1986

Patricia Tries Again Revisited

Wojciech Szpankowski

Purdue University, spa@cs.purdue.edu

Report Number:

86-625

PATRICIA TRIES AGAIN REVISITED

Wojciech Szpankowski

CSD-TR-625
September 1986

PATRICIA TRIES AGAIN REVISITED

*Wojciech Szpankowski**
Department of Computer Science
Purdue University
West Lafayette, IN 47907

Abstract

This paper studies the average complexity of Patricia tries from the successful and unsuccessful search point of view. It is assumed that the Patricia trie is built over a V-element alphabet, and keys are strings of elements from the alphabet. The occurrence of the ith element from the alphabet in a key is given by a probability p_i, $i = 1, 2, \ldots, V$. We also assume that n keys are stored in the Patricia trie. These assumptions determine the so called Bernoulli model. Let S_n and U_n denote the successful search and unsuccessful search in the Patricia, respectively. We prove that the m-th moment of the successful search, $E(S_n)^m$, satisfies $\lim_{n\to\infty} E(S_n)^m / ln^m n = 1/h_1^m$, where $h_1 = -\sum_{i=1}^{V} p_i \ln p_i$. In particular, we show that the variance of S_n is $\mathrm{var}\, S_n = c \ln n + O(1)$ (c is a constant dependent on p_i, $i = 1, 2, \ldots, V$) for an asymmetric Patricia, and $\mathrm{var}\, S_n = O(1)$ for a symmetric Patricia (e.g. if $V = 2$ $\mathrm{var}\, S_n = 1.00$). The unsuccessful search U_n, is studied only for binary symmetric Patricia tries. We prove that $\lim_{n\to\infty} E(U_n)^m / lg^m n = 1$. In particular, the variance of U_n, is given by $\mathrm{var}\, U_n = 0.8790$.

1. INTRODUCTION

Digital searching is a well-known technique for storing and retrieving information using lexicographical (digital) structure of words. Let A be an alphabet containing V elements, $A = \{\sigma_1, \ldots, \sigma_V\}$ and we define a set S which consists of finite numbers, say n, of (possible infinite) strings (keys) from A. A *trie or radix search trie* is a V-ary digital search tree in which edges are labelled by elements from A and leaves (external nodes) contain the keys [2], [7], [10], [12], [14], [18]. The access path from the root to the leaf is a minimal prefix of the information contained in the leaf. The radix trie has an annoying flaw: there is "one-way branching" which leads to the creation of extra nodes in the tree. D.R. Morrison discovered a way to avoid this problem in a structure which he named the *Patricia trie*. In such a tree all nodes have branching degree greater equal than two. This is achieved by collapsing one-way branches on internal nodes. For more details see [7], [10], [14], [18]. The Patricia tree finds many applications e.g. in

*and Technical University of Gdainsk, Poland

lexicographical order [16], dynamic hashing algorithms [4], and most recently in conflict resolution algorithms [9], [21]. (For more examples see [10], [14].)

Two quantities are of interest for tries in general, and in particular for the Patricia trie, namely successful search and unsuccessful search. *A successful search* occurs whenever a new key added to the Patricia is already in the trie. If a new key is not in the Patricia, then an *unsuccessful search* occurs. The average complexity analysis of the Patricia is very scarce, (see [7], [13], [14]) and in fact restricted to binary symmetric Patricia, that is, $V = 2$ is assumed with equal probability of occurrence of the elements over a binary alphabet. Moreover, only average values were studied. These simplifications are dropped in this paper. For the successful search, we assume that a sequence of elements from the alphabet A is an independent sequence of Bernoulli trials (*Bernoulli model*), and the probability of occurrence of an element $\sigma_i \in A$ in a key is equal to p_i, $i = 1, 2, \ldots, V$. Under these assumptions we study all moments of the successful search, S_n. It is shown that $\lim_{n\to\infty} E(S_n)^m / ln^m n = 1/h_1^m$, where $h_1 = -\sum_{i=1}^{V} p_i \ln p_i$ and m is an integer. In particular, we prove that the variance of S_n is $(h_2 - h_1^2)h_1^{-3}\ ln\ n + O(1)$ for an asymmetric Patricia, where $h_2 = \sum_{i=1}^{V} p_i\ ln^2 p_i$. Note that this implies that the variance in the symmetric Patricia is equal to O(1) (e.g. for $V = 2$ the variance var $S_n = 1.00$, for $V = 3$ we find var $S_n = 0.6309$ and so on). These results extend the works of Knuth [14], Flajolet and Sedgewick [7], and Kirschenhofer and Prodinger [13].

The results for the unsuccessful search are even more scarce, and to the author's knowledge only the mean value of U_n was obtained by Knuth [14]. The problem is also much more intricate, therefore, symmetric binary Patricia tries are assumed. However, asymptotic analysis of all moments of the unsuccessful search is discussed. It is proved that that the m-th moment of U_n satisfies $\lim_{n\to\infty} E(U_n)^m / lg\ n = 1$ for any integer m, where $lgn = \log_2 n$. In particular, the variance of U_n is equal to 0.8790.

The paper is organized as follows. In the next section we present some notation and preliminary results. In particular, we derive a recurrence equation for the generating function of the V-ary asymmetric Patricia trie. Then, in Section 3, we study the successful search. At the beginning we present some recurrence equations, show how to solve them using the results obtained by the author in [19], [22], and we also give final results of this section. In the last part of the section we show how to prove the main results. In a similar way, we organize Section 4 where unsuccessful search for binary symmetric Patricia trie is discussed.

2. NOTATIONS AND PRELIMINARY RESULTS

Let us consider a family T_n of Patricia tries with n keys (records) built over and alphabet $A = \{\sigma_1, \ldots, \sigma_V\}$. A key is a string of (possible infinite) elements from A, such that the i-th element $\sigma_i \in A$ occurs independently of other elements, and with probability p_i, $i = 1, 2, \ldots, V$, $\sum_{i=1}^{V} p_i = 1$. A trie $t \in T_n$ built over A is called V-ary asymmetric Patricia trie since the alphabet contains V elements which are distributed according to the probabilities p_i, $i = 1, 2, \ldots, V$. The keys are stored in external nodes, while internal nodes determine branching strategy. The degree of each internal node is greater equal than two ,that is, one-way branches are collapsed on internal modes by including in the nodes the number of bits to skip over before making the next decision (for details see [7], [10], [19], [18]). Two parameters of tries are of particular interest: *successful search* and *unsuccessful search*. The successful search, called also the depth of a leaf, is the number of internal nodes in the trie on the path from the root to a given key (external node) if the key *is* already in the trie. If the new key *is not* in the trie, then an unsuccessful search occurs. The unsuccessful search is not simply related to the successful search, since unsuccessful searches are more likely to occur at external nodes near the root.

We study properties of successful and unsuccessful searches in a random family of Patricia tries T_n. Let S_n and U_n (random variables) denote the successful search and the unsuccessful search in T_n. The m-th factorial moments of S_n and U_n are defined as follows

$$s_n^{\underline{m}} \stackrel{def}{=} E\{S_n(S_n-1)(S_n-2)\cdots(S_n-m+1)\}, \tag{2.1}$$

$$u_n^{\underline{m}} = E\{U_n(U_n-1)(U_n-2)\cdots(U_n-m+1)\}, \tag{2.2}$$

where the expectations in (2.1) and (2.2) is taken over all tries in T_n and over all external nodes in a given trie $t \in T_n$. It is shown (see next sections) that these moments are related to m-th derivatives of the so called generating functions of T_n. Let $H_n(z)$ denote this generating function with the coefficient at z^k being the expected number of external nodes at level k in our family of trees.

