The Stack-Size of Tries
A Combinatorial Study

Markus E. Nebel

Johann Wolfgang Goethe-Universität
Fachbereich Informatik
Frankfurt am Main
Germany

Abstract

In this paper we introduce a class of extended binary trees that resembles all possible tree-structures of binary tries. Assuming a uniform distribution of those trees we prove that for α being the number of internal nodes the average stack-size is given by $\sqrt{\frac{3}{2}\pi\alpha}$. Since this result is quite similar to that for ordinary extended binary trees an attempt to find an explanation for that similarity using a quantitative level is made.

Keywords: Average case analysis, combinatorial structures, stack-size, tries.

1 Introduction

In the present paper we want to analyze the average stack-size of a class of generalized extended binary trees. Those trees result from ordinary extended binary trees by coloring their leaves in such a way that (under an appropriate interpretation of the different colors) all possible tree-structures of binary tries are resembled.
Given any binary tree[1] T, the stack-size $s(T)$ of T is given by

$$s(T) := \begin{cases} 1 & : \ T \text{ is either a leaf or empty} \\ \max(s(T.l), s(T.r) + 1) & : \ \text{else} \end{cases}$$

where $T.l$ (resp. $T.r$) denotes the left (resp. right) subtree of T. Like the height of a tree or the Horton-Strahler number, the stack-size is related to the recursion-depth needed to traverse the tree. When a *preorder*-traversal (see e.g. [Knu68]) is implemented by a recursive procedure the height of the traversed tree equals the number of stack-cells which are needed to store the return-addresses of the recursive procedure calls. By applying a technique known as end recursion removal (see [Sed88]) it is possible to optimize the space requirement. Then the stack-size of the tree describes the amount of stack space needed. The stack-size of a tree is also related to the evaluation of arithmetic expressions. If an expression is represented by its syntax tree (a binary tree where each internal node corresponds to an operator and each leaf represents an operand) then the stack-size of this tree equals the number of cells that are needed to store

[1] For the definition of s, T can either be an extended binary tree or a binary trie.

Figure 1: An example for a set of keys K, the resulting trie and the corresponding generalized extended binary tree.

intermediate results in order to evaluate the expression using a simple traversal strategy (see e.g. [Kem84] for details). From a pure combinatorial point of view the stack-size determines the height of the tree when only right edges contribute. Sometimes it is therefore also called *right-height* of the tree[2]. However, we prefer to use the notion stack-size in order to stress the relation to computer science.

A trie is a binary tree which is used to store the set of keys $K = \{k_1, \ldots, k_n\}$ in the following manner: Each key k_i, considered as a string of 0's and 1's due to its binary representation, defines a path in the binary tree (0 indicates a left turn and 1 a right turn); the trie defined by $k_1, \ldots, k_n$ is the smallest binary tree T for which the paths truncated at the leaves of T are all pairwise different. Thus each leaf of T *stores* exactly one of the keys k_i, $1 \leq i \leq n$. Note that it is not necessary that T is an extended binary tree. T might have internal nodes with only one successor. However, the stack-size $s(T)$ remains unchanged when we make T an extended binary tree.

Here we want to look at the combinatorics of binary tries by regarding the tree-structures that can be generated by the trie algorithm. More detailed, we want to study the stack-size of those tree-structures. Let a *generalized extended binary tree* be an extended binary tree with colored leaves. Leaves are colored black (represented as ■) or white (represented as □) such that each black leaf is the brother of an internal node. If we now assume a white leaf to store a key and a black leaf to represent a NIL-pointer, the class of generalized extended binary trees resembles all possible tree-structures of binary tries, see Figure 1 as an example. Note that by definition it is impossible that two leaves ■ belong to the same father or that a leaf □ is the brother of a leaf ■ since this would correspond to a subtrie storing no key at all or to a key that could be stored on a lower level of the trie, respectively. Both situations are avoided by the trie algorithm. Even if it is unnatural for the data structure trie we will assume in the sequel that all generalized extended binary trees with the same number of internal nodes are equally likely. The same is assumed for ordinary extended binary trees. As usual, $[x^n]f(x)$ is used to represent the coefficient at x^n in the Taylor expansion of $f(x)$ at $x = 0$. Just for the sake of simplicity a generalized extended binary tree with α internal nodes will be called α-trie. In order emphasize the combinatorial nature of our model the name $\mathcal{C}$-tries will be used to denote the class of all generalized extended binary trees. An ordinary

[2] Note that some authors consider the opposite case where the left son instead of the right one contributes. In that case we talk about the *left-height*. However, because of symmetry, such a modification of the definition would not affect our results.

Figure 2: The construction of an extended binary tree. $\{\square, \blacksquare\}$ represent a leaf that might be a NIL-pointer within a corresponding α-trie.

extended binary tree will be called extended binary tree.

2 The results

Basics

Let T be an α-trie and let T' denote the extended binary tree which is deduced from T by changing each leaf $\blacksquare$ into $\square$. Then the stack-size of T' is equal to $s(T)$. Let $\eta(T')$ denote the number of different α-tries T which are all transformed into the same extended binary tree T', and assume a constant behavior of $\eta(T')$ over the set of extended binary trees. In that case we could deduce the solution to our problem from the well studied stack-size of extended binary trees. Thus, before we start with complicated computations for the stack-size itself it makes sense to examine $\eta(T')$ in detail.
In order to analyze η we set up the ordinary generating function $B(x, w)$ for the extended binary trees where each internal node is marked by the variable x and each leaf that might either be $\square$ or $\blacksquare$ within a corresponding α-trie is marked by w. Since an extended binary tree is constructed symbolically as shown in Figure 2, $B(x, w)$ fulfills the functional equation $B(x, w) = x + 2xwB(x, w) + xB(x, w)^2$. Thus we find

$$B(x, w) = \frac{1 - 2xw - \sqrt{1 - 4xw + 4x^2w^2 - 4x^2}}{2x}.$$

To determine the average number of leaves that might either be white or black we have to take the partial derivative with respect to w and set w equal to 1 afterwards. We find

$$\frac{\partial}{\partial w} B(x, w)|_{w=1} = -\frac{\sqrt{1-4x} + 2x - 1}{\sqrt{1-4x}} = \frac{1}{\sqrt{1-4x}} - \frac{2x}{\sqrt{1-4x}} - 1.$$

We could use the binomial theorem to get an explicit representation of the coefficient $[x^\alpha]\frac{\partial}{\partial w}B(x, w)|_{w=1}$ but it is also quite simple to get an asymptotic equivalent by applying the transfer lemmata of [FlO90] (all we need is the fact that $[z^n](1-z)^\gamma \sim n^{-\gamma-1}/\Gamma(-\gamma)$). We find that the coefficient in question behaves like

$$\frac{1}{2}\frac{\alpha^{\frac{1}{2}-1}4^\alpha}{\Gamma(\frac{1}{2})}, \; \alpha \to \infty.$$

To get the average number of appropriate leaves we have to divide this quantity by the number of extended binary trees with α internal nodes. We get $\frac{1}{2}\alpha$ as a result. To obtain further knowledge on η we now determine the related variance. Thus we compute $\frac{\partial^2}{\partial w^2}B(x,w)|_{w=1}$, use the transfer lemmata to estimate its coefficients and divide by the number of extended binary trees of size α to get the second factorial moment $\frac{1}{4}\alpha^2$. This, together with the expected value, can be used to compute the variance. Again, we find $\frac{1}{2}\alpha$. Since $\eta(T')$ is equal to 2^ν where ν is the number of leaves of T' that might be a $\square$ or a $\blacksquare$ within a corresponding α-trie, η cannot have a constant behavior for all trees of size α. As a consequence, the stack-size of $\mathcal{C}$-tries cannot be deduced from the extended binary trees in that obvious way. Thus a detailed analysis is required.

