prefix $\operatorname{lcp}(\mathfrak{X})$ is stored in the attribute I_ε of the root, and for every $a \in \mathcal{A}$, $\alpha \in \mathcal{A}^*$ the attribute $I_{a\alpha}$ will be the attribute I_α of $\operatorname{pat} T(\mathfrak{X}^{\operatorname{lcp}(\mathfrak{X})a})$. While the patricia trie technically consists of the tree and the map $v \mapsto I_v$ of the common prefixes, we usually regard it as a tree and ignore the additional structure when convenient.

We use $\operatorname{pat} T(\mathfrak{X})$ as suggestive notation to emphasize the fact that the patricia trie can also be constructed from the trie. For a trie T, the patricia trie $\operatorname{pat}(T)$ for the same strings is constructed by iteratively merging every node $v \in \mathcal{A}^*$ with exactly one child va for $a \in \mathcal{A}$ with its child and prepending the character a to the attribute I_v of the merged node. Conversely, by adding a parent with only one child for every character in I_v, for every node v one can construct the trie from the patricia trie. With simple induction one can show that the two definitions are actually equivalent, so that we have $\operatorname{pat}(T(\mathfrak{X})) = \operatorname{pat} T(\mathfrak{X})$.

Let $\mathcal{T}_n$ denote the random trie generated by $n \in \mathbb{N}$ i.i.d. infinite strings, where each character is i.i.d. sampled from the distribution p and $\mathcal{P}_n = \operatorname{pat}(\mathcal{T}_n)$ is the patricia trie from the same strings. We also consider a Poisson version, where we construct the trie from a $\operatorname{Poi}(\lambda)$-distributed number N_λ of strings for $\lambda \geq 0$, with N_λ independent of the strings. We define $\widetilde{\mathcal{T}}_{:\lambda} = \mathcal{T}_{N_\lambda}$ and $\widetilde{\mathcal{P}}_\lambda := \mathcal{P}_{N_\lambda} = \operatorname{pat}(\widetilde{\mathcal{T}}_\lambda)$.

The properties that we study are represented as so-called additive functionals. Given a function $\varphi\colon \mathfrak{T} \to \mathbb{R}^n$, $n \geq 1$ with $\varphi(\varnothing) = 0$, we define the corresponding *additive functional* $\Phi\colon \mathfrak{T} \to \mathbb{R}^n$ as

$$\Phi(T) = \sum_{\alpha \in \mathcal{A}^*} \varphi(T^\alpha) = \sum_{v \in T} \varphi(T^v). \tag{2}$$

This can be written recursively as

$$\Phi(T) = \varphi(T) + \sum_{a \in \mathcal{A}} \Phi(T^a), \qquad \Phi(\varnothing) = 0,$$

which also shows that every function $\Phi\colon \mathfrak{T} \to \mathbb{R}^n$ with $\Phi(\varnothing) = 0$ can be represented as an additive functional. The term 'additive functional' for Φ is thus mainly defined by its relation to φ, which is called the *toll function* of Φ. For example, if we define

$$\varphi_i(T) := \mathbf{1}\{T \text{ has more than 1 node}\},$$

then the corresponding additive functional Φ_i counts the internal nodes of a tree. For a patricia trie $\operatorname{pat} T(\mathfrak{X})$, we assume that $\Phi(\operatorname{pat} T(\mathfrak{X}))$ only depends on the patricia trie regarded as an $\mathcal{A}$-tree and ignores the additional attributes I_v.

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

We use the notation $\xrightarrow{d}$ for convergence in distribution and $\overset{d}{\approx}$ for approximation in distribution, i.e. for two sequences X_n and Y_n of random variables on the same Borel space S, $X_n \overset{d}{\approx} Y_n$ holds if and only if, for all bounded, continuous functions $f\colon S \to \mathbb{R}$,

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| \longrightarrow 0 \quad \text{as } n \to \infty. \tag{3}$$

If $S = \mathbb{R}$, we say that this approximation *holds with all moments* if (3) is also true for $f(x) = |x|^r$, $r > 0$ and $f(x) = x^s$, $s \in \mathbb{N}$. Compare Appendix B of [10]. Furthermore, we write H for the *entropy* of the source p, that is,

$$H := \sum_{a \in \mathcal{A}} p_a \log\left(\frac{1}{p_a}\right).$$

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

$$\rho(s) := \sum_{a \in \mathcal{A}} p_a^s. \tag{4}$$

For natural $s \in \mathbb{N}$, we may interpret this as the probability for s strings to start with the same character.

3. Patricia trie and trie

We now relate additive functionals on patricia tries to those on tries.

Proposition 1. *An additive functional* Φ *on patricia tries defines an additive functional* $\widehat{\Phi} := \Phi \circ \mathrm{pat}$ *on tries by pullback with* pat. *The toll function* $\widehat{\varphi}(T)$ *of* $\widehat{\Phi}$ *is given by*

$$\widehat{\varphi}(T) = \begin{cases} 0 & T\text{'s root has exactly one child}, \\ \varphi(\mathrm{pat}\,(T)) & \text{else}. \end{cases}$$

Proof. Each node $\widehat{v}$ of the patricia trie $\mathrm{pat}\,(T)$ was created by compressing nodes in the trie. Associate $\widehat{v}$ 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 do not have exactly one child.