There is no explicit formula for $H_n(z)$ but a rather sophisticated recurrence. To find it, let us denote by $\mathbf{j}=(j_1, j_2, \ldots, j_V)$ a vector such that $j_1+j_2+\cdots+j_V=n$. Also let $\binom{n}{\mathbf{j}} \stackrel{def}{=} \begin{bmatrix} n \\ j_1, \ldots, j_V \end{bmatrix} = \frac{n!}{j!\, j_2! \cdots j_V!}$ be a multinomial coefficient, and let $\sum_{\{j_\Sigma=n\}} f(j_1, \ldots, j_V)$ denote a sum of $f(j_1, \ldots, j_V)$ over all $\mathbf{j}$ such that $j_1+j_2+\cdots+j_V=n$ for a given function $f(\cdot)$. Then the following recurrence on $H_n(z)$ may be established.

Lemma 1. For any natural n the generating function $H_n(z)$ of the random family of Patricia tries, T_n, satisfies the recurrence

$$H_0(z)=0, \qquad H_1(z)=1$$
$$H_n(z) = z\sum_{\{j_\Sigma=n\}} \binom{n}{\mathbf{j}} p_1^{j_1}\cdots p_V^{j_V}\,[H_{j_1}(z)+\cdots+H_{j_V}(z)\,] - (z-1)[p_1^n+p_2^n+\cdots+p_V^n\,] \tag{2.3}$$

Proof: Consider V subtrees of the root, each with $j_1, j_2 \cdots j_V$ keys, $j_1+j_2+\cdots+j_V=n$. Then, for a given trie $t \in T_n$

$$H_n(z) = [H_{j_1}(z)+\cdots+H_{j_V}(z)][z+\delta_{j_1,n}(1-z)+\delta_{j_2,n}(1-z)+\cdots+\delta_{j_V,n}(1-z)].$$

where $\delta_{j,k}$ is the Kronecker delta. In the second squared bracket the first z shows the fact that the subtrees are one level below the root, and the other terms are responsible for avoiding one-way branches (e.g. if $j_1 = n$, then the leftmost branch would be one-way branch, but $z + \delta_{j_1,n}(1-z) = 1$, hence the subtree starts at the root). Taking now the expectation of the last recurrence over all tries in T_n, and noting that in our Bernoulli model the probability of $j_1, \ldots, j_V$ keys in the subtrees is equal to $\binom{n}{\mathbf{j}} p_1^{j_1} \cdots p_V^{j_V}$, we finally obtain (2.3).

□

3. SUCCESSFUL SEARCH

In this section we analyze the successful search, that is, we derive an asymptotic approximations for all moments of S_n. We start with some initial results follow by our final results. In the further part of this section, we show how to derive these results.

3.1. Initial and final results

Let L_n denote an external path in a trie $t \in T_n$, that is, it is the sum of all paths from the root to external nodes. We generalize the definition of L_n as follows. Let $S_n(i)$ be a path from root to the i-th external node. For a given integer m we define

$$L_n^{\underline{m}} = \sum_{i=1}^{n} S_n(i)[S_n(i)-1][S_n(i)-2] \cdots [S_n(i)-m+1] \qquad (3.1)$$

and let $l_n^{\underline{m}} = EL_n^{\underline{m}}$. The quantity $l_n^{\underline{m}}$ is not exactly the m-th factorial moment of L_n, but it is closely related to it. We call $l_n^{\underline{m}}$ the m-th semi-factorial moment of the external path length.

Denote now by $H_n^{(m)}(1)$ the m-th derivative of $H_n(z)$ at $z = 1$. Then the following is easily to establish (see [14], [22])

Property 1. For integers n and m the below relationships hold

$$H_n(1) = n \qquad l_n^{\underline{m}} = H_n^{(m)}(1) \qquad (3.2)$$

$$s_n^{\underline{m}} = l_n^{\underline{m}}/n \qquad (3.3)$$

□

Using Lemma 1 and (3.2) we derive a recurrence equation for l_n^m, hence by (3.3) also on s_n^m. We shall work at the beginning with l_n^m. For simplicity of computations, assume now that $V = 2$ and $p_1 = p$, $p_2 = 1 - p_1 = q$. Then for $m = 1$ we find immediately that

$$l_n^1 = n(1 - p^n - q^n) + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} [l_k^1 + l_{n-k}^1]. \tag{3.4}$$

Computing the second derivative of $H_n(z)$ one shows that

$$l_n^2 = 2[\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (l_n^1 + l_{n-k}^1) - l_n (p^n + q^n)] + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} [l_n^2 + l_{n-k}^2].$$

But, the first term of the above is by (3.4) equal to $l_n^1 - n$, hence

$$l_n^2 = 2(1 - p^n - q^n)[l_n^1 - n] + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} [l_k^2 + l_{n-k}^2]. \tag{3.5}$$

Note that (3.4) and (3.5) is a system of recurrences, i.e., to find l_n^2 we need l_n^1. Generalizing the above, we can prove that

Lemma 2. For any integers m and n, the m-th semi-factorial moment of L_n satisfies the following recurrence

$$l_0^m = l_1^m = 0$$

$$l_n^m = m!(1 - \sum_{i=1}^{V} p_i^n) \sum_{k=1}^{m} (-1)^{m-k} \frac{l_n^{k-1}}{(k-1)!} + \sum_{\{j_x = n\}} \binom{n}{\mathrm{j}} p_1^{j_1} \cdots p_V^{j_V} [l_{j_1}^m + \cdots + l_{j_V}^m] \tag{3.6}$$

where in (3.6) we have defined $l_n^0 = n$.

Proof: The proof uses induction arguments applied to (2.3), and is left to the reader.

□

As noted before, (3.6) is a system of recurrences. To compute l_n^m we need l_n^1,

$l_n^2, \ldots, l_n^{m-1}$ from the previous recurrences. Note also that (3.6) has a common pattern and the recurrences differ only by the first term in (3.6) which we call the *additive term* and denote by a_n. This type of recurrence has been solved by Szpankowski in [19], [22] (see also [14]). Since we use it extensively in the further part of the paper we quote below, without proofs, some of these results.

Solution of a recurrence equation

Let $x_0, x_1, \ldots, x_n$ be a sequence of numbers satisfying the following linear recurrence

$$\begin{aligned} &\text{given} \quad x_0 = x_1 = 0 \\ &\text{solve} \quad x_n = a_n + \sum_{\{j_\Sigma = n\}} \begin{bmatrix} n \\ \mathbf{j} \end{bmatrix} p_1^{j_1} \cdots p_V^{j_V} [x_{j_1} + \cdots + x_{j_V}] \end{aligned} \tag{3.7}$$

where a_n is any sequence of numbers. We call a_n an additive term of the recurrence (3.7). To solve (3.7) we introduce the so called *binomial inverse relations*. Let us, for a given sequence a_n, define a new sequence $\hat{a}_n$ as

$$\hat{a}_n = \sum_{k=0}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} a_k \qquad a_n = \sum_{k=0}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \hat{a}_k \tag{3.8}$$

(The second equation justify the name binomial inverse relations). For more details see Riordan [23]. Using the above we proved in [19] that

Theorem 1. The recurrence (3.7) possesses the following solution

$$x_n = \sum_{k=2}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \frac{\hat{a}_k + k a_1 - a_0}{1 - \sum_{i=1}^{V} p_i^k}. \tag{3.9}$$

□

From the numerical point of view the solution (3.9) is not better than the recurrence (3.7), however, (3.9) might be used to derive asymptotic approximation for x_n. In most computer science applications a_n is a linear combination of the following terms $\begin{bmatrix} n \\ r \end{bmatrix} c^n$, where r is an integer, and c is a constant. From [23] [14], we know that

$$a_n = \begin{bmatrix} n \\ r \end{bmatrix} c^n \rightarrow \hat{a}_n = \begin{bmatrix} n \\ r \end{bmatrix} (-c)^r (1-c)^{n-r}. \tag{3.10}$$

Since the recurrence (3.7) is linear, from the asymptotic point of view we need an approximation for large n of the following

$$T_{n,r}(c) \stackrel{def}{=} \sum_{k=2}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \begin{bmatrix} k \\ r \end{bmatrix} \frac{c^k}{1 - \sum_{i=1}^{V} p_i^k}. \tag{3.11}$$