A detailed analysis

Let us start this section with the observation that we have to consider the number of internal nodes of an α-trie when analyzing the average stack-size. Even if a specification of the number of white leaves would have a closer connection to the data structure trie and to the notion of size usually used there, we are faced with the situation that there are infinitely many $\mathcal{C}$-tries with a fixed number of white leaves and a limited stack-size.
Later in this section we will need the number of $\mathcal{C}$-tries with α internal nodes. Therefore this number is quantified first.

Lemma 1 *Let $\delta_{n,m}$ denote Kronecker's delta. The number $|T_\alpha|$ of $\mathcal{C}$-tries with α internal nodes is given by*

$$|T_\alpha| = \sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1}\binom{n}{\alpha+1-n}2^{2n-\alpha-1}(-3)^{\alpha+1-n} - \delta_{\alpha,0}.$$

Proof: Let x mark an internal node. The construction process of an α-trie is shown in Figure 2. But here a leaf $\blacksquare$ has to be interpreted as a second possibility for building an α-trie. Thus, for $T(x)$ the ordinary generating function of the $\mathcal{C}$-tries with at least one internal node, Figure 2 translates into $T(x) = x + 4xT(x) + xT^2(x)$. Therefore $T(x) = \frac{1-4x-\sqrt{1-8x+12x^2}}{2x}$ holds. There is only one $\mathcal{C}$-trie with no internal nodes and we have to add 1 in order to take this tree into account. In that way we find

$$T(x) = \frac{1-2x-\sqrt{1-8x+12x^2}}{2x}. \qquad (1)$$

We conclude by expanding $T(x)$. $\square$

We continue our investigations by deriving the ordinary generating function $A_k(x)$ of $\mathcal{C}$-tries that have a stack-size less than $k+1$. If we quantify the number

Figure 3:
All possible decompositions of an α-trie with β white leaves and a stack-size of at most i. The number inside the triangle corresponds to the number of white leaves it has to possess, the number below a triangle determines the stack-size of the subtree represented by it.

of white leaves in a (sub)trie, it is easy to distinguish between the different types of leaves □ and ■. In this setting, an α-trie with β white leaves and a stack-size of at most k can be decomposed into the cases of Figure 3. Let $L_{i,j}(x)$ denote the ordinary generating function of $\mathcal{C}$-tries with j white leaves and a stack-size of at most i. Then those cases translate into the following set of equations:

$$\begin{aligned} L_{i,1}(x) &= 1,\, i \geq 1, \\ L_{1,j}(x) &= \delta_{j,1},\, j \geq 1, \\ L_{i,j}(x) &= xL_{i,j}(x) + xL_{i-1,j}(x) + x \sum_{\substack{\beta_1+\beta_2=j \\ \beta_1\cdot\beta_2\neq 0}} L_{i,\beta_1}(x)L_{i-1,\beta_2}(x) \\ &= \left(xL_{i-1,j}(x) + x \sum_{\substack{\beta_1+\beta_2=j \\ \beta_1\cdot\beta_2\neq 0}} L_{i,\beta_1}(x)L_{i-1,\beta_2}(x) \right)(1-x)^{-1}. \end{aligned}$$

To solve this system of equations we introduce the bivariate generating function $\mathsf{A}_i(x,w) := \sum_{j\geq 1} L_{i,j}(x)w^j$. We find $\mathsf{A}_1(x,w) = w$ and for $i \geq 2$

$$\mathsf{A}_i(x,w) = \frac{w + \frac{x}{1-x}\mathsf{A}_{i-1}(x,w) - \frac{xw}{1-x}}{1 - \frac{x}{1-x}\mathsf{A}_{i-1}(x,w)}.$$

Now $\mathsf{A}_k(x,1) = A_k(x)$ holds and thus

$$\begin{aligned} A_1(x) &= 1 \\ A_k(x) &= -1 + \frac{3x-2}{-1+x+xA_{k-1}(x)}. \end{aligned}$$

Therefore, $A_k(x)$ is the k-th approximant of a continued fraction of the pattern $A_k(x) = -1 + \frac{c_1}{c_2+xA_{k-1}(x)}$, $A_1(x) = c_3$ (in our case $c_1 = 3x-2$, $c_2 = -1+x$ and $c_3 = 1$ holds). This suggests to express the generating function as a quotient of polynomials $X_k(x)$ and $Y_k(x)$ with

$$\frac{X_k(x)}{Y_k(x)} = -1 + \frac{c_1}{c_2 + x\frac{X_{k-1}(x)}{Y_{k-1}(x)}} = \frac{(c_1-c_2)Y_{k-1}(x) - xX_{k-1}(x)}{c_2Y_{k-1}(x) + xX_{k-1}(x)},$$

$X_1(x) = c_3$ and $Y_1(x) = 1$. We translate this representation into the following equations of matrices

$$\begin{aligned} \begin{pmatrix} X_1(x) \\ Y_1(x) \end{pmatrix} &= \begin{pmatrix} c_3 \\ 1 \end{pmatrix}, \\ \begin{pmatrix} X_k(x) \\ Y_k(x) \end{pmatrix} &= \begin{pmatrix} -x & c_1 - c_2 \\ x & c_2 \end{pmatrix} \begin{pmatrix} X_{k-1}(x) \\ Y_{k-1}(x) \end{pmatrix}, \; k \geq 2. \end{aligned} \tag{2}$$

Now (2) is solved by introducing the generating function $F(q) = \sum_{n\geq 1} \binom{X_n(x)}{Y_n(x)} q^n$ which has the closed form representation

$$F(q) = \begin{pmatrix} \frac{q(-c_3+qc_2c_3-qc_1+qc_2)}{-1+qc_2-qx+q^2xc_1} \\ \frac{-q(qxc_3+qx+1)}{-1+qc_2-qx+q^2xc_1} \end{pmatrix}.$$