Compressing the nodes does not change the ancestor–descendant relationship, so the fringe tree of $\widehat{v}$ consists of the compressed nodes of the fringe tree of v. Thus $\mathrm{pat}\,(T)^{\widehat{v}}$ is the same as $\mathrm{pat}\,(T^v)$, except that in $\mathrm{pat}\,(T^v)$ there is no common prefix in the root ($I_\varepsilon = \varepsilon$). Because the toll function φ is not allowed to depend on the common prefixes, we nevertheless have $\varphi(\mathrm{pat}\,(T)^{\widehat{v}}) = \widehat{\varphi}(T^v)$. Nodes with exactly one child are ignored by $\widehat{\varphi}$, so summing over all nodes yields the equation. □

Since then $\Phi(\mathcal{P}_n) = \widehat{\Phi}(\mathcal{T}_n)$, all results about additive functionals on tries can be extended to patricia tries by just exchanging the toll function φ with $\widehat{\varphi}$ if the conditions are still satisfied. There is an alternative way to write $\widehat{\varphi}$: by defining

$$\varphi_p(T) := \mathbf{1}\{T \text{ does not have exactly one child}\}, \tag{5}$$

we have $\widehat{\varphi}(T) = \varphi_p(T)\varphi(\mathrm{pat}\,(T))$ for all toll functions φ.

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

4. Limit theorems

We reformulate the main Theorem 3.9 of [10] for patricia tries. Let Φ be an additive functional on patricia tries. We call the functional *increasing* if, for every two sets $\mathfrak{X} \subseteq \mathfrak{Y} \subset \mathcal{A}^{\mathbb{N}}$ of strings, we have $\Phi(\mathrm{pat}\,(T(\mathfrak{X}))) \leq \Phi(\mathrm{pat}\,(T(\mathfrak{Y})))$. In other words, Φ is increasing if and only if $\Phi \circ \mathrm{pat} \circ T = \widehat{\Phi} \circ T$ is monotone. Note that this is not the same as Φ being increasing as an additive functional on tries as in [10]: in patricia tries, new leaves can also grow out of edges, making functionals such as 'number of edges that connect two nodes with more than three children' not increasing.

Let Φ be an additive functional on patricia tries and let $\widehat{\Phi}$ be the corresponding pullback on tries as defined in Proposition 1, with toll function $\widehat{\varphi}$. Let $\chi := \varphi(\bullet)$ be the value every leaf contributes to Φ. The leaves have to be treated as a special case in the analysis of patricia tries (and tries) because their behavior is substantially different from internal nodes. However, their number is simply n in $\mathcal{P}_n$ and N_λ in $\widetilde{\mathcal{P}}_\lambda$, and we use χ only to accommodate for this fact.

To express the asymptotics for mean and variance, we need the following functions, where the terms are much simpler for $\chi = 0$:

$$f_E(\lambda) := \mathbb{E}\big[\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big] - \chi\lambda \mathrm{e}^{-\lambda}, \tag{6}$$

$$\begin{aligned} f_V(\lambda) := {} & 2\,\mathrm{Cov}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda), \widehat{\Phi}(\widetilde{\mathcal{T}}_\lambda)\big) - \mathrm{Var}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big) \\ & + 2\chi\lambda \mathrm{e}^{-\lambda}\big(\mathbb{E}\widehat{\Phi}(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big) - \chi^2\lambda \mathrm{e}^{-\lambda}(1 - \lambda \mathrm{e}^{-\lambda}), \end{aligned} \tag{7}$$

$$f_C(\lambda) := \mathrm{Cov}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big) + \chi\lambda(\lambda - 1)\,\mathrm{e}^{-\lambda}. \tag{8}$$

In the convenient case of $\chi = 0$, this simplifies to

$$f_E(\lambda) = \mathbb{E}\big[\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big], \tag{9}$$

$$f_V(\lambda) = 2\,\mathrm{Cov}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda), \widehat{\Phi}(\widetilde{\mathcal{T}}_\lambda)\big) - \mathrm{Var}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big), \tag{10}$$

$$f_C(\lambda) = \mathrm{Cov}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda), N_\lambda\big). \tag{11}$$

Now for $X = E, V, C$ we define the Mellin transform of f_X as

$$f_X^*(s) := \int_0^\infty t^{s-1} f_X(t)\,\mathrm{d}t, \tag{12}$$

for $s \in \mathbb{C}$ where the integral converges absolutely.

For certain distributions p, there are typically oscillations in mean and variance of additive functionals. These occur if there is a number d_p such that

$$\{\log p_a \mid a \in \mathcal{A}\} \subseteq d_p\mathbb{Z}. \tag{13}$$

We then define d_p to be the biggest such number. If there is no such number, we define $d_p := 0$. An alternative characterization is to define d_p as the infimum of positive elements of the additive subgroup generated by $\{\log p_a \mid a \in \mathcal{A}\}$.