Let $h_k = (-1)^k \sum_{i=1}^{V} p_i \ ln^k \ p_i$. Then in [19] we have proved that (for some details of the derivation see also this section)

Theorem 2. For any r, c and large n the following holds

$$T_{n,r}(c) = \begin{cases} nc\{\frac{ln(nc) + \gamma - \delta_{n,0}}{h_1} + \frac{h_2}{2h_1^2} + (-1)^r f_r(nc)\} + O(1) & r = 0, 1 \\ \\ (-1)^r \ nc\{\frac{1}{r(r-1)h_1} + f_r(nc)\} + O(1) & r \geq 2 \end{cases} \tag{3.12}$$

where $\gamma = 0.571$ is the Euler constant, and $f_r(n)$ is a fluctuating function with a small amplitude defined as

$$f_r(n) = - \sum_{\{z_k^r \neq 0, r-1\}} \frac{\Gamma(z_k^r) n^{r-z_k^r}}{\sum_{i=1}^{V} p_i^{r+1-z_k} \ln p_i}. \tag{3.13}$$

The number z_k^r, $k = 0, 1, \ldots,$ $r = 0, 1, \ldots, n$ are roots of the equation

$$1 - \sum_{i=1}^{V} p_i^{r-z} = 0. \tag{3.14}$$

□

It is shown [6], [7], [10], [12], [19] that the function $f_r(n)$ has a very small amplitude and may be safely ignored in practice.

Final results of this section

Using Property 1, Lemma 2 and Theorem 1,2 we prove our main results of this section. Let us, in addition to h_n defined before, introduce a new notation $\bar{h}_n = (-1)^n \sum_{i=1}^{V} p_i \, ln^n \, (1-p_i)$. Then

Proposition 1.

(i) The mean of the successful search S_n for large n is given by

$$ES_n = \frac{1}{h_1} \{ln \; n + \rho + F_1(n)\} + O(n^{-1}) \tag{3.15}$$

where $\rho \stackrel{def}{=} \gamma - \bar{h}_1 + \dfrac{h_2}{2h_1}$ and $F_1(n)$ is a fluctuating function with a small amplitude.

(ii) The variance, var S_n, of the successful search for large n satisfies

$$\text{var } S_n = \frac{h_2 - h_1^2}{h_1^3} \, ln \; n + \alpha - 2\beta + F_2(n) + O(n^{-1}), \tag{3.16}$$

where

$$\alpha = \frac{1}{h_1^2}\left[\frac{\pi^2}{6} + \gamma^2 + \frac{3}{2}\,\frac{h_2^2}{h_1^3} + \frac{2\gamma h_2}{h_1} - \frac{2}{3}\,\frac{h_3}{h_1} + h_2 + \bar{h}_2 + 2h_1\,\bar{h}_1\right] - 2(h_1 + \bar{h}_1)\,\frac{\gamma h_1 + h_2}{h_1^3} + \frac{\rho}{h_1}\left(1 - \frac{\rho}{h_1}\right) \tag{3.17}$$

and

$$\beta = \frac{1}{h_1} \sum_{\lambda=1}^{V} \sum_{\upsilon=1}^{V} p_\upsilon p_\lambda \sum_{l=0}^{\infty} \sum_{\{i_\Sigma = l\}} \binom{l}{\mathrm{i}} \prod_{\mu=1}^{V} p_\mu^{i_\mu} \ln\{1 + \frac{p_\lambda(1-p_\upsilon)}{1-p_\lambda} \prod_{\mu=1}^{V} p_\mu^{i_\mu}\}. \tag{3.18}$$

In particularly, for V-ary symmetric Patricia tries $h_2 = h_1^2$ and (3.16) is reduced to (the coefficient at ln n is equal to zero)

$$\text{var } S_n = \frac{\pi^2}{6\ln^2 V} + \frac{1}{12} - \frac{2}{\ln V}\ln\{\prod_{l=1}^{\infty}\left(1 + \frac{1}{V^l}\right)\} + F_2(n) + O(n^{-1}). \tag{3.19}$$

(iii) The n-th moment $E(S_n)^m$, of the successful search satisfies

$$\lim_{n \to \infty} \frac{E(S_n)^m}{ln^m n} = \frac{1}{h_1^m} \tag{3.20}$$

□

We now compare the successful search for regular tries and Patricia tries. Let $S_n^{[T]}$ and $S_n^{[P]}$ denote S_n for the regular trie and the Patricia trie respectively. Then by the proposition and the results from [22] we easily see that $ES_n^{[T]} - ES_n^{[P]} = \bar{h}_1/h_1$. The variance of S_n for a regular trie is given by (3.16) with $\beta = 0$ [22]. Table 1 compares the variances for tries and the Patricia tries in the symmetric case. For $V = 2$ (symmetric case) the result was previously obtained by Kirschenhofer and Prodinger [13] (see also [12]).

V	$\mathrm{var}\, S_n^{[T]}$	$\mathrm{var}\, S_n^{[P]}$
2	3.507	1.000
3	1.446	0.630
4	0.939	0.500
5	0.718	0.430
6	0.596	0.387

Finally, let us mention that the recurrence for the variance is very slowly "convergent" to its asymptotic approximation (at least 150 terms must be computed), so the above results are particularly useful.

3.2 The first two moments of S_n

The mean value of S_n is l_n^1/n, where l_n^1 is given by (3.4). But (3.4) is a recurrence of type (3.7) with $a_n = n(1 - \sum_{i=1}^{V} p_i^n)$. Using (3.8) we find $\hat{a}_n = n \sum_{i=1}^{V} p_i^{n-1}$ for $n \geq 2$, hence by Theorem 1 the solution of (3.4) is

$$l_n^1 = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \binom{k}{1} \frac{\sum_{i=1}^{V} p_i (1-p_i)^{k-1}}{1 - \sum_{i=1}^{V} p_i^k} \tag{3.21}$$

To find an asymptotic approximation of (3.21) we apply Theorem 2. Note that

$$l_n^1 = \sum_{i=1}^{V} \frac{p_i}{1-p_i} T_{n,1}(1-p_i)$$

hence, by (3.3) and (3.12) we obtain our result (3.15) with

$$F_1(n) = \sum_{i=1}^{V} p_i f_1[n(1-p_i)]. \tag{3.22}$$

The second moment l_n^2 is more difficult to compute. As before assume for simplicity $V=2$. Note that by (3.5) l_n^2 satisfies the recurrence (3.7), but now $a_n = (1-p^n-q^n)(l_n^1 - n)$. Hence, the inverse sequence $\hat{a}_n$ is (we use (3.8))

$$\hat{a}_n = 2\hat{l}_n^1 + 2(\delta_{n,1} - npq^{n-1} - npq^{n-1}) - 2I(p^n l_n^1) - 2I(q^n l_n^1). \tag{3.23}$$

where $I(p^n l_n^1)$ is the inverse relation to $p^n l_n^1$. Two problems are encountered in obtaining the exact expression for $\hat{a}_n$. Namely, we must know $\hat{l}_n^1$, and - what is more intricate - the inverse sequence to $p^n l_n^1$. The first problem is easy. We know that l_n^1 is given by (3.21). But by the definition of the inverse relations $l_n^1 = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \hat{l}_k^1$ and $\hat{l}_0 = \hat{l}_1 = 0$. Hence combining this with (3.21) we immediately find

$$\hat{l}_k^1 = k \frac{\sum_{i=1}^{V} p_i (1-p_i)^{k-1}}{1 - \sum_{i=1}^{V} p_i^k}. \tag{3.24}$$

Let us compute now $I(p^n l_n^1)$. We first prove

Lemma 3. Let a_n and $\hat{a}_n$ are given, and let $b_n = p^n a_n$, where $0 \le p < 1$. Then

$$\hat{b}_n = \sum_{j=0}^{n} \binom{n}{j} \hat{a}_j p^j (1-p)^{n-j}. \tag{3.25}$$