To determine $A_k(x) = [q^k]F(q)$ we compute the partial fraction decomposition of both entries of the above vector with respect to q. Let $\rho_1(x)$ and $\rho_2(x)$ denote the two roots of $-1 + qc_2 - qx + q^2xc_1 = 0$ with respect to q. Then we find

$$\frac{q(-c_3 + qc_2c_3 - qc_1 + qc_2)}{-1 + qc_2 - qx + q^2xc_1} = \frac{A}{\rho_1(x) - q} + \frac{B}{\rho_2(x) - q} + \frac{c_2c_3 - c_1 + c_2}{xc_1}$$

and

$$\frac{-q(qxc_3 + qx + 1)}{-1 + qc_2 - qx + q^2xc_1} = \frac{\bar{A}}{\rho_1(x) - q} + \frac{\bar{B}}{\rho_2(x) - q} - \frac{c_3 + 1}{c1}$$

for

$$\begin{aligned} A &= \frac{c_1(x + c_3x - c_2) + c_2(c_3 + 1)(c_2 - x)}{2c_1^2x^2} \\ &\quad + \frac{-c_2(c_3 + 1)(c_2 - x)^2 + 2xc_1^2 + c_1(c_2^2 - c_2x(4 + 3c_3) + (c_3 + 1)x^2)}{2c_1^2x^2\sigma}, \end{aligned}$$

$$\begin{aligned} B &= \frac{c_1(x + c_3x - c_2) + c_2(c_3 + 1)(c_2 - x)}{2c_1^2x^2} \\ &\quad + \frac{c_2(c_3 + 1)(c_2 - x)^2 - 2xc_1^2 - c_1(c_2^2 - c_2x(4 + 3c_3) + (c_3 + 1)x^2)}{2c_1^2x^2\sigma}, \end{aligned}$$

$$\bar{A} = \frac{c_1 - (c_3 + 1)(c_2 - x)}{2xc_1^2} + \frac{(c_3 + 1)(c_2 - x)^2 + c_1(-c_2 + (3 + 2c_3)x)}{2xc_1^2\sigma}$$

and

$$\bar{B} = \frac{c_1 - (c_3 + 1)(c_2 - x)}{2xc_1^2} + \frac{c_1(c_2 - (3 + 2c_3)x) - (c_3 + 1)(c_2 - x)^2}{2xc_1^2\sigma}.$$

Here, σ is a short cut for $\sqrt{c_2^2 - 2c_2 x + x^2 + 4xc_1}$. Applying the binomial theorem gives us $[q^n]\frac{A}{\rho_1(x)-q} = \frac{A}{\rho_1^{(n+1)}(x)}$. Analogously we derive the other coefficients. This gives us the following representation of the generating function in question:

$$A_k(x) = \frac{X_k(x)}{Y_k(x)} = \frac{A\rho_2^{k+1}(x) + B\rho_1^{k+1}(x)}{\bar{A}\rho_2^{k+1}(x) + \bar{B}\rho_1^{k+1}(x)}.$$

Now, for $\kappa := 2x - 3x^2$, $\rho_1(x) = \frac{1+\sqrt{1-4\kappa}}{-2\kappa}$ and $\rho_2(x) = \frac{1-\sqrt{1-4\kappa}}{-2\kappa}$ holds. Introducing the substitutions $u := (1-\epsilon)/(1+\epsilon)$ with $\epsilon := \sqrt{1-4\kappa}$ and $S_k(u) := \frac{1-u^{k-1}}{1-u^k}(1+u)$, the application of numerous algebraic manipulations finally yields

$$A_k(x) = \frac{1 - xS_k(u)}{1 - 2xS_k(u)}. \tag{3}$$

We now have to determine $[x^\alpha]A_k(x)$. Therefore we expand (3) to

$$(1 - xS_k(u))\sum_{i\geq 0}(2x)^i S_k^i(u) = \sum_{i\geq 0} 2^i x^i S_k^i(u) - \sum_{i\geq 0} 2^i x^{i+1} S_k^{i+1}(u).$$

Now for $\kappa = x$, $S_k(u)$ is the well-known generating function of those extended binary trees that have a stack-size less than k. Thus, it is possible to use an old result due to R. Kemp which gives a representation for the i-th power of $S_k(u)$:

Lemma 2 ([Kem80]) *Let*

$$\mathsf{S}_k(x) := \frac{1-u^k}{1-u^{k+1}}(1+u)$$

with $u = (1-\sqrt{1-4x})/(1+\sqrt{1-4x})$ *be the generating function of those extended binary trees that have a stack-size of at most* k*. Then for* $i \geq 1$

$$\begin{aligned} \mathsf{S}_{k-1}^i(x) &= \sum_{n\geq 0} x^n \sum_{\lambda\geq 0}\sum_{h\geq 0}(-1)^\lambda \binom{i}{\lambda}\binom{i-1+h}{i-1} \\ &\quad \times \left[\binom{2n+i-1}{n-(k-1)\lambda - kh} - \binom{2n+i-1}{n-(k-1)\lambda-kh-1}\right] \end{aligned}$$

holds. □

Define $\varphi(i,k,n) := \sum\limits_{\lambda\geq 0}\sum\limits_{l\geq 0}(-1)^\lambda\binom{i}{\lambda}\binom{i-1+l}{i-1}\left[\binom{2n+i-1}{n-(k-1)\lambda-kl} - \binom{2n+i-1}{n-(k-1)\lambda-kl-1}\right]$.
Then

$$\begin{aligned} \sum_{i\geq 0} 2^i x^i S_k^i(u) &= \sum_{i\geq 0} 2^i x^i \left[\sum_{n\geq 0}\kappa^n \varphi(i,k,n) + \delta_{i,0}\right] \\ &= \sum_{i\geq 0}\sum_{n\geq 0}\sum_{m\geq 0}\binom{n}{m} 2^{n-m+i}(-3)^m x^{i+n+m}\varphi(i,k,n) + \delta_{i,0}, \end{aligned}$$

and

$$\begin{aligned}\sum_{i\geq 0} 2^i x^{i+1} S_k^{i+1}(u) &= \sum_{i\geq 0} 2^i x^{i+1} \sum_{n\geq 0} \kappa^n \varphi(i+1,k,n) \\ &= \sum_{i\geq 0}\sum_{n\geq 0}\sum_{m\geq 0} \binom{n}{m} 2^{n-m+i}(-3)^m x^{i+n+m+1}\varphi(i+1,k,n).\end{aligned}$$

Now we can pick the coefficient at x^α to quantify the number of $\mathcal{C}$-tries with α internal nodes and a stack-size of at most k. We find that $[x^\alpha]A_k(x)$ is given by 1 for $\alpha = 0$ and by

$$\sum_{i\geq 0}\sum_{n\geq 0} \binom{n}{\alpha-i-n} 2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\varphi(i,k,n)$$

$$-\sum_{i\geq 0}\sum_{n\geq 0} \binom{n}{\alpha-i-n-1} 2^{2n+2i-\alpha+1}(-3)^{\alpha-i-n-1}\varphi(i+1,k,n)$$

for $\alpha \geq 1$. Applying some fundamental simplifications lead to the following lemma:

Lemma 3 *The number $S_{k,\alpha}$ of $\mathcal{C}$-tries with α internal nodes and a stack-size $\leq k$ is 1 for $\alpha = 0$ and*

$$\frac{1}{2}\sum_{i\geq 0}\sum_{n\geq 0} \binom{n}{\alpha-i-n} 2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\varphi(i,k,n)$$

for $\alpha \geq 1$. □

To quantify the average stack-size of $\mathcal{C}$-tries we have to determine

$$|T_\alpha|^{-1} \sum_{1\leq k\leq \alpha+1} k(S_{k,\alpha} - S_{k-1,\alpha}) = (\alpha+1) - |T_\alpha|^{-1} \sum_{1\leq k\leq\alpha} S_{k,\alpha}.$$