If $d_p = 0$, there are no oscillations, and we define

$$\psi_X(t) := f_X^*(-1) \tag{14}$$

as constant functions. If $d_p > 0$, we define ψ_X as bounded, continuous, d_p-periodic functions with the Fourier series

$$\psi_X(t) \sim \sum_{m=-\infty}^{\infty} f_X^*\left(-1 - \frac{2\pi m}{d_p}\mathrm{i}\right) \mathrm{e}^{2\pi \mathrm{i} m t/d_p}. \tag{15}$$

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

Moreover, if $X = E$ or if $f_X'(\lambda) = O(\lambda^{-\varepsilon_1})$ as $\lambda \to \infty$ for some $\varepsilon_1 > 0$, Theorem 3.1 of [10] shows that the Fourier series (15) converges absolutely, and thus '$\sim$' may be replaced by '$=$' in (15). Lemma 3.6 of [10] further shows that $\psi_C(t) = \psi_E(t) + \psi_E'(t)$ for all values of t, so that we actually just need to compute ψ_E and ψ_V. Typically, the values

$$f_X^*\left(-1 - \frac{2\pi m}{d_p}\mathrm{i}\right)$$

are small for $m \neq 0$, so we can still consider the mean term $f_X^*(-1)$ as 'average asymptotic value'.

With these values and definitions, we can finally state our version of Theorem 3.9 of [10].

Theorem 1. *Let φ_+, φ_- be bounded toll functions on patricia tries so that their additive functionals Φ_+, Φ_- are increasing. Then $\Phi := \Phi_+ - \Phi_-$ is also an additive functional on patricia tries. Let $\widehat{\varphi}$, $\widehat{\Phi}$ be the toll function and additive functional on tries corresponding to Φ as in Proposition 1. Then, with the functions f_X^*, ψ_X defined in (6)–(15), we have the following.*

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

$$\frac{\Phi(\widetilde{\mathcal{P}}_\lambda) - \mathbb{E}[\Phi(\widetilde{\mathcal{P}}_\lambda)]}{\sqrt{\lambda}} \overset{d}{\approx} \mathcal{N}(0, \widetilde{\sigma}^2(\lambda)), \qquad \frac{\Phi(\mathcal{P}_n) - \mathbb{E}[\Phi(\mathcal{P}_n)]}{\sqrt{n}} \overset{d}{\approx} \mathcal{N}(0, \sigma^2(n)) \tag{16}$$

with all moments, where

$$\widetilde{\sigma}^2(\lambda) = \chi^2 + H^{-1}\psi_V(\log\lambda),$$
$$\sigma^2(n) = H^{-1}\psi_V(\log n) - H^{-2}\psi_C(\log n)^2 - 2\chi H^{-1}\psi_C(\log n).$$

(ii) *The expected values satisfy*

$$\mathbb{E}[\Phi(\mathcal{P}_n)] - \mathbb{E}[\Phi(\widetilde{\mathcal{P}}_n)] \in o(\sqrt{n}), \tag{17}$$

and we may thus replace $\mathbb{E}[\Phi(\mathcal{P}_n)]$ with $\mathbb{E}[\Phi(\widetilde{\mathcal{P}}_n)]$ in (16).

(iii) *If $\operatorname{Var}\Phi(\mathcal{P}_n) \in \Omega(n)$, then*

$$\frac{\Phi(\widetilde{\mathcal{P}}_\lambda) - \mathbb{E}[\Phi(\widetilde{\mathcal{P}}_\lambda)]}{\sqrt{\operatorname{Var}\Phi(\widetilde{\mathcal{P}}_\lambda)}} \xrightarrow{d} \mathcal{N}(0,1), \qquad \frac{\Phi(\mathcal{P}_n) - \mathbb{E}[\Phi(\mathcal{P}_n)]}{\sqrt{\operatorname{Var}\Phi(\mathcal{P}_n)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{18}$$

with all moments for all values of d_p.

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

$$\frac{\mathbb{E}[\Phi(\widetilde{\mathcal{T}}_\lambda)]}{\lambda} = \frac{\psi_E(\log\lambda)}{H} + o(1), \qquad \frac{\mathbb{E}[\Phi(\widetilde{\mathcal{T}}_n)]}{n} = \frac{\psi_E(\log n)}{H} + o(1). \tag{19}$$

Remark 1. The assumption of boundedness on ϕ_+, ϕ_- can, as in [10], be weakened to show that for some $\varepsilon > 0$,

$$\mathbb{E}\big[\phi_\pm(\widetilde{\mathcal{T}}_\lambda)\big] = O(\lambda^{1-\varepsilon}), \qquad \operatorname{Var}(\phi_\pm(\widetilde{\mathcal{T}}_\lambda)) = O(\lambda^{1-\varepsilon}),$$

and that there exists some $r > 2$ such that

$$\mathbb{E}\big|\phi_\pm(\widetilde{\mathcal{T}}_\lambda) - \mathbb{E}\phi_\pm(\widetilde{\mathcal{T}}_\lambda)\big|^r = O(\lambda^{r/2}).$$

The convergence (18) then only holds for moments $s < r$; see Theorem 5.8 of [10].

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

Remark 2. Theorem 1 also applies to linear combinations of additive functionals that meet the conditions of Theorem 1, so one can use the Cramér–Wold device to show a multivariate version of Theorem 1 (see the proof of Theorem 5.6 of [10]).

From his version of Theorem 1, Janson derives a weak law of large numbers, while leaving as an open question if one also had almost sure (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 have the following strong law of large numbers for patricia tries. The same proof can be applied to tries, answering the open question.