Proof. Using well known relationships for binomial coefficients (see Riordan [23]) we find

$$b_n = \sum_{k=0}^{n}(-1)^k \binom{n}{k} a_k\, p^k = \sum_{k=0}^{n}(-1)^k \binom{n}{k} p^k \sum_{j=0}^{k}(-1)^j \binom{k}{j} \hat{a}_j =$$

$$\sum_{j=0}^{n} \binom{n}{j} \hat{a}_j\, p^n \sum_{k=0}^{n-j}(-1)^k \binom{n-j}{k} (1/p)^{n-k-j} = \sum_{j=0}^{n} \binom{n}{j} \hat{a}_j\, p^j (1-p)^{n-j}.$$

□

Using Lemma 3 and (3.24) (once again for $V = 2$) we immediately obtain that

$$I(p^n l_n^1) = \sum_{j=0}^{n} \binom{n}{j} \binom{j}{1} p^j (1-p)^{n-j} \frac{pq^{j-1} + qp^{j-1}}{1 - \sum_{i=1}^{V} p_i^j} \tag{3.26}$$

Therefore, using (3.26), (3.23), (3.24) and (3.9) from Theorem 1, we finally derive an exact solution for l_n^2.

Theorem 3. The second semi-factorial moment l_n^2 has the following solution

$$l_n^2 = 2 \sum_{k=2}^{n}(-1)^k \binom{n}{k} \binom{k}{1} \frac{\{\sum_{i=1}^{V} p_i (1-p_i)^{k-1}\}\{\sum_{i=1}^{V} p_i^k\}}{(1 - \sum_{i=1}^{V} p_i^k)^2} -$$

$$2 \sum_{k=2}^{n}(-1)^k \binom{n}{k} \frac{1}{1 - \sum_{i=1}^{V} p_i^k} \sum_{j=2}^{k} \binom{k}{j} \binom{j}{1} \frac{\{\sum_{i=1}^{V} p_i^j (1-p_i)^{k-j}\}\{\sum_{i=1}^{V} p_i (1-p_i)^{j-1}\}}{1 - \sum_{i=1}^{V} p_i^j} \tag{3.27}$$

Proof. The details of algebraic manipulations are left to the reader.

□

Asymptotic approximation for var S_n

We compute here the asymptotic approximation for the variance, var S_n, of the successful search. For this we need asymptotic expression for (3.27). Let us denote by A_n and B_n the first and the second term in (3.27), that is, $l_n^2 = 2A_n - 2B_n$.

The expression for A_n does not fall into (3.11) (since the denominator is raised to the power two), hence Theorem 2 cannot be applied. However, for asymptotic approximation of A_2 we need to evaluate the following

$$T_n^{(2)}(c) = \sum_{k=2}^{n} (-1)^k \left\{ {n \atop k} \right\} \left\{ {k \atop 1} \right\} \frac{c^k}{(1 - \sum_{i=1}^{V} p_i^k)^2} \tag{3.28}$$

which generalized (3.11). We use the Mellin transform technique [4], [11], [14], [22]. Performing some algebra over (3.28) we can prove that (for details see [22])

$$T_n^{(2)}(c) = -\frac{nc}{2\pi i} \int_{-1/2-\infty}^{-1/2+i\infty} \frac{\Gamma(z)(nc)^{-z}}{(1 - \sum_{i=1}^{V} p_i^{1-z})^2}\, dz + O(1) \tag{3.29}$$

where $\Gamma(z)$ is the gamma function [1]. The evaluation of the integral is standard: we must compute residues of the function under the integral to the right of the line ($1/2 - i\infty$, $-1/2 + i\infty$). Let z_k $k = 0, \pm1, \pm2, \ldots$ denote roots of the denominator, that is,

$$1 - \sum_{i=1}^{V} p_i^{1-z} = 0. \tag{3.30}$$

Note that $z_0 = 0$, while the other roots are complex. For more detailed treatment of (3.30) see [6], [12], [22]. The most difficult to handle is the pole $z_0 = 0$ since it is the double pole of the denominator and, in addition, singular point of the gamma function. But, the following Taylor expansions of the functions under the integral are available [1], [4], [11]

$$\Gamma(z) = z^{-1} - \gamma + \frac{1}{2}\left[\frac{\pi^2}{6} + \gamma^2\right] z + O(z^2)$$

$$(nc)^{-z} = 1 - z \ln nc + \frac{z^2}{2} \ln^2 nc + O(z^3)$$

$$(1 - \sum_{i=1}^{V} p_i^{1-z})^2 = z^2[b_0 + b_1 z + b_2 z^2] + O(z^5)$$

where $b_0 = h_1^2$, $b_1 = h_1 h_2$, $b_2 = \frac{1}{4} h_2^2 + \frac{1}{3} h_1 h_3$ (for details see [22]), and as before $h_k = (-1)^k \sum_{i=1}^{V} p_i \ln^k p_i$. The algorithm to compute the residue at $z_0 = 0$ is provided in [11],

[22]. Let $g(z)$ denote the function under the integral in (3.29). Then, after some algebra

$$res_{z_0} g(z) = \frac{1}{2h_1^2} \ln^2 nc + \varepsilon \ln nc + \delta \tag{3.31}$$

where

$$\varepsilon = \frac{\gamma h_1 + h_2}{h_1^3}$$
$$\delta = \frac{1}{h_1^2}\left[\frac{\pi^2}{12} + \frac{\gamma^2}{2} + \frac{3h_2^2}{4h_1^2} + \frac{\gamma h_2}{h_1} - \frac{h_3}{3h_1}\right]. \tag{3.32}$$

On the other hand, the residues of $g(z)$ at z_k, $k \neq 0$ are

$$res_{z_k} g(z) = \frac{\Gamma(z_k)(nc)^{-z_k}}{h_1^2(z_k)} \tag{3.33}$$

where $h_1(z_k) \overset{def}{=} -\sum_{i=1}^{V} p_i^{-z_k} \ln p_i$. So finally by (3.29)–(3.33) we obtain

$$T_n^{(2)} = \frac{1}{2h_1^2} \ln^2 nc + \varepsilon \ln nc + \delta + f^{(2)}(nc) + O(1) \tag{3.34}$$

where

$$f^{(2)}(n) = \sum_{\substack{k \neq -\infty \\ k \neq 0}}^{\infty} \frac{\Gamma(z_k) n^{-z_k}}{h_1^2(z_k)}$$

Using the above we prove

Lemma 4. The coefficient A_n (the first term in (3.27)) for large n is equal to

$$A_n = n\left\{\frac{1}{2h_1^2} \ln^2 n + \left[\varepsilon - \frac{h_1 + \bar{h}_1}{h_1^2}\right] \ln n + \eta + F_A(n)\right\} + O(1) \tag{3.35a}$$

where

$$\eta = \frac{1}{2h_1^2}(h_2 + \bar{h}_2 + 2h_1 \bar{h}_1) - \varepsilon(h_1 + \bar{h}_1) + \delta \tag{3.35b}$$

$$\bar{h}_k = (-1)^k \sum_{i=1}^{V} p_i \ln^k(1 - p_i) \quad \text{and}$$
$$F_A(n) = \sum_{i=1}^{V} \sum_{j=1}^{V} p_i\, p_j\, f^{(2)}[n\, p_j(1 - p_i)] \tag{3.36}$$

Proof: To prove (3.35) it is enough to note that (3.27) and (3.28) imply

$$A_n = \sum_{i=1}^{n} \frac{p_i}{1-p_i} \sum_{j=1}^{V} T_n^{(2)} \left[p_j(1-p_i)\right]$$

Using (3.34) after some algebra one finds (3.35).

□

To evaluate the second term in (3.27), namely B_n, we need some additional computations.