For this purpose we introduce $\phi(n,i,\lambda,a) := \sum_{1\leq k\leq\alpha}\sum_{l\geq 0}\binom{i-1+l}{i-1}\binom{2n+i-1}{n-(k-1)\lambda-kl-a}$. Then $\sum_{1\leq k\leq\alpha} S_{k,\alpha}$ reads

$$\frac{1}{2}\sum_{\substack{i\geq 0\\ n\geq 0}}\binom{n}{\alpha-i-n} 2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\sum_{\lambda\geq 0}(-1)^\lambda\binom{i}{\lambda}\,[\phi(n,i,\lambda,0)-\phi(n,i,\lambda,1)].$$

A simple rearrangement of the terms in the sum $\phi(n,i,\lambda,a)$ shows, that

$$\phi(n,i,\lambda,a) = (1+\delta_{\lambda,0}(\alpha-1))\binom{2n+i-1}{n-a}$$

$$+\sum_{v\geq 1}\binom{2n+i-1}{n-v-a}\sum_{d|(v+\lambda)}\binom{i-1+d-\lambda}{i-1}.$$

Thus $(\alpha+1)-|T_\alpha|^{-1}\sum_{1\leq k\leq\alpha}S_{k,\alpha}$ is given by

$$(\alpha+1)-\frac{1}{2}|T_\alpha|^{-1}(\Theta_\alpha^{(1)}+\Theta_\alpha^{(2)}+\Theta_\alpha^{(3)}+\Theta_\alpha^{(4)})$$

with

$$\begin{aligned}
\Theta_\alpha^{(1)} &= \sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\\
&\quad\times\alpha\left[\binom{2n+i-1}{n}-\binom{2n+i-1}{n-1}\right],\\
\Theta_\alpha^{(2)} &= \sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\\
&\quad\times\sum_{v\geq 1}\left[\binom{2n+i-1}{n-v}-\binom{2n+i-1}{n-v-1}\right]\sum_{d|v}\binom{i-1+d}{i-1},\\
\Theta_\alpha^{(3)} &= \sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\\
&\quad\times\sum_{\lambda\geq 1}(-1)^\lambda\binom{i}{\lambda}\left[\binom{2n+i-1}{n}-\binom{2n+i-1}{n-1}\right]
\end{aligned}$$

and

$$\begin{aligned}
\Theta_\alpha^{(4)} &= \sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\sum_{\lambda\geq 1}(-1)^\lambda\binom{i}{\lambda}\\
&\quad\times\sum_{v\geq 1}\left[\binom{2n+i-1}{n-v}-\binom{2n+i-1}{n-v-1}\right]\sum_{d|(v+\lambda)}\binom{i-1+d-\lambda}{i-1}.
\end{aligned}$$

Using Lemma 1 we have $\Theta_\alpha^{(1)}=2\alpha|T_\alpha|$ and $\Theta_\alpha^{(3)}=-2|T_\alpha|$. Hence the average stack-size can be written as

$$2+\frac{1}{2}|T_\alpha|^{-1}\sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\sum_{\lambda\geq 0}(-1)^{\lambda+1}\binom{i}{\lambda}$$
$$\times\sum_{v\geq 1}\left[\binom{2n+i-1}{n-v}-\binom{2n+i-1}{n-v-1}\right]\sum_{d|(v+\lambda)}\binom{i-1+d-\lambda}{i-1}.$$

Decreasing the index v by λ the innermost sum can be split into a difference of two sums. We obtain the representation $2+h_\alpha^{(1)}-h_\alpha^{(2)}$ with

$$h_\alpha^{(1)}=\frac{1}{2}|T_\alpha|^{-1}\sum_{i\geq 0}\sum_{n\geq 0}\binom{n}{\alpha-i-n}2^{2n+2i-\alpha}(-3)^{\alpha-i-n}\sum_{\lambda\geq 0}(-1)^{\lambda+1}\binom{i}{\lambda}$$

$$\times \sum_{v \geq 1} \left[\binom{2n+i-1}{n-v+\lambda} - \binom{2n+i-1}{n-v+\lambda-1} \right] \sum_{d|v} \binom{i-1+d-\lambda}{i-1},$$

$$h_\alpha^{(2)} = \frac{1}{2} |T_\alpha|^{-1} \sum_{i \geq 0} \sum_{n \geq 0} \binom{n}{\alpha - i - n} 2^{2n+2i-\alpha} (-3)^{\alpha-i-n} \sum_{\lambda \geq 0} (-1)^{\lambda+1} \binom{i}{\lambda}$$

$$\times \sum_{1 \leq v \leq \lambda} \left[\binom{2n+i-1}{n-v+\lambda} - \binom{2n+i-1}{n-v+\lambda-1} \right] \sum_{d|v} \binom{i-1+d-\lambda}{i-1}.$$

Now regard the last sum. The term

$$\sum_{d|v} \binom{i-1+d-\lambda}{i-1}$$

is zero for $d < \lambda$. Since $d \leq v \leq \lambda$, the sum $h_\alpha^{(2)}$ collapses and we get $h_\alpha^{(2)} = 1$. This gives the following theorem:

Theorem 1 *The average stack-size of $\mathcal{C}$-tries with α internal nodes is explicitly given by*

$$1 + \frac{1}{2} |T_\alpha|^{-1} \sum_{i \geq 0} \sum_{n \geq 0} \binom{n}{\alpha - i - n} 2^{2n+2i-\alpha} (-3)^{\alpha-i-n} \sum_{\lambda \geq 0} (-1)^{\lambda+1} \binom{i}{\lambda}$$

$$\times \sum_{v \geq 1} \left[\binom{2n+i-1}{n-v+\lambda} - \binom{2n+i-1}{n-v+\lambda-1} \right] \sum_{d|v} \binom{i-1+d-\lambda}{i-1}.$$