Theorem 2. *Let φ be a bounded toll function on patricia tries as in Theorem 1. Then, as $n \to \infty$,*

$$\frac{\Phi(\mathcal{P}_n)}{n} - H^{-1}\psi_E(\log(n)) - \chi \longrightarrow 0 \quad a.s. \tag{20}$$

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

$$\frac{\Phi(\mathcal{P}_n)}{n} \longrightarrow H^{-1}f_E^*(-1) + \chi \quad a.s.$$

Proof. Note that $\mathbb{E}[\Phi(\mathcal{P}_n)]$ is asymptotically $H^{-1}\psi_E(\log\lambda) + \chi$ according to (19). So (20) is equivalent to showing that $n^{-1}Z_n := n^{-1}(\Phi(\mathcal{P}_n) - \mathbb{E}\Phi(\mathcal{P}_n)) \to 0$ a.s. Writing out the quantifiers of 'not converging', we have

$$\left\{\frac{Z_n}{n} \not\to 0\right\} = \bigcup_{m\in\mathbb{N}} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right\} = \bigcup_{m\in\mathbb{N}} \limsup_{n\to\infty} \left\{\left|\frac{Z_n}{n}\right| > \frac{1}{m}\right\}. \tag{21}$$

Since we have convergence of all moments of $Z_n/\sqrt{n}$ in (16), 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

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

hence the Borel–Cantelli lemma 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 (21), subadditivity implies the assertion. □

5. Applications

With Theorem 1, we can calculate the asymptotic distribution of random fringe trees in patricia tries. Furthermore, we show another application, the independence number, to demonstrate how to use this theorem without a closed formula for the functions f_E, f_V.

5.1. Fringe patricia tries

Our main application is the study of random fringe trees $\mathcal{P}_n^*$ of patricia tries. We want to first study the size $|\mathcal{P}_n^*|_e$, as measured in number of leaves or equivalently strings. We count the number of subtrees of size $k \ge 1$ with the functional

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

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

Note that subtrees of size 1 are the leaves, so $\varphi_1 = \varphi_\bullet = \widehat{\varphi}_\bullet$. This case behaves differently from the others and is trivial, 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 Theorems 1 and 2.

We present two ways to get results about fringe patricia tries. One way is to directly calculate $f_{E,k}, f^*_{E,k}$ etc. Another way is to directly link fringe trees on tries with those on patricia tries. For those, we first have to understand what differs from tries: the common prefixes.

Lemma 1. *Let T be a fixed $\mathcal{A}$-nary tree where no node has outdegree 1 and $k := |T|_e$. Let $T_e \subset \mathcal{A}^*$ be its leaves and let T_i be its internal nodes. The probability $p_T := \mathbb{P}(\mathcal{P}_k = T)$ of a random patricia trie of size k being T is given by*

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

Conditioned on the tree structure of $\mathcal{P}_k$ being given by T, the common prefixes 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

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

Thus the length $|I|_v$ is $\mathrm{Geom}_0\,(1 - \rho(|T^v|_e))$-distributed, where we use Geom_0 to refer to the geometric distribution with support $\mathbb{N}_0$ and Geom_1 to the one with support $\mathbb{N}^+$.

We will prove Lemma 1 at the end of this section. The trie can be reconstructed from the tree structure and the common prefixes of the patricia trie. As pat is compressing chains of nodes into the common prefixes, the inverted process basically expands those common prefixes into a chain of nodes. Every fringe patricia trie T^v of size k then corresponds to $|I_v| + 1$ fringe tries of size k. By Lemma 1, $\Phi_k(\mathcal{T}_n)$ is thus distributed as the sum of $\Phi_k(\mathcal{P}_n)$ $\mathrm{Geom}_1\,(1 - \rho(k))$-distributed, independent random variables. This immediately links mean and variance of the number of fringe trees of size k:

$$\begin{aligned}
\mathbb{E}[\Phi_k(\mathcal{T}_n)] &= \frac{1}{1 - \rho(k)} \mathbb{E}[\Phi_k(\mathcal{P}_n)], \\
\mathrm{Var}\,(\Phi_k(\mathcal{T}_n)) &= \mathbb{E}[\mathrm{Var}\,(\Phi_k(\mathcal{T}_n) \mid \Phi_k(\mathcal{P}_n))] + \mathrm{Var}\,(\mathbb{E}[\Phi_k(\mathcal{T}_n) \mid \Phi_k(\mathcal{P}_n)]) \\
&= \frac{\rho(k)}{(1 - \rho(k))^2} \mathbb{E}[\Phi_k(\mathcal{P}_n)] + \frac{1}{(1 - \rho(k))^2} \mathrm{Var}\,(\Phi_k(\mathcal{P}_n)).
\end{aligned}$$

For fixed k, mean and variance for patricia tries are hence just linear combinations of mean and variance for tries, so this will also hold for $f_{E,k}, f_{V,k}$ and their Mellin transforms $f^*_{E,k}, f^*_{V,k}$.

Nevertheless, we show how to calculate them directly, starting with the mean. From Proposition 1 and (5) we have $\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda) = \mathbf{1}\{N_\lambda = k\}\varphi_p(\widetilde{\mathcal{T}}_\lambda)$, so the mean function is given by

$$\begin{aligned}
f_{E,k}(\lambda) &= \mathbb{E}\big[\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\big] \\
&= \mathbb{P}(N_\lambda = k)\mathbb{E}\big[\varphi_p(\widetilde{\mathcal{T}}_\lambda) \mid N_\lambda = k\big] \\
&= \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \left(1 - \sum_{a \in \mathcal{A}} p_a^k\right) \\
&= \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} (1 - \rho(k)) \qquad (22)
\end{aligned}$$