Let

$$B_n = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{B_k'}{1 - \sum_{i=1}^{V} p_i^k} \tag{3.37a}$$

where, after some simple algebra, one obtains

$$B_k' = \sum_{j=2}^{k} \binom{k}{j} \binom{j}{1} \frac{\sum_{\lambda=1}^{V} p_\lambda^j \,(1-p_\lambda)^{k-j} \sum_{\upsilon=1}^{V} p_\upsilon (1-p_\upsilon)^{j-1}}{1 - \sum_{i=1}^{V} p_i^j} \tag{3.37b}$$

Now we develop the denominator in a geometric series and for $\mathrm{i} = (i_1, i_2, \ldots, i_V)$ such that $i_1 + i_2 + \cdots + i_V = l$ (l is an integer) we denote $c_{\mathrm{i}} = \prod_{\mu=1}^{V} p_\mu^{i_\mu}$. Then (3.37b) becomes

$$B_k' = k \sum_{\lambda=1}^{V} \sum_{\upsilon=1}^{V} p_\upsilon p_\lambda \sum_{l=0}^{\infty} \sum_{\{i_z = l\}} \binom{l}{\mathrm{i}} c_{\mathrm{i}} \{[c_{\mathrm{i}}\, p_\lambda (1-p_\upsilon) + 1 - p_\lambda]^{k-1} - (1-p_\lambda)^{k-1}\}$$

But by (3.37a) and our notation (3.11) one finds that

$$B_k' = \sum_{\lambda=1}^{V} \sum_{\upsilon=1}^{V} p_\upsilon p_\lambda \sum_{l=0}^{\infty} \sum_{\{i_z = l\}} \binom{l}{\mathrm{i}} c_{\mathrm{i}} \left\{ \frac{T_{k,1}[c_{\mathrm{i}}\, p_\lambda(1-p_\upsilon) + 1 - p_\lambda]}{c_{\mathrm{i}}\, p_\lambda (1-p_\upsilon) + 1 - p_\lambda} - \frac{T_{k,1}(1-p_\lambda)}{1-p_\lambda} \right\}$$

Hence by Theorem 2 (see (3.12)) we finally obtain

$$B_n = n\,\beta = n \left\{ \frac{1}{h_1} \sum_{\lambda=1}^{V} \sum_{\upsilon=1}^{V} p_\lambda p_\upsilon \sum_{l=0}^{\infty} \sum_{\{i_z = l\}} \binom{l}{\mathrm{i}} c_{\mathrm{i}} \ln[1 + p_\lambda (1-p_\upsilon) c_{\mathrm{i}} / (1-p_\lambda)] + F_B(n) \right\} \tag{3.38}$$

where

$$F_B(n) = \sum_{\lambda=1}^{V} \sum_{\upsilon=1}^{V} p_\upsilon p_\lambda \sum_{l=0}^{\infty} \sum_{\{i_k = l\}} \binom{l}{\mathbf{i}} c_1 \{f_1[n\,(c_1 p_\lambda(1-p_\upsilon) + 1 - p_\lambda)] - f_1(n(1-p_\lambda)\} \quad (3.39)$$

and the constant β is defined as B_n/n. Note that we prove in fact that $B_n = O(n)$. This we shall use later to derive higher moments of S_n.

Finally, taking into account (3.3), (3.27), (3.35) and (3.38) we prove

Theorem 4. The second factorial moment s_n^2 of the successful search S_n is given by

$$s_n^2 = 2\,A_n/n - 2B_n/n = \frac{1}{h_1^2}\,ln^2 n + 2\,[\varepsilon - \frac{h_1 + \bar{h}_1}{h_1^2}]\ln n + 2\eta - 2\beta + 2F_A(n) - 2F_B(n) + O(n^{-1}) \quad (3.40)$$

where η and β are given by (3.35b) and (3.38), and $F_A(n)$, $F_B(n)$ are presented in (3.36) and (3.39).

□

To compute the variance of S_n we note that var$S_n = s_n^2 + ES_n - (ES_n)^2$, hence the Proposition 1(ii) follows with $\alpha = 2\eta + \frac{\rho}{h_1}(1 - \frac{\rho}{h_1})$ (see definition of ρ just after (3.15) and for η see (3.35b)). The function $F_2(n)$ in (3.17) is

$$F_2(n) = 2F_A(n) - 2F_B(n) + F_1(n) - [F_1(n)]^2 \quad (3.41)$$

where the terms of (3.41) are defined in (3.36), (3.39) and (3.22). Note also that for the symmetric Patricia trie $h_1 = \ln V$ and $h_k = h_1^k$. In particular, the coefficient at $\ln n$ disappears and var $S_n = \alpha - 2\beta + O(n^{-1})$ where $\alpha = \frac{\pi^2}{6\,\ln^2 V} + \frac{1}{12}$ and 2β is the third term in (3.19). It is interesting that for a regular symmetric V-ary trie var $S_n = \alpha + O(n^{-1})$ with α the same as for the Patricia.

3.3 Higher moments of S_n

In Lemma 2 we have established the recurrence equation (3.6) for the m-th semi-factorial moment of the external path length. By Property 1 the m-th factorial moment of S_n, s_n^m, is equal to l_n^m/n. Note that the recurrence (3.6) is of type (3.7) with the additive term a_n equal to

$$a_n = m! \sum_{k=1}^{m} (-1)^{m-k} \frac{l_n^{k-1}}{(k-1)!} - m! \left\{ \sum_{i=1}^{V} p_i^n \right\} \left\{ \sum_{k=1}^{n} (-1)^{m-k} \frac{l_n^{k-1}}{(k-1)!} \right\}. \qquad (3.42)$$

Let us denote the first and the second term in (3.42) by $a_n^{(1)}$ and $a_n^{(2)}$, respectively, that is, $a_n = a_n^{(1)} + a_n^{(2)}$. But (3.6) is a linear recurrence, hence $l_n^m = \overline{l}_n^m - \underline{l}_n^m$, where $a_n^{(1)}$ contributes to $\overline{l}_n^m$ and $a_n^{(2)}$ to $\underline{l}_n^m$. We prove that $\underline{l}_n^m = O(n)$, hence $s_n^m = \overline{l}_n^m/n + O(1)$.

Lemma 5. For any n and m $l_n^m \le \overline{l}_n^m$ and

$$\overline{l}_n^m = \frac{n}{h_1^m} \ln^m n + n \frac{m}{h_1^m} \ln^{m-1} n \left[\gamma + \frac{m}{2}\frac{h_2}{h_1} - (m-1)h_1 - h_1^m F(n)\right] + O(n\ ln^{m-2} n) \qquad (3.43)$$

where $F(n)$ is a fluctuating function with a small amplitude.

Proof: We recognize that $\overline{l}_n^m$ is the m-th semi-factorial moment of the external path length in a regular digital trie (see [22]). Then (3.43) is proved in [22].

□

We now prove that $\underline{l}_n^m = O(n)$. Note that by $l_n^m \le \overline{l}_n^m$, and by (3.43) we can find such constants $\xi_0, \xi_1, \ldots, \xi_m$ that $a_n^{(2)}$ is bounded, and

$$a_n^{(2)} \le \sum_{i=1}^{V} p_i^n \sum_{r=0}^{m} \xi_k \binom{n}{r} \stackrel{def}{=} A_n^m$$

Let now y_n be a sequence

$$y_n^m = A_n^m + \sum_{\{j_\Sigma = n\}} \binom{n}{\mathbf{j}} p_1^{j_1} \cdots p_V^{j_V} [y_{j_1}^m + \cdots + y_{j_V}^m] \qquad (3.44)$$

By linearity of the recurrence for l_n^m we have for each m

$$l_n^m \le y_n^m \qquad (3.45)$$

Hence, to prove that l_n^m is $O(n)$ it is enough to show that $y_n^m = O(n)$. But

Lemma 6. For any m and large n

$$y_n^m = O(n)$$

Proof: The recurrence (3.44) is of type (3.7), hence Theorem 1 and 2 may be applied. By (3.12) after some algebra we find

$$y_n^m = \frac{n}{h_1} \sum_{i=1}^{V} \left\{ \xi_0 [p_i + (1-p_i)\ln(1-p_i)] + \xi_1 p_i \ln(1-p_i) + \sum_{r=2}^{m} \xi_r \frac{p_i^r}{r(r-1)(1-p_i)^{r-1}} \right\} = O(n)$$

□

Finally by Lemma 5, 6 and (3.45) we prove our last result of this section.