Here, a representation for $|T_\alpha|$ is stated in Lemma 1. □

Asymptotic behavior

To see how the average stack-size behaves it is necessary to derive some asymptotic results since the representation in Theorem 1 does not give us sufficient information. To do this, we could try to estimate the exact solution of the last section. For that reason we refer to [Kem80] where it was necessary to estimate the sum

$$T_{n,r}^{-1} \sum_{0 \leq \lambda \leq r} (-1)^\lambda \binom{r}{\lambda} \sum_{m \geq 1} \left[\binom{2n-r-3}{n-r-m+\lambda-1} - \binom{2n-r-3}{n-r-m+\lambda-2} \right] \tag{4}$$

$$\times \sum_{d|m} \binom{d+r-\lambda-1}{r-1},$$

for $T_{n,r}$ the number of planted plane trees with a root of degree r and n nodes. Under the assumption of r being constant, a demanding computation led to the asymptotic

$$\sqrt{\pi n} - \frac{1}{2}r - \frac{1}{2} + \mathcal{O}(\ln(n)/n^{0.5-\varepsilon}),$$

for some $\varepsilon > 0$. Unfortunately, since we have to sum over different instances of (4)[3] this assumption cannot be guaranteed in our case. Further it is not obvious how to get rid of it. Thus, neither the results of [Kem80] nor the idea of the computation can be used. Therefore, we return to (3) and consider the sum $T(x) + \sum_{k\geq 1}(T(x) - A_k(x))$ where $T(x)$ is given in (1). This sum resembles in terms of generating functions the sum of k times the number of $\mathcal{C}$-tries T with $s(T) = k$ taken over all possible k. Using our generating functions we find the representation

$$T(x) + \frac{\sqrt{1-4\kappa}}{x}\sum_{k\geq 1} g^k(x) \sum_{d|k}\left(\frac{s(x)}{h(x)}\right)^d,$$

where

$$g(x) = \frac{1-\sqrt{1-4\kappa}}{1+\sqrt{1-4\kappa}},$$
$$s(x) = \kappa\sqrt{1-4\kappa},$$

and

$$h(x) = -2(-1+2x)(-1+3x)(-1+6x) + \sqrt{1-4\kappa}(2-7\kappa).$$

Returning to our substitution u, we get the somehow simplified form

$$\underbrace{u + \sqrt{1-u+u^2}}_{=:f_1(u)} + \underbrace{\frac{3(1-u)}{1+u-\sqrt{1-u+u^2}}}_{=:f_2(u)}$$
$$\times \sum_{k\geq 1} u^k \sum_{d|k}\left(1 + \underbrace{\frac{2(u-1)(-1+u+\sqrt{1-u+u^2}}{u}}_{=:f_3(u)}\right)^d$$

which, by means of the binomial theorem and the well known relation of Stirling numbers of the first kind $\mathcal{S}_k^{(v)}$ and binomial coefficients (e.g. [Kem84], B8),

$$\binom{x}{r} = \frac{1}{r!}\sum_{0\leq v\leq r} \mathcal{S}_r^{(v)} x^v,$$

can be transformed into

$$f_1(u) + f_2(u)\sum_{r\geq 0}\sum_{v\geq 0}\frac{1}{r!}\mathcal{S}_r^{(v)} f_3^r(u) \sum_{k\geq 1} u^k \sigma_v(k).$$

[3]Not really (4), but something quite similar to it.

Here, $\sigma_v(k)$ is the sum of the v-th powers of the positive divisors of k. Now, setting $u := \exp(-t)$ and applying the well known identity

$$\exp(-tj) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \Gamma(s) j^{-s} t^{-s} ds$$

for some c in the fundamental strip of the Mellin transform of $\exp(-tj)$ and $\Gamma(s)$ the complete gamma function, it is possible to express the sum over the divisor function $\sigma_v(k)$ by means of the Riemann Zeta function (see [Apo76] for details). We obtain

$$f_1(e^{-t}) + f_2(e^{-t})\sum_{v\geq 0}\sum_{r\geq 0}\frac{1}{r!}\mathcal{S}_r^{(v)} f_3^r(e^{-t})\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Gamma(s)t^{-s}\zeta(s)\zeta(s-v)ds.$$

It is standard to expand the generating function about its dominant singularity (which is $x = \frac{1}{6}$ in our case) and to use $\mathcal{O}$-transfer (see [FlO90]) to derive an asymptotic for the coefficients. Thus, we consider $t = 0$ and use residue calculus to evaluate the integral. For $v = 0$ the sum of the residues is given by

$$t^{-1}(\gamma - \log(t)) + \frac{1}{4} - \frac{1}{144}t - \frac{1}{86400}t^3 + \mathcal{O}(t^4).$$

The sum $f_2(e^{-t})\sum_{r\geq 0}\frac{1}{r!}\mathcal{S}_r^{(0)} f_3(e^{-t})^r$ possesses the expansion

$$3t + \frac{7}{8}t^3 - \frac{29}{640}t^5 + \mathcal{O}(t^6).$$

For $v = 1$, we find

$$\frac{1}{6}\pi^2 t^{-2} - \frac{1}{2}t^{-1} + \frac{1}{24} + \mathcal{O}(t)$$

for the integral and

$$-6t^2 - t^4 - \frac{3}{160}t^6 + \mathcal{O}(t^8)$$

for the leading factor. For $v > 1$, we only have contributions of the order $\mathcal{O}(t^4)$. As we will see those can be neglected. Further we have

$$f_1(e^{-t}) = \frac{3}{2} - \frac{3}{2}t + \frac{9}{8}t^2 - \frac{7}{16}t^3 + \mathcal{O}(t^4).$$

Thus, we find the following expansion about $t = 0$:

$$-3\log(t) + \frac{9}{4}t + \frac{9}{32}t^3 + \mathcal{O}(t^4).$$

Now we have to extract the coefficients. Since $t = -\log(u) = -\log(\frac{1-\epsilon}{1+\epsilon})$, we conclude that $t \sim 2(1-4\kappa)^{\frac{1}{2}}$. Thus $[x^\alpha](-3\log(t)) \sim -3[x^\alpha]\log(\frac{4}{3}\sqrt{6}(1 -$

α	exact	asymptotic	exact/asymptotic
4	3.473684211	3.216607528	1.079921682
8	5.095040934	4.827460250	1.055428874
12	6.402595151	6.144884822	1.041939001
16	7.521408593	7.276965056	1.033591413
20	8.514998606	8.283129563	1.027992927
24	9.417908737	9.197223110	1.023994811
28	10.25124129	10.04038524	1.021000793
32	11.02901804	10.82679550	1.018677968
36	11.76101957	11.56648925	1.016818441
40	12.45457331	12.26686849	1.015301772
⋮	⋮	⋮	⋮
100	20.355424941	20.22303764	1.006546361
200	29.305882012	29.20730125	1.003380000
300	36.186729411	36.10442411	1.002280000

Table 1: Some exact and asymptotic values of the average stack-size.

$6x)^{\frac{1}{2}}) = \frac{3}{2}\frac{6^\alpha}{\alpha}$. By the application of transfer-lemmata we can determine the contribution of the other terms. We find in total, that for $\mathcal{C}$-tries with α internal nodes our sum behaves as

$$\frac{3}{2}\frac{6^\alpha}{\alpha} + \frac{9}{2}\left(\frac{2}{3}\right)^{\frac{1}{2}} \frac{\alpha^{-\frac{1}{2}-1}6^\alpha}{\Gamma(-\frac{1}{2})} + \frac{9}{2}\left(\frac{2}{3}\right)^{\frac{3}{2}} \frac{\alpha^{-\frac{3}{2}-1}6^\alpha}{\Gamma(-\frac{3}{2})}.$$

This quantity has to be divided by the asymptotical number of $\mathcal{C}$-tries of size α. Applying an $\mathcal{O}$-transfer to (1) we find that

$$|T_\alpha| \sim \sqrt{\frac{3}{2}}\frac{\alpha^{-\frac{3}{2}}6^\alpha}{\sqrt{\pi}}.$$

This gives us the following theorem:

Theorem 2 *Under the uniform model the average stack-size of a $\mathcal{C}$-trie with α internal nodes is asymptotically*

$$\sqrt{\frac{3}{2}\pi\alpha} + \frac{3}{2}\frac{(1-\alpha)}{\alpha}.$$

□

In Table 1 we find some exact and asymptotic values of the above average together with their quotient.