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

with $\rho(k)$ as defined in (4). By the definition of the Gamma function Γ, the Mellin transform is given by

$$f^*_{E,k}(s) = (1-\rho(k))\frac{\Gamma(k+s)}{k!}, \quad f^*_{E,k}(-1) = \frac{1-\rho(k)}{k(k-1)}. \tag{23}$$

Now to the variance. The function $f_{V,k}$ is given by

$$f_{V,k}(\lambda) := 2\operatorname{Cov}\big(\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda), \widehat{\Phi}_k(\widetilde{\mathcal{T}}_\lambda)\big) - \operatorname{Var}\big(\widehat{\varphi}(\widetilde{\mathcal{T}}_\lambda)\big).$$

Note that if $\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)$ is 1, that is, if the root has more than one child and there are k strings in the trie, all other nodes must have less than k strings. So then $\widehat{\Phi}_k(\widetilde{\mathcal{T}}_\lambda) = 1$. We then use monotone convergence to show

$$\begin{aligned} f_{V,k}(\lambda) &= \mathbb{E}\big[\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\big] - 2\mathbb{E}\big[\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\big]\mathbb{E}\big[\widehat{\Phi}_k(\widetilde{\mathcal{T}}_\lambda)\big] + \mathbb{E}\big[\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda)\big]^2 \\ &= f_{E,k}(\lambda) - 2\sum_{\alpha\in\mathcal{A}^*} f_{E,k}(\lambda)\mathbb{E}\big[\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda\alpha)\big] + f^2_{E,k}(\lambda) \\ &= f_{E,k}(\lambda) - 2\sum_{\alpha\in\mathcal{A}^*} f_{E,k}(\lambda) f_{E,k}(p_\alpha\lambda) + f^2_{E,k}(\lambda). \end{aligned} \tag{24}$$

The last equality is because $\widetilde{\mathcal{T}}_\lambda\alpha$ is the trie of the strings starting with α (at least if it exists), a fact which can be directly deduced from the definition in (1). Since the number of strings starting with α is Poi (λp_α)-distributed, $\widehat{\varphi}_k(\widetilde{\mathcal{T}}_\lambda\alpha)$ is distributed as $\widehat{\varphi}_k(\widetilde{\mathcal{T}}_{[\lambda p_\alpha]})$. Note that the last term is the term in the sum for $\alpha = \varepsilon$. Sums like in (24) show up often in the analysis of f_V, so Janson defines the shorthand $\sum_\alpha{}^* := \sum_{\alpha\in\mathcal{A}^*} + \sum_{\alpha\in\mathcal{A}^*\setminus\{\varepsilon\}}$, which summarizes our last two terms. We can then substitute $f_{E,k}$ using (22):

$$f_{V,k}(\lambda) = \frac{1-\rho(k)}{k!}\lambda^k \mathrm{e}^{-\lambda} - \sum_\alpha{}^*\left(\frac{1-\rho(k)}{k!}\right)^2 \lambda^{2k} p_\alpha^k \mathrm{e}^{-\lambda(1+p_\alpha)}.$$

The Mellin transform, which interchanges with the sum because of monotone convergence, can then be expressed using the Gamma function:

$$f^*_{V,k}(s) = \frac{1-\rho(k)}{k!}\Gamma(k+s) - \sum_\alpha{}^*\left(\frac{1-\rho(k)}{k!}\right)^2 p_\alpha^k \Gamma(s+2k)(1+p_\alpha)^{-s-2k}, \tag{25}$$

$$f^*_{V,k}(-1) = \frac{1-\rho(k)}{k(k-1)} - (2k-2)!\left(\frac{1-\rho(k)}{k!}\right)^2 \sum_\alpha{}^* \frac{p_\alpha^k}{(1+p_\alpha)^{2k-1}}. \tag{26}$$

Theorem 3. *Theorems 1 and 2 apply to the number $\Phi_k(\mathcal{P}_n)$ of fringe trees with $k \geq 2$ strings in a patricia trie of n strings. In particular, we have*

$$\frac{\Phi_k(\mathcal{P}_n) - \mathbb{E}\Phi_k(\mathcal{P}_n)}{\sqrt{n}} \overset{d}{\approx} \mathcal{N}(0, \sigma^2(\log n))$$

with convergence of all (absolute) moments. The variance $\sigma^2(\log n)$ is bounded and periodic. In the $d_p = 0$ case, we have convergence

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

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

a.s. and with all moments, and the term $\sigma^2(\log n)$ *is constant and given by*

$$\sigma^2(\log n) = \frac{1-\rho(k)}{Hk(k-1)} - \frac{(2k-2)!}{H}\left(\frac{1-\rho(k)}{k!}\right)^2 \sum_{\alpha}^{*} \frac{p_\alpha^k}{(1+p_\alpha)^{2k-1}}.$$

If further $|\mathcal{A}| = 2$*, this means that*

$$\mathbb{P}(|\mathcal{P}_n^*|_e = k \mid \mathcal{P}) \longrightarrow \frac{1-\rho(k)}{2Hk(k-1)} \tag{27}$$

a.s. and with all moments. In the $d_p > 0$ *case, small oscillations around the above values occur; see Theorem* 1.