Theorem 6. For $m > 2$ the m-th factorial moment of the successful search S_n is

$$s_n^m = \frac{1}{h_1^m} \ln^m n + \frac{n}{h_1^m} \ln^{m-1} n [\gamma + \frac{m}{2}\frac{h_2}{h_1} - (m-1)h_1 - h_1^m F(n)] + O(\ln^{m-2} n) \qquad (3.46)$$

□

This, in a trivial way, implies our Proposition 1(iii).

4. UNSUCCESSFUL SEARCH

The unsuccessful search is neither simply related to the external path length nor to the successful search, since unsuccessful searches are more likely to occur at external nodes near the root. This makes the analysis much more difficult. Therefore, we consider only binary symmetric Patricia tries, that is, $V = 2$ and $p_1 = p_2 = 0.5$ is assumed, however, we derive asymptotic approximation for all moments of the unsuccessful search. The organization of this section is the same as for Section 3.

4.1. Initial and final results

Let $H_n^{(m)}(½)$ denote the m-th derivative of the generating function $H_n(z)$ (see (2.3) in Lemma 1) at $z = ½$. Then one proves

Property 2. For any integer m the following holds

$$H_n(½) = 1 \tag{4.1a}$$

$$u_n^m = \frac{1}{2^m} H_n^{(m)}(½) \tag{4.1b}$$

Proof. For binary symmetric Patricia tries ($V = 2,\ p_1 = p_2 = 0.5$) the generating function $H_n(z)$ from Lemma 1 becomes

$$H_n(z) = z\ 2^{1-n} \sum_{k=1}^{n} \binom{n}{k} H_k(z) - 2^{1-n}(z-1)\, H_n(z), \quad n \geq 2 \tag{4.2}$$

Substituting $z = ½$ in (4.2) one proves (4.1a). The average value of the unsuccessful search, u_n^1 is $\sum_{l=0}^{\infty} l\ H_l 2^{-1} = ½\, H_n^{(1)}(½)$, since we end up at a given external node on level l with probability 2^{-l} (by H_l we denote the number of external nodes at level l). For $m = 2$ we have $u_n^2 = \sum_{l=0}^{\infty} l(l-1)H_l 2^{-l} = (½)^2 H_n(½)$, and so on. This proves (4.16).

□

To compute the factorial moments u_n^m we must find the derivatives of the recurrence (4.2). Differentiating (4.2) and using (4.1b) for $m = 1$ one immediately obtains

$$\begin{gathered} u_0^1 = u_1^1 = 0 \\ u_n^1(2^n - 2) = 2^n - 2 + \sum_{k=1}^{n-1} \binom{n}{k} u_k^1 \end{gathered} \tag{4.3}$$

For the second factorial moment we must differentiate (4.2) twice. Then

$$u_n^2(2^n-2)=2\sum_{k=1}^{n-1}\binom{n}{k}u_k^1+\sum_{k=1}^{n-1}\binom{n}{k}u_k^2$$

But the first term of the above might be computed from (4.3). Then finally

$$u_n^2(2^n-2)=2(2^n-2)(u_n^1-1)+\sum_{k=1}^{n-1}\binom{n}{k}u_k^2 \tag{4.4}$$

Generalizing the above, we find

Lemma 7. For any integers n and m, the m-th factorial moment of U_n satisfies

$$u_0^m=u_1^m=0$$

$$u_n^m(2^n-2)=m(2^m-2)\left[u_n^{m-1}+\sum_{k=1}^{m}(-1)^k(m-k)u_n^{m-k}\right]+\sum_{k=1}^{n-1}\binom{n}{k}u_k^m \tag{4.5}$$

and $u_n^0 \stackrel{def}{=} 1$

Proof. The proof uses induction arguments and is left to the reader.

□

Note that (4.5) is a system of recurrences. To compute u_n^m we need $u_n^1,\ u_n^2,\ldots,\ u_n^{m-1}$. But recurrences of type (4.5) have a common pattern, and for various m they differ only by additive terms. To solve (4.5) we shall use extensively a solution of a recurrence discussed in Szpankowski [20]. Below, we summarize the most important results from [20].

Solution of a recurrence equation

Let $x_0, x_1, \ldots, x_n$ be a sequence of number such that

$$\begin{aligned}&\text{given}\quad x_0=x_1=0,\ x_2,\ldots,x_N\\&\text{solve}\quad x_n(2^n-2)=2^n a_n+\sum_{k=1}^{n-1}\binom{n}{k}x_k\qquad n>N\end{aligned} \tag{4.6}$$

where N is an integer, and a_n is a given, but otherwise arbitrary sequence. It turns out that the solution of (4.6) depends on the so called *Bernoulli inverse relations* (see Riordan [23]). Define for an a_n a new sequence $\tilde{a}_n$ as

$$\tilde{a}_n = \sum_{k=0}^{n} \left\{ \begin{matrix} n \\ k \end{matrix} \right\} B_k \; a_{n-k} \tag{4.7a}$$

where B_k are the Bernoulli numbers defined as the coefficients of the Taylor expansion of $z(e^z - 1)^{-1}$, that is,

$$\frac{z}{e^z - 1} = \sum_{k=0}^{\infty} B_k \; \frac{z^k}{k!} \tag{4.8}$$

For more details about Bernoulli numbers see [1], [15], [20], [23]. The sequence $\tilde{a}_n$ and a_n are called inverse pair since [23]

$$a_n = \sum_{k=0}^{n} \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \frac{\tilde{a}_{n-k}}{k+1} = \frac{1}{n+1} \sum_{k=0}^{n} \left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} \tilde{a}_k \tag{4.7b}$$

hence also $\tilde{\tilde{a}}_n = a_n$. Using the above, we have proved in [20] that

Theorem 7. The solution of (4.6) is given by

$$x_n = b_n + \frac{1}{n+1} \sum_{k=2}^{n} \left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} \frac{\tilde{b}_k}{2^{k-1} - 1} \tag{4.9}$$

where

$$b_n = a_n + g_n \; \chi_{(n \le N)} \tag{4.10a}$$

$$g_k = x_k(1 - 2^{-k}) - a_k - 2^{-k} \sum_{i=1}^{k} \left\{ \begin{matrix} k \\ i \end{matrix} \right\} x_i \quad k = 1, 2, \ldots, N \tag{4.10b}$$

$$\tilde{b}_k = \tilde{a}_k - a_0 B_k + \sum_{i=1}^{N} \left\{ \begin{matrix} k \\ i \end{matrix} \right\} g_i \; B_{k-i} \tag{4.11}$$

and $\chi_{(n \le N)}$ is the indicator function.

□

Note also that (4.9) and (4.7b) imply that the inverse sequence $\tilde{x}_n$, to x_n defined in the recurrence

(4.6) is

$$\tilde{x}_n = \tilde{b}_n + \frac{\tilde{b}_n}{2^{n-1}-1} \qquad n \geq 2 \tag{4.12}$$

To find asymptotic approximation of x_n we need more information about a_n and $\tilde{a}_n$. In particular, it is proved that [20], [23]

$$a_n = \left[\begin{matrix} n \\ r \end{matrix} \right] q^n \rightarrow \tilde{a}_n = \left[\begin{matrix} n \\ r \end{matrix} \right] q^r B_{n-r}(q) \tag{4.13}$$

where r is an integer, and $0 < q < 1$, while $B_n(q)$ denotes the Bernoulli polynomial defined as [1], [15], [20]

$$\frac{ze^{tz}}{e^z - 1} = \sum_{k=0}^{\infty} B_k(t) \frac{z^k}{k!} \tag{4.14}$$

Some more inverse pairs the reader may find is [23]. For example, it is easy to prove that

$$a_n = \frac{1}{n+1} \qquad \tilde{a}_n = \delta_{n0} \tag{4.15a}$$

$$a_n = 1 \qquad \tilde{a}_n = B_n + \delta_{n1} \tag{4.15b}$$

Let us now restrict our considerations to a_n given by (4.13), and define

$$R_{n,r}(q) = \frac{1}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \left[\begin{matrix} k \\ r \end{matrix} \right] \frac{B_{k-r}(q)}{2^{k-1}-1} \tag{4.16}$$

The following is proved in [20].