If we compare the leading term of this average value with the average stack-size of ordinary extended binary trees (which is given by $\sqrt{\pi\alpha}$ ([BKR72], [Kem84], Theorem 5.3)) the only difference is the factor $\sqrt{\frac{3}{2}}$. Even if it seems to be obvious that the coloring of leaves (i.e. the change of ordinary extended binary trees into generalized ones which model the possible structures of tries) should not severely affect the average stack-size, the similar behavior of both classes of trees asks for a detailed investigation. A first attempt for that can be found in the following section.

3 Essay of explanation

In this section we try to find a relation between extended binary trees and our generalized variant which explains their similar behavior concerning the average stack-size. In detail we try to find an explanation for the factor $\sqrt{\frac{3}{2}}$. If we could find such a relation it might make the analysis of the last section superfluous and could enable us to derive our result from [BKR72] or [Kem84], Theorem 5.3. Starting point of our consideration is the structural equivalence of both classes of trees which only differ in the following fact: Within the class of $\mathcal{C}$-tries (generalized extended binary trees) some of the leaves $\square$ might be $\blacksquare$ as well. This implies that for each extended binary tree there might be more than one α-trie of the same structure. All those α-tries differ in the color of their leaves only. At the beginning of Section 2 we have computed the average number of leaves of an extended binary tree that might be a $\blacksquare$ within an α-trie. The result told us that this number is of large variation and thus a direct derivation of the average stack-size of $\mathcal{C}$-tries from that of extended binary trees seemed impossible. But what happens if there is no variation of this number for all trees of the same stack-size. In that case it might be possible to derive the average stack-size of $\mathcal{C}$-tries from that of extended binary trees or at least to explain the factor $\sqrt{\frac{3}{2}}$.

Let $\mathcal{B}$ denote the set of all extended binary trees and let $\mathfrak{l}(T)$ represent the number of leaves $\square$ of $T \in \mathcal{B}$ that might be a $\blacksquare$ within an α-trie of T's structure. To compute the average behavior of $\mathfrak{l}(T)$ we derive the generating function $B_k(x,w) := \sum_{T\in\mathcal{B},\, s(T)\le k} x^{|T|}w^{\mathfrak{l}(T)}$, where $|T|$ is the number of internal nodes of T. For $B_k(x,w)$ we have to distinguish the cases shown in Figure 4. Note that $\{\square,\blacksquare\}$ is used to picture the case of a leaf $\square$ that might be a $\blacksquare$ as well. The cases of Figure 4 translate into the equations

$$\begin{aligned} B_1(x,w) &= 1, \\ B_k(x,w) &= 1 + x + xw(B_k(x,w)-1) + xw(B_{k-1}(x,w)-1) \\ &\quad + x(B_k(x,w)-1)(B_{k-1}(x,w)-1),\ k\ge 2. \end{aligned}$$

Figure 4: The construction of an extended binary tree with a stack-size $\leq k$. The label on top of the root of each tree specifies the stack-size of it. Each triangle represents a subtree with at least one internal node, the number below a triangle determines its stack-size.

Thus, for $k \geq 2$, we find

$$\begin{aligned} B_k(x,w) &= \frac{1+2x+xw(B_{k-1}(x,w)-2)-xB_{k-1}(x,w)}{1-xw-x(B_{k-1}(x,w)-1)} \\ &= 1-w+\frac{-w-x+w^2x}{-1-x+wx+xB_{k-1}(x,w)}. \end{aligned}$$

Let $\mathcal{S}_k(x) = \sum_{T\in\mathcal{B},\, s(T)\leq k} x^{|T|}$ be the generating function of Lemma 2. Then $B_k(x,1) = \mathcal{S}_k(x)$ holds. We introduce $\mathcal{B}_k(x) := \frac{\partial}{\partial w}B_k(x,w)|_{w=1}$ and use the identity $\mathcal{S}_k(x) = 1/(1-x\mathcal{S}_{k-1}(x))$ (see [BKR72]) which gives us

$$\begin{aligned} \mathcal{B}_1(x) &= 0 \\ \mathcal{B}_k(x) &= -x(-(1+2x)\mathcal{S}_{k-1}(x)+xS_{k-1}^2(x)+1-\mathcal{B}_{k-1}(x))\mathcal{S}_k^2(x). \end{aligned}$$

It is quite easy to solve this linear recurrence. We find

$$\mathcal{B}_k(x) = -x^{k-1}\prod_{i=1}^{k-1}\mathcal{S}_{i+1}^2(x)\sum_{j=1}^{k-1}\frac{x^{-j+1}\mathcal{S}_{j+1}^2(x)(-(1+2x)\mathcal{S}_j(x)+x\mathcal{S}_j^2(x)+1)}{\prod_{l=1}^{j}\mathcal{S}_{l+1}^2(x)}.$$

To simplify this solution we first have a look at the product

$$\prod_{i=1}^{k-1}\mathcal{S}_{i+1}^2(x) = \prod_{i=1}^{k-1}\left[2\frac{(1+\sqrt{1-4x})^{i+1}-(1-\sqrt{1-4x})^{i+1}}{(1+\sqrt{1-4x})^{i+2}-(1-\sqrt{1-4x})^{i+2}}\right]^2$$

which is telescopic and collapses to

$$4^{k-1}\left[\frac{(1+\sqrt{1-4x})^2-(1-\sqrt{1-4x})^2}{(1+\sqrt{1-4x})^{k+1}-(1-\sqrt{1-4x})^{k+1}}\right]^2.$$

This gives the following representation of $\mathcal{B}_k(x)$:

$$\begin{aligned} \mathcal{B}_k(x) &= -(4x)^{k-1}[(1+\sqrt{1-4x})^{k+1}-(1-\sqrt{1-4x})^{k+1}]^{-2} \\ &\quad\times\sum_{j=1}^{k-1}x^{-j+1}4^{-j+1}[(1+\sqrt{1-4x})^{j+1}-(1-\sqrt{1-4x})^{j+1}]^2 \\ &\quad\times(-(1+2x)\mathcal{S}_j(x)+x\mathcal{S}_j^2(x)+1). \end{aligned}$$

Inserting the closed form representation for $\mathcal{S}_k(x)$ and performing some algebraic simplifications yields

$$\begin{aligned}\mathcal{B}_k(x) &= 4^{k+1}x^k\left(\left(1+\sqrt{1-4x}\right)^{k+1}-\left(1-\sqrt{1-4x}\right)^{k+1}\right)^{-2} \\ &\quad\times\Bigg(1-k-4x+2kx+2x^{-k+2}(1-4x)^{-\frac{1}{2}} \\ &\quad\times\left[\left(\frac{1+\sqrt{1-4x}}{2}\right)^{2k}-\left(\frac{1-\sqrt{1-4x}}{2}\right)^{2k}\right]\Bigg).\end{aligned} \tag{5}$$