Proof. The theorem is just a straightforward application of Theorem 1, with the values for $f_{E,k}^*(-1)$ and $f_{V,k}^*(-1)$ from (23) and (26) respectively. The statement for $|\mathcal{A}| = 2$ follows from the fact that for $|\mathcal{A}| = 2$ all internal nodes have the same outdegree of 2, so that $|\mathcal{P}_n| = 2n-1$ deterministically and therefore

$$\mathbb{P}(|\mathcal{P}_n^*|_e = k \mid \mathcal{P}_n) = \frac{1}{2n-1}\mathbb{E}[\Phi_k(\mathcal{P}_n) \mid \mathcal{P}_n] \quad \text{a.s.}$$

□

Remark 3. It is possible to get results like (27) for $|\mathcal{A}| > 2$ too. Theorem 4.7 of [10], which gives the asymptotic normality and asymptotics for

$$\mathcal{P}(|\mathcal{T}_n^*|_e = k \mid \mathcal{T}_n) = \frac{\Phi_k(\mathcal{T}_n)}{n + \Phi_i(\mathcal{T}_n)},$$

holds similarly for patricia tries, with just the functionals for size and fringe tree count replaced by their patricia trie counterparts. The quantity

$$J := \sum_{a\in\mathcal{A}} (1-p_a)\log\left(\frac{1}{1-p_a}\right),$$

describes the expected size of a patricia trie, which is (ignoring oscillations)

$$\frac{J+H}{H}n + \mathrm{o}(n)$$

(see [2]). The distribution of fringe trees is then asymptotically

$$\mathcal{P}(|\mathcal{P}_n^*|_e = k \mid \mathcal{P}_n) = \frac{1-\rho(k)}{(J+H)k(k-1)} + \text{oscillations} + \mathrm{o}(1).$$

See [6] for the detailed terms of the oscillations and the variance of $\mathcal{P}(|\mathcal{P}_n^*|_e \mathcal{P}_n) = k \mid \mathcal{P})$ using this analysis.

Compare this to the asymptotic distribution of fringe tries, which oscillates around

$$\frac{1}{(1+H)k(k-1)}$$

(see [10, Theorem 4.4]). In the trie, the distribution only depends on the entropy H, whereas in the patricia trie there is an additional dependence on the source in the form of ρ. Other

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

known asymptotic distribution of fringe trees include the random recursive tree, which has a limit of $1/k(k+1)$ [1]. While these probabilities decay with k^{-2} for large k, the probabilities for conditioned Galton–Watson decay more slowly, with order $k^{-3/2}$; see [1].

Remark 4. For a fixed $\mathcal{A}$-ary tree T, one can count the number Φ_T of fringe trees equal to T. Because the probability of $\widetilde{\mathcal{P}}_\lambda$ being T is the probability of the $\widetilde{\mathcal{P}}_\lambda$ having k strings and the patricia trie of the k strings being T, the mean function is given by $f_{E,T}(\lambda) = f_{E,k}(\lambda)\mathbb{P}(\mathcal{P}_k = T)$. The rest of the analysis is largely the same as for Φ_k and one can get an equivalent of Theorem 3 for Φ_T and the tree structure of $\mathcal{P}_n^*$.

We finish this section with the proof of Lemma 1.

Proof of Lemma 1. In a trie 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 being T' is $p_{T'} := k! \prod_{v \in T_e} p_v$. Having no node of outdegree 1 makes T a valid patricia trie. Write $\mathrm{pat}^{-1}(T)$ for the set of all tries T' such that $\mathrm{pat}\, T'$ has the tree structure T. The probability p_T is the sum of the $p_{T'}$ of the tries $T' \in \mathrm{pat}^{-1}(T)$ that are given by assigning common prefixes to the inner nodes. Fixing the common prefixes $I_v \in \mathcal{A}^*$ 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, which corresponds to a leaf $\widehat{w} \in T'$. This leaf $\widehat{w}$, when seen as a string, consists of the characters in w interleaved with all common prefixes I_v of the true ancestors $v \in T$ of w. To be precise, if $w = a_1 \cdots a_r$, then $\widehat{w} = I_\varepsilon a_1 I_{a_1} a_2 I_{a_1 a_2} a_3 \cdots I_{a_1 \ldots a_{r-1}} a_r$. Thus, writing $v \le w$ for 'v is ancestor of w', we have $p_{\widehat{w}} = p_w \prod_{v<w} p_{I_v}$. Multiplying p_w over all leaves, p_{I_v} gets multiplied as often as there are leaves in T^v. So the probability $p_{T'}$ is given by

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

Note that this term is the product of the proposed term for p_T and the proposed point masses of $q_{|T^v|_e}$. So if we show that the q_i for $i \ge 2$ really are distributions and sum up to 1, this already shows the term for p_T, q_i and the independence given the tree structure. Summing over strings of length $n \ge 0$, we have

$$\sum_{\alpha \in \mathcal{A}^n} q_i(\{\alpha\}) = (1-\rho(i)) \sum_{\alpha \in \mathcal{A}^n} p_\alpha^i = (1-\rho(i)) \left(\sum_{a \in \mathcal{A}} p_a^i \right)^n = (1-\rho(i))\rho(i)^n.$$

We have thus shown that the lengths are really $\mathrm{Geom}_0(1-\rho(i))$, which makes q_i indeed a well-defined distribution since $\rho(i) < 1$ for $i > 1$. □

5.2. Independence number

Given a graph $G = (V, E)$, an *independent set* $I \subseteq V$ is a set of nodes such that there are no edges between two nodes in I. The *independence number* $\alpha(G)$ of a tree G is the maximal cardinality of an independent set.