Theorem 8. For large n the below holds

$$R_{n,0}(q) = (½ + \delta_{q,1} - q)(lg\, n - ½ + \frac{\gamma}{\ln 2}) + \frac{\zeta'(1 - q + \delta_{q,1})}{\ln 2} + f_0(n) + O(n^{-1}) \tag{4.17a}$$

$$R_{n,1}(q) = lgn - ½ + \frac{\gamma}{\ln 2} - \frac{\psi(1 - q + \delta_{q,1})}{\ln 2} + f_1(n) + O(n^{-1}) \tag{4.17b}$$

$$R_{n,r}(q) = \frac{1}{r \ln 2} \zeta(r, 1 - q + \delta_{q,1}) + \frac{1}{r!} f_r(n) + O(n^{-1}), \quad r \geq 2 \tag{4.17c}$$

where $\psi(x)$ is the psi function, $\zeta(z,q)$ is the generalized Riemman zeta function ($\zeta(z) = \zeta(z,1)$) [1], [11], [24], and

$$f_r(n) = \frac{1}{\ln 2} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \zeta(r + 2\pi ik/\ln 2)\Gamma(r + 2\pi ik/\ln 2)\exp[-2\pi ik \; lgn] \tag{4.17d}$$

and $lgn = lg_2 n$.

□

The function $f_r(n)$ is a fluctuating function with a small amplitude and may be safely ignored in practice [14], [20].

Final results of this section

Using the recurrence (4.5), and Theorems 7 and 8 we prove our main result of this section.

Proposition 2.

(i) The mean of the unsuccessful search is

$$EU_n = lgn - \theta + F_1(n) + O(n^{-1}) \tag{4.18}$$

where

$$\theta = \frac{\ln \pi - \gamma}{\ln 2} - ½ = 0.31875 \tag{4.19}$$

and $F_1(n)$ is a fluctuating function with a small amplitude.

(ii) The variance, var U_n, of U_n satisfies

$$\begin{aligned} \text{var } U_n &= 4(\alpha - \beta - \theta - 2) - \theta - \theta^2 + F_2(n) + O(n^{-1}) \\ &\approx 0.87904 \end{aligned} \tag{4.20}$$

where

$$\alpha = ½\,\theta + \frac{23}{24} + \frac{1}{\ln^2 2}\left[\frac{\pi^2}{24} + \frac{\gamma^2}{4} - \frac{\gamma \ln 2\pi}{2} - \zeta_2\right] \tag{4.21a}$$

$$\zeta_2 = ½\,\zeta''(0) = \frac{\gamma^2}{4} + \frac{c_1}{2} - \frac{\pi^2}{48} - \frac{\ln^2(2\pi)}{4} \tag{4.21b}$$

with $c_1 = -0.0728158$ and

$$\beta = \theta + \tfrac{1}{2} + \frac{2}{\ln 2} \sum_{k=2}^{\infty} \frac{\zeta(k) 2^{-k}}{k} \left\{ \sum_{\lambda=2}^{\infty} 2^{-k(\lambda-1)} \left[\sum_{i=1}^{2^{\lambda-1}-1} i^k - \frac{2^{\lambda-1}}{k+1} + \tfrac{1}{2} \right] - \frac{1}{2(k+1)} \right\} \approx 0.48738, \tag{4.21c}$$

and $F_2(n)$ is a fluctuating function with a small amplitude.

(iii) The m-th moment, $E(U_n)^m$,of U_n satisfies

$$\lim_{n \to \infty} \frac{E(U_n)^m}{lg^m\, n} = 1. \tag{4.22}$$

□

4.2. The first two moments of U_n

The mean value of U_n is given by recurrence (4.3). This falls into our general recurrence (4.6) with $N = 2$ and $a_n = 1 - 2^{1-n}$, and $b_n = a_n$ and $\bar{b}_n = 2[B_k - B_k(\tfrac{1}{2})] = 4B_k[1 - 2^{-k}]$ (we use here the following identity $B_n(\tfrac{1}{2}) = B_n(2^{1-n} - 1)$ [1]). Hence, by Theorem 7 we obtain

$$u_n^1 = 2 - \frac{4}{n+1} + 2\delta_{n0} + \frac{2}{n+1} \sum_{k=2}^{n} \binom{n+1}{k} \frac{B_k}{2^{k-1} - 1} \tag{4.23}$$

Note also that by (4.12) and (4.15) the inverse sequence to u_n^1 is given by

$$\bar{u}_n^1 = 4B_n + 2\delta_{n1} - 4\delta_{n0} + (1 - \delta_{n0} - \delta_{n1}) \frac{2B_n}{2^{n-1} - 1} \tag{4.24}$$

The asymptotic approximation of (4.23) follows directly from (4.16) and Theorem 8. Namely, for large n

$$U_n^1 = 2R_{n,0}(1) = lg - \theta + f_0(n) + O(n^{-1})$$

as needed in Proposition 2(i).

The analysis of the second moment is much more intricate. The recurrence is given by (4.4). This recurrence does *not* fall exactly in our general recurrence (4.6). Therefore, let us split the additive term into two parts, namely

$$2(2^n - 2)[u_n^1 - 1] = 2 \cdot 2^n [u_n^1 - 1 + 2^{1-n}] + 4u_n^1$$

and let the solution of (4.4) derived from the first part of the above be denoted as $U_n^{(1)}$, while the solution that follows from the second part is defined as $U_n^{(2)}$. By linearity of (4.4) we have

$$u_n^2 = U_n^{(1)} - U_n^{(2)} \tag{4.25}$$

Note that for $U_n^{(1)}$ we have the following recurrence

$$U_0^{(1)} = U_1^{(1)} = 0$$
$$(2^n - 2)U_n^{(1)} = 2^n \cdot 2[u_n^1 - 1 + 2^{1-n}] + \sum_{k=1}^{n-1} \left[\begin{matrix} n \\ k \end{matrix} \right] U_k^{(1)}$$

and this falls into (4.6) with $a_n = 2[u_n^1 - 1 + 2^{1-n}]$. Since $\tilde{a}_n = 2[\tilde{u}_n^1 - B_n - \delta_{n1} + 2B_n(½)]$, and by (2.24) we immediately obtain from Theorem 7 (we use the following identity $B_n(½) = B_n(2^{1-n} - 1)$ [1])

$$U_n^{(1)} = \frac{8}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k}{2^{k-1} - 1} + \frac{4}{n+1} \sum_{k=2}^{n} \left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} \frac{B_k}{(2^{k-1} - 1)^2} \tag{4.26}$$

The analysis of $U_n^{(2)}$ is much more difficult. Note that the recurrence for $U_n^{(2)}$ is *not* exactly of type (4.6), since the additive term is $4u_n^1$ (not $2^n a_n$ as required). But $u_n^1 = 2^n (2^{-n} u_n^1)$, and for the solution, we need the inverse sequence to $2^{-n} u_n^1$. But

Lemma 8. Let $A_n = q^n a_n$, and $\tilde{a}_n$ is given. Then

$$\tilde{A}_n = \sum_{j=0}^{n} \left[\begin{matrix} n \\ j \end{matrix} \right] \tilde{a}_j q^{j-1} \frac{B_{n+1-j}(q) - B_{n+1-j}}{n+1-j} \tag{4.27}$$