From [Knu68], p. 93, we know that

$$\underbrace{(1-4x)^{-\frac{1}{2}}\left[\left(\frac{1+\sqrt{1-4x}}{2}\right)^{2k}-\left(\frac{1-\sqrt{1-4x}}{2}\right)^{2k}\right]}_{=:p_{2k}(x)} = \sum_{i=0}^{2k-1}\binom{2k-i-1}{i}(-x)^i$$

holds. Thus only the denominator of (5) can have singularities. From [BKR72] we know that the equation

$$\left(1+\sqrt{1-4x}\right)^{k+1}-\left(1-\sqrt{1-4x}\right)^{k+1}=0$$

possesses the solutions $x_j=[4\cos^2(j\pi/(k+1))]^{-1}$, $1\le j<\frac{k+1}{2}$. The solution of smallest modulus is given by $x_1=[4\cos^2(\pi/(k+1))]^{-1}$. Furthermore, for $x=[4\cos^2(\Theta)]^{-1}$, $p_k(x)$ is given by

$$\sin(k\Theta)/(\sin(\Theta)(2\cos(\Theta))^{k-1}).$$

Thus

$$\begin{aligned}\mathcal{B}_k([4\cos^2(\Theta)]^{-1}) &= [4\cos^2(\Theta)]^{-k}(1-k-\cos^{-2}(\Theta)+k[2\cos^2(\Theta)]^{-1} \\ &\quad +2[4\cos^2(\Theta)]^{k-2}[\sin(2k\Theta)/(\sin(\Theta)(2\cos(\Theta))^{2k-1})]) \\ &\quad \times\sin^2(\Theta)(2\cos(\Theta))^{2k}\sin^{-2}((k+1)\Theta)(1-\cos^{-2}(\Theta))^{-1}\end{aligned}$$

holds. In order to apply Darboux's Theorem we regard

$$\left(1-\frac{\cos^2(\pi/(k+1))}{\cos^2(\Theta)}\right)^{-2}\underbrace{\left(1-\frac{\cos^2(\pi/(k+1))}{\cos^2(\Theta)}\right)^{2}\mathcal{B}_k([4\cos^2(\Theta)]^{-1})}_{=:g_k([4\cos^2(\Theta)]^{-1})}.$$

By Darboux's Theorem (e.g. see [Kem84], Theorem 4.12) we know that the coefficient $[x^\alpha]\mathcal{B}_k(z)$ is asymptotically given by

$$\alpha[4\cos^2(\pi/(k+1))]^\alpha g_k([4\cos^2(\pi/(k+1))]^{-1}).$$

Figure 5: The convergence of $\mathcal{A}_k(\alpha)$ to $\mathcal{A}_k(\alpha) - \mathcal{A}_{k-1}(\alpha)$.

To determine $g_k([4\cos^2(\pi/(k+1))]^{-1})$ we consider the limit

$$\lim_{\Theta \to \pi/(k+1)} g_k([4\cos^2(\Theta)]^{-1})$$

which evaluates to

$$\frac{2}{(k+1)^2} \tan^2(\pi/(k+1))(2 + (k-1)\cos(2\pi/(k+1))).$$

Thus we have the following lemma:

Lemma 4 *Let T be an extended binary tree and let $\mathfrak{l}(T)$ denote the number of leaves of T that might be black within an α-trie of T's structure. Then $\sum_{T\in\mathcal{B},\, s(T)\le k,\, |T|=\alpha} \mathfrak{l}(T)$ is asymptotically given by*

$$\mathcal{A}_k(\alpha) := \alpha[4\cos^2(\pi/(k+1))]^\alpha \frac{2}{(k+1)^2} \tan^2(\pi/(k+1))(2+(k-1)\cos(2\pi/(k+1))),$$

$\alpha \to \infty$, k fixed. □

To find the average behavior, this number must be divided by the number $\mathcal{T}_k(\alpha)$ of trees $T \in \mathcal{B}$ with $s(T) \le k$ and $|T| = \alpha$. It is well known that this number behaves as [BKR72]

$$\frac{4^{\alpha+1}}{k+1} \tan^2(\pi/(k+1)) \cos^{2\alpha+2}(\pi/(k+1)) \tag{6}$$

for $\alpha \to \infty$ and fixed k. Thus, we find as a result

Theorem 3 *For large α and fixed k, the average number of leaves of an extended binary tree T with $s(T) \le k$ and $|T| = \alpha$ that could be black within an α-trie of T's structure is asymptotically given by*

$$\alpha \frac{2 + (k-1)\cos(2\pi/(k+1))}{2(k+1)\cos^2(\pi/(k+1))}.$$

□

Remarks:

- Since with respect to k the essential singularity of $\mathcal{B}_k(x)$ is strictly monotonically decreasing, the asymptotic of Theorem 3 also holds for $s(T) = k$.
- An asymptotic for the case $s(T) = k$ of higher precision is given by $(\mathcal{A}_k(\alpha) - \mathcal{A}_{k-1}(\alpha))/(\mathcal{T}_k(\alpha) - \mathcal{T}_{k-1}(\alpha))$, but as you can see in Table 2 the improvement is rather small.

In Figure 5 we have illustrated the quotient $(\mathcal{A}_k(\alpha)-\mathcal{A}_{k-1}(\alpha))/\mathcal{A}_k(\alpha)$ in order to show the rate of convergence. It is remarkable that this convergence seems to be rather slow in contrast to the similarity of both asymptotics for the average value.

We now determine the second factorial moment of our parameter. For that purpose we take the second partial derivative of $B_k(x,w)$ with respect to w and set $w=1$ afterwards. A lengthy computation proves that $\mathcal{B}_k^{(2)}(x) := \frac{\partial^2}{\partial w^2}B_k(x,w)|_{w=1}$ is given by

$$\mathcal{B}_k^{(2)}(x) = \mathcal{D}(x)^{-1}(\mathcal{N}_1(x)+k\mathcal{N}_2(x)+k^2\mathcal{N}_3(x))$$

for

$\mathcal{N}_1(x) =$

$$\frac{4^3x^3((1-\varepsilon)^{3k+3}+(1+\varepsilon)^{3k+3})+(1-\varepsilon)^{k+2}(-a-b)-(1+\varepsilon)^{k+2}(a-b)}{8\varepsilon^3}$$

with $\varepsilon=\sqrt{1-4x}$, $a=4^kx^k(64(5x-1)(1+(x-4)x))$ and $b=4^kx^k(32\varepsilon(-3+2x(11+(x-19)x)))$,