While it is an algorithmically challenging problem on general graphs, on a tree T, the independence number $\alpha(T)$ can be calculated recursively; see also [9] for references on the asymptotic properties of the independence number of other random tree models. The independence number only grows by at most 1 after adding a node. For the toll function φ_α of α

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

this means that $\varphi_\alpha(T) := \alpha(T) - \alpha(T\backslash\{\varepsilon\}) \in \{0, 1\}$. If $\varphi_\alpha(T) = 0$, the largest independent set of T is as large as the largest one of $T\backslash\{\varepsilon\}$. Therefore we can find a maximal independence set excluding the root. If $\varphi_\alpha(T) = 1$, the largest independent set of T is not an independent set of $T\backslash\{\varepsilon\}$, therefore every maximal independent set must include the root ε. We call such nodes $\beta \in T$ with $\varphi_\alpha(T^\beta) = 1$ *essential*.

An independent set containing the root cannot contain any of its children $b \in \mathcal{A}$. Hence it is upper-bounded by

$$1 + \sum_{b\in\mathcal{A}} \alpha(T^b\backslash\{b\}) = 1 + \alpha(T\backslash\{\varepsilon\}) - \sum_{b\in\mathcal{A}} \varphi_\alpha(T^b).$$

Furthermore, an independent set not containing the root has maximal cardinality $\alpha(T\backslash\{\varepsilon\})$. Thus we have

$$\varphi_\alpha(T) = \max\left\{0, 1 - \sum_{b\in\mathcal{A}} \varphi_\alpha(T^b)\right\}; \tag{28}$$

a node is exactly then essential if none of its children are. Another related quantity is the *matching number*, the maximal cardinality of a matching, which is just given by $|T| - \alpha(T)$ for trees.

The toll function φ_α is bounded and α is increasing. We can thus use Theorem 1 to get the following theorem.

Theorem 4. *Theorem 1 applies to the independence number $\alpha(\mathcal{P}_n)$ of a random patricia trie. Furthermore,* $\operatorname{Var}\alpha(\mathcal{P}_n) = \Omega(n)$ *holds, and*

$$\frac{\alpha(\mathcal{P}_n) - \mathbb{E}[\alpha(\mathcal{P}_n)]}{\sqrt{\operatorname{Var}(\alpha(\mathcal{P}_n))}} \xrightarrow{d} \mathcal{N}(0, 1) \tag{29}$$

with all moments. If $d_p = 0$, the expectation is asymptotically

$$\frac{\mathbb{E}[\alpha(\mathcal{P}_n)]}{n} = 1 + H^{-1} f^*_{E,\alpha}(-1) + \mathrm{o}(1),$$

and if $d_p > 0$, it is

$$\frac{\mathbb{E}[\alpha(\mathcal{P}_n)]}{n} = 1 + \frac{1}{H} \sum_{m=-\infty}^{\infty} f^*_{E,\alpha}\left(-1 - \frac{2\pi m\mathrm{i}}{d_p}\right) \mathrm{e}^{2\pi\mathrm{i}m \log n/d_p} + \mathrm{o}(1). \tag{30}$$

The values of $f^*_{E,\alpha}$ in the asymptotic expansions can be calculated to arbitrary precision. For example in the binary, symmetric case, the expected ratio $\mathbb{E}[\alpha(\mathcal{P}_n)]/(2n)$ of essential nodes oscillates around

$$\frac{1}{2} + \frac{1}{2H} f^*_{E,\alpha}(-1) \in (0.60225, 0.60316). \tag{31}$$

Proof. The limit (29) follows from Theorem 1(iii) if we show that $\operatorname{Var}\alpha(\mathcal{P}_n) = \Omega(n)$. This can be done by an extension of Lemma 3.14 of [10], as shown in [6]. A rough sketch is as follows. We condition on all nodes but those included in fringe trees of size 5 ('5-fringes') and furthermore on whether the roots of the 5-fringes are essential. This already fixes the essentiality of all nodes outside the 5-fringes. 5-fringes with non-essential root can have either

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

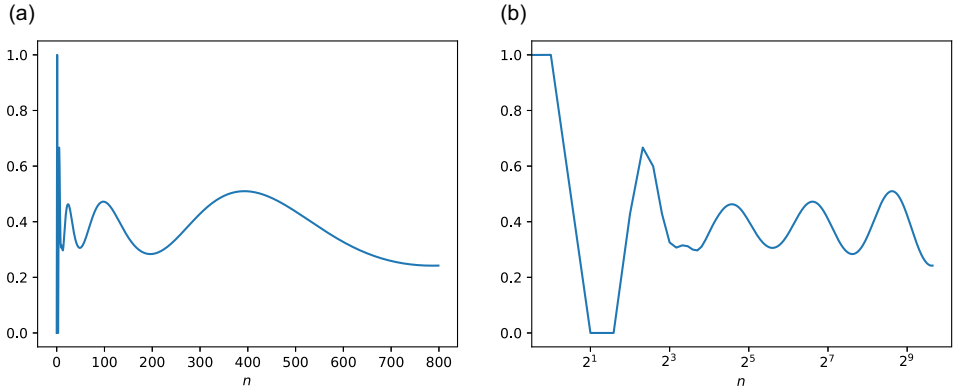


FIGURE 1. The first values of $\alpha_n := \mathbb{E}\varphi_\alpha(\mathcal{P}_n)$ in the symmetric, binary case, on (a) a normal and (b) a logarithmic x scale, showing the oscillations.