$\mathcal{N}_2(x) = \dfrac{4^{k+1}x^k(-\varepsilon((1-\varepsilon)^k+(1+\varepsilon)^k)c+((1-\varepsilon)^k-(1+\varepsilon)^k)d}{1-4x}$

with $c=(-3+4(x-1)x(2x-5))$, $d=(-3+4x(7+x(26x-23)))$,

$\mathcal{N}_3(x) =$

$$\frac{(1-2x)^2 4^{k+1}x^k((1-4x)((1+\varepsilon)^k-(1-\varepsilon)^k)+\varepsilon((1-\varepsilon)^k+(1+\varepsilon)^k))}{-1+4x}$$

and

$\mathcal{D}(x) = ((1+\varepsilon)^{k+1}-(1-\varepsilon)^{k+1})^3.$

We set $x=[4\cos^2(\Theta)]^{-1}$ within ε in order remove the square roots and afterwards $\Theta=\pi/(k+1)$ for those parts that do not get singular. We find that $\mathcal{B}_k^{(2)}(x)=\sin^{-3}((k+1)\Theta)\mathcal{R}_k(x)$ with

$$\mathcal{R}_k(x) = \Bigg(\frac{1}{4\sin^3(2\pi/(k+1))} + \frac{\cos^3(\pi/(k+1))(2(1-5x)(1+x^2-4x)+(1-4x-4x^2+8x^3)\sin^2(\pi/(k+1)))}{\sin^3(\pi/(k+1))} - k\frac{\cos^5(\pi/(k+1))(8x-64x^2+96x^3)}{\sin(\pi/(k+1))} + k^2\frac{\cos^2(2\pi/(k+1))}{2\sin(2\pi/(k+1))}\Bigg).$$

To apply the theorem of Darboux we need a representation of the pattern $(1-z/z_\lambda)^{-\omega_\lambda}g_\lambda(z)$ for z_λ being the singular point and $g_\lambda(z)$ analytic near z_λ. For

k	α	as1.	as2.	ex.
10	10	4.725674713	4.294738942	8.058823529
	30	14.17702414	14.62539900	15.43112546
	50	23.62837357	23.88451696	24.64011046
	70	33.07972300	33.26497837	34.00313618
	100	47.25674713	47.38070781	48.10891257
20	20	9.816090348	9.638733210	18.02702703
	40	19.63218070	18.98365281	23.59014547
	60	29.44827105	22.10571473	32.11052278
	80	39.26436140	40.93564480	41.34600893
	100	49.08045174	50.01890270	50.85703475
30	40	19.81986873	19.65245893	29.53666026
	60	29.72980310	29.41844700	36.08046869
	80	39.63973748	39.09113905	44.35521268
	100	49.54967184	48.53224651	53.31545713
40	40	19.89361458	19.80877001	38.01298701
	60	29.84042188	29.70005855	41.71019744
	80	39.78722917	39.57846751	48.57581765

Table 2: The rough [as1.] and the improved [as2.] asymptotical values for the case $s(T) = k$ together with the corresponding exact [ex.] account.

that purpose we determine

$$\lim_{\Theta \to \pi/(k+1)} \left(1 - \frac{\cos^2(\pi/(k+1))}{\cos^2(\Theta)}\right)^3 \frac{1}{\sin^3((k+1)\Theta)}$$

which evaluates to

$$8 \frac{\sin^3(\pi/(k+1))}{\cos^3(\pi/(k+1))(k+1)^3}.$$

Thus $[x^\alpha]\mathcal{B}_k^{(2)}(x)$ possesses the asymptotical representation

$$4\alpha^2 [4\cos^2(\pi/(k+1))]^\alpha \frac{\sin^3(\pi/(k+1))}{\cos^3(\pi/(k+1))(k+1)^3} \mathcal{R}_k([4\cos^2(\pi/(k+1))]^{-1}).$$

The application of numerous simplifications such as well known identities for trigonometric functions gives us

$$[z^\alpha]\mathcal{B}_k^{(2)}(x) \sim \frac{4^\alpha \alpha^2 \cos^{2\alpha-4}(\pi/(k+1)) \sin^2(\pi/(k+1))}{(k+1)^3} \times \big((-2 + \cos(2\pi/(k+1)))^2 - k(1 - 4\cos(2\pi/(k+1))$$

$$
\begin{aligned}
& \quad + \cos(4\pi/(k+1))) + k^2\cos^2(2\pi/(k+1))\big) \\
= & \quad \frac{4^\alpha \alpha^2 \cos^{2\alpha-4}(\pi/(k+1))\sin^2(\pi/(k+1))(2+(k-1)\cos(2\pi/(k+1)))^2}{(k+1)^3}.
\end{aligned}
$$

To determine the second factorial moment this quantity has to be divided by (6) which yields:

Theorem 4 *For large α and fixed k, the second factorial moment of the number of leaves of an extended binary tree T with $s(T) \leq k$ and $|T| = \alpha$ that could be black within an α-trie of T's structure is asymptotically given by*

$$
\alpha^2 \frac{(2+(k-1)\cos(2\pi/(k+1)))^2}{4(k+1)^2\cos^4(\pi/(k+1))}.
$$

Note that the same arguments as those for the first moment imply the validity of Theorem 4 also for the case $s(T) = k$. Thus, asymptotically the second factorial moment of the number of leaves in an extended binary tree of stack-size k that might be black within a corresponding α-trie behaves like the square of the first moment. This implies the variance being equal to the first moment and thus not being zero. In conclusion, our attempt to explain the similar behavior of ordinary extended binary trees and our generalized variant with respect to the stack-size failed.

4 Concluding remarks

In this paper we have proved that within the uniform model the average stack-size of an α-trie is asymptotically given by $\sqrt{\frac{3}{2}\pi\alpha}$. This is quite similar to the average stack-size of extended binary trees which is given by $\sqrt{\pi\alpha}$. This similarity is obviously implied by our model which just considers all possible tree-structures of a trie as equally likely and thus leads to a generalization of extended binary trees; the attempt to explain the similarity in detail failed. However, further investigations are sensible. Since for the data structure trie keys are only stored in external nodes, a result which also considers the number of leaves would be of interest. Furthermore, the assumption of a more realistic probability model such as the Bernoulli- or Poisson model (see [Mah92]) could be the starting point of further observations which then would not be a combinatorial study but an investigation of tries as a data structure.

References

[Apo76] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, 1976

[BKR72] N. G. de Bruijn, D. E. Knuth, S.O. Rice, *The Average Height of Planted Plane Trees*, Graph Theory and Computing (R. C. Read, ed.), Academic Press, 1972

[FlO90] P. Flajolet and A. Odlyzko, *Singularity Analysis of Generating Functions*, SIAM J. Disc. Math. **3**, No. 2, 216-240, 1990

[Kem80] R. Kemp, *The Average Height of r-Tuply Rooted Planted Plane Trees*, Computing **25**, 209-232, Springer, 1980

[Kem84] R. Kemp, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner Series in Computer Science, 1984

[Knu68] D. E. Knuth, *The Art of Computer Programming*, Volume 1: Fundamental Algorithms, Addison-Wesley, 1968

[Mah92] H. M. Mahmoud, *Evolution of Random Search Trees*, Wiley-Interscience Series, 1992

[Sed88] R. Sedgewick, *Algorithms*, Addison-Wesley, 1988