5 or 6 essential nodes, so the conditioned variance is not zero. By Section 5.1 there are $\Theta(n)$ 5-fringes, so we have

$$\operatorname{Var}\alpha(\mathcal{P}_n) \geq \mathbb{E}[\Phi_5(\mathcal{P}_n)]\mathbb{E}[\operatorname{Var}(\alpha(\mathcal{P}_5) \mid \varphi_\alpha(\mathcal{P}_5))] = \Theta(n).$$

The asymptotic expansions (30) and (31) are the definitions of $\psi_{E,\alpha}$ from (14) and (15) substituted into (17). Calculating the mean and variance is hindered by the recursive nature of (28), but we can estimate the asymptotic mean to arbitrary precision by counting only essential nodes with small fringe trees.

We calculate $\alpha_n := \mathbb{E}[\varphi_\alpha(\mathcal{P}_n)]$ for $0 \leq n \leq N$ with $N \in \mathbb{N}$. Then $\mathbb{E}[\alpha(\mathcal{P}_n)]$ can be bounded by

$$0 \leq \mathbb{E}[\alpha(\mathcal{P}_n)] - \left(n + \sum_{k=2}^{N} \mathbb{E}[\varphi_\alpha(\mathcal{P}_k)]\mathbb{E}[\Phi_k(\mathcal{P}_n)]\right) \leq \mathbb{E}[\Phi_{>N}(\mathcal{P}_n)].$$

We can now use the simultaneous convergence of Φ_i and Φ_k, $2 \leq k \leq N$ from Section 5.1. For $\Phi_{>N}$ we have $\Phi_{>N} = \Phi_i - \sum_{k=2}^{N} \Phi_k$ and the fact that the limits of $\mathbb{E}[\Phi_k]$, $k \geq 2$ sum up to the limit of $\mathbb{E}[\Phi_i]$, and obtain in the $d_p = 0$ case

$$0 \leq \lim_{n\to\infty} \frac{\mathbb{E}[\alpha(\mathcal{P}_n)]}{n} - \left(1 + \sum_{k=2}^{N} \frac{(1-\rho(k))\alpha_k}{k(k-1)H}\right) \leq \sum_{k=N+1}^{\infty} \frac{1-\rho(k)}{k(k-1)H} \leq \frac{1}{NH}. \tag{32}$$

In the periodic $d_p > 0$ case, the same can be done for each value of $f^*_{E,\alpha}$.

The values α_k can be calculated by conditioning on the first step; this gives a recursive equation. For the symmetric, binary case, this is given by

$$\alpha_n = \sum_{k=1}^{n-1} \binom{n}{k} \frac{(1-\alpha_k)(1-\alpha_{n-k})}{2^n - 2}$$

for $n \geq 2$ and $\alpha_1 = 1$, $\alpha_0 = 0$. See Figure 1 for a plot of the first values of α. Calculating these values up to $N = 800$ (higher values gave overflows in double-precision floats), we obtain from (32) the bounds

$$0.60225 < \frac{1}{2} + \frac{1}{2H} f^*_{E,\alpha}(-1) < 0.60316$$

for the asymptotic mean of the proportion of essential nodes. □

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press

Acknowledgements

I wish to thank my PhD advisor Ralph Neininger for insightful discussions and guidance on the presentation of this work.

Funding information

There are no funding bodies to thank relating to this creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALDOUS, D. J. (1991). Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Prob.* **1**, 228–266.
[2] BOURDON, J. (2001). Size and path length of Patricia tries: Dynamical sources context. *Random Structures Algorithms* **19**, 289–315.
[3] CLÉMENT, J., FLAJOLET, P. AND VALLÉE, B. (2001). Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica* **29**, 307–369.
[4] DRMOTA, M. (2009). *Random Trees: An Interplay between Combinatorics and Probability*. Springer, Vienna.
[5] FUCHS, M., HWANG, H.-K. AND ZACHAROVAS, V. (2014). An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoret. Comput. Sci.* **527**, 1–36.
[6] ISCHEBECK, J. (2022). Central limit theorems for fringe trees in patricia tries. Master's thesis, Universitätsbibliothek Johann Christian Senckenberg.
[7] JACQUET, P. AND SZPANKOWSKI, W. (2015). Algorithms and data structures. In *Analytic Pattern Matching: From DNA to Twitter*, pp. 135–153. Cambridge University Press, Cambridge.
[8] JANSON, S. (2016). Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Structures Algorithms* **48**, 57–101.
[9] JANSON, S. (2020). On the independence number of some random trees. *Electron. Commun. Prob.* **25**, 1–14.
[10] JANSON, S. (2022). Central limit theorems for additive functionals and fringe trees in tries. *Electron. J. Prob.* **27**, 1–63.
[11] KNUTH, D. (1973). *The Art of Computer Programming*, Vol. 3, *Sorting and Searching*. Addison-Wesley, Boston.
[12] MORRISON, D. R. (1968). PATRICIA: Practical algorithm to retrieve information coded in alphanumeric. *J. Assoc. Comput. Mach.* **15**, 514–534.
[13] SZPANKOWSKI, W. (1990). Patricia tries again revisited. *J. Assoc. Comput. Mach.* **37**, 691–711.
[14] WAGNER, S. (2015). Central limit theorems for additive tree parameters with small toll functions. *Combinatorics Prob. Comput.* **24**, 329–353.

https://doi.org/10.1017/jpr.2025.10034 Published online by Cambridge University Press