Proof. In the proof we use identities from [1], [23] and (4.7). We have

$$\tilde{A}_n = \sum_{k=0}^{n} \left[\begin{matrix} n \\ j \end{matrix} \right] B_{n-k} q^k a_k = \sum_{k=0}^{n} \left[\begin{matrix} n \\ k \end{matrix} \right] B_{n-k}\, q^k \sum_{j=0}^{k} \left[\begin{matrix} k \\ j \end{matrix} \right] \frac{1}{k+1-j} \tilde{a}_j =$$

$$\sum_{j=0}^{n} \left[\begin{matrix} n \\ j \end{matrix} \right] \tilde{a}_j q^j \sum_{k=0}^{n-1} \left[\begin{matrix} n-j \\ k \end{matrix} \right] \frac{1}{k+1} B_{n-j-k} q^k = \sum_{j=0}^{n} \left[\begin{matrix} n \\ j \end{matrix} \right] \tilde{a}_j q^{j-1} \frac{1}{n-j+1} \sum_{k=1}^{n-j+1} \left[\begin{matrix} n+1-j \\ k \end{matrix} \right] B_{n+1-j-k} q^k =$$

$$\sum_{j=0}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \bar{a}_j \, q^{j-1} \, \frac{B_{n+1-j}(q) - B_{n+1-j}}{n+1-j}$$

□

Let now $A_n = 2^{-n} u_n^1$. Then, using (4.27) and (4.24) one shows that

$$\bar{A}_n = 2[B_n(½) - B_n] - 8\,\frac{B_{n+1}(½) - B_{n+1}}{n+1} + 2V_n$$

where

$$V_k = \sum_{j=2}^{k} \left(\begin{matrix} k \\ j \end{matrix} \right) \frac{B_j}{2^{j-1}-1} \, 2^{1-j} \, \frac{B_{k+1-j}(½) - B_{k+1-j}}{k+1-j}$$

Finally, applying Theorem 7 and the above we obtain

Theorem 9. The solution to u_n^2 given by recurrence (4.4) is $u_n^2 = U_n^{(1)} - U_n^{(2)}$ where $U_n^{(1)}$ is done in (4.26) and

$$U_n^{(2)} = 4 \cdot 2^{-n} u_n^1 + \frac{8}{n+1} \sum_{k=2}^{n} \begin{bmatrix} n+1 \\ k \end{bmatrix} \frac{B_k(½) + B_k}{2^{k-1}-1} - \frac{32}{n+1} \sum_{k=2}^{n} \begin{bmatrix} n+1 \\ k \end{bmatrix} \frac{B_{k+1}(½) - B_{k+1}}{(k+1)(2^{k-1}-1)} + \tag{4.29}$$

$$\frac{8}{n+1} \sum_{k=2}^{n} \begin{bmatrix} n+1 \\ k \end{bmatrix} \frac{1}{2^{k-1}-1} \sum_{j=2}^{k} \left(\begin{matrix} k \\ j \end{matrix} \right) \frac{B_j}{2^{j-1}-2} \, 2^{1-j} \, \frac{B_{k+1-j}(½) - B_{k+1-j}}{k+1-j}$$

□

Asymptotic approximation for $U_n^{(1)}$

Let

$$R_n^{(2)} = \frac{1}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k}{(2^{k-1}-1)^2} \tag{4.30}$$

Then by (4.26), (4.16) and the above $U_n^{(1)} = 8\, R_{n,0}(1) + 4\, R_n^{(2)}$. But by Theorem 8 we have

$$R_{n,0}(1) = ½ \, lg \; n \; - ½ \, \theta - 1 + f_0(n) + O(n^{-1}) \tag{4.31}$$

Hence to find approximation for $U_n^{(1)}$ we need asymptotic analysis for $R_n^{(2)}$. But applying the Mellin transform to (4.30) one proves (see [3], [20].)

$$R_n^{(2)} = \frac{1}{2\pi i} \int_{-\frac{1}{2}+i\infty}^{-\frac{1}{2}+i\infty} \frac{\zeta(z)\Gamma(z)n^{-z}}{(2^{-z}-1)^2}\, dz \tag{4.32}$$

and to evaluate $R_n^{(2)}$ we need residues of the function under the integral. Note that the roots of the denominator are $z_k = 2\pi ik/\ln 2$, with $z_0 = 0$. In addition, there is a single pole of the zeta function $\zeta(z)$, at $z = 1$ [1]. The root $z_0 = 0$ is a triple pole since $z_0 = 0$ is also a singular point of the gamma function , $\Gamma(z)$. It turns out also that this pole makes the biggest contribution to the asymptotic analysis of (4.32). To find the residue at $z_0 = 0$ we use the following Taylor expansions [1], [4], [11], [24].

$$n^{-z} = 1 - z\ \ln n + \frac{\ln^2 n}{2}\, z^2 + O(z^3)$$

$$\Gamma(z) = z^{-1} - \gamma + \tfrac{1}{2}\,[\frac{\pi^2}{6} + \gamma^2]\, z + O(z^2)$$

$$\frac{1}{(2^{-z}-1)} = \frac{1}{\ln^2 2}\, z^{-2} + \frac{1}{\ln 2}\, z^{-1} + \frac{5}{12} + O(z)$$

and

$$\zeta(z) = -\tfrac{1}{2} - \frac{z}{2}\ln 2\pi + \zeta_2\, z + O(z^2)$$

where $\zeta_2 = \frac{1}{2}\,\zeta''(0)$. The value of ζ_2 is computed in [3, p.204], and

$$\zeta_2 = \tfrac{1}{2}\,\zeta''(0) = \frac{\gamma^2}{4} + \frac{c_1}{2} - \frac{\pi^2}{48} - \frac{1}{4}\ln^2(2\pi) = -1.0032 \tag{4.33}$$

where $c_1 = -0.0728158$. Multiplying the above and taking the coefficient at z^{-1} one finds the desired residue. In a similar way the residues at z_k might be obtained. Computations show that

$$R_n^{(2)} = \frac{1}{4}\, lg^2 n - \tfrac{1}{2}\,[2.5 + \theta]\, lg\ n + \delta + F_0(n) + f_0(n) + O(n^{-1})$$

where

$$\delta = \tfrac{1}{2}\,\theta + \frac{23}{24} + \frac{1}{\ln^2 2}[\frac{\pi^2}{24} + \frac{\gamma^2}{4} - \frac{\gamma\ln 2\pi}{2} - \zeta_2]$$

$$F_0(n) = \frac{1}{\ln^2 2} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} [\zeta(z_k))\Gamma'(z_k) + \Gamma(z_k)\zeta'(z_k) - \ln n\,\Gamma(z_k)]\exp[-2\pi ik\ lg\ n] \tag{4.34}$$

Finally using the above and (4.31) we find

$$U_n^{(1)} = lg^2 n - [1+2\theta] lg\ n + 4(\delta - \theta - 1) + 12 f_0(n) + 4F_0(n) + O(n^{-1}) \qquad (4.35)$$

where θ and δ are defined in (4.19) and (4.34).

Upper and lower bounds for $U_n^{(2)}$

We now study $U_n^{(2)}$. It is easy to show that $U_n^{(2)}$ satisfies the following recurrence

$$(2^n - 2)U_n^{(2)} = 4u_n^1 + \sum_{k=1}^{n-1} \begin{bmatrix} n \\ k \end{bmatrix} U_k^{(2)} \quad n \geq 2 \qquad (4.36)$$

with $U_0^{(2)} = U_1^{(2)} = 0$. Since (4.36) is not of type (4.6), hence we explore some other methods to evaluate $U_n^{(2)}$. In this subsection, we give a tight lower bound and a tight upper bound on $U_n^{(2)}$. In fact, we prove that $U_n^{(2)} = O(1)$.

Note that by Proposition 2 (i) $u_n^1 = lg\ n + O(1)$, hence we can find such constants ξ_0, ξ_1 and ξ_2 that $\xi_0 \leq u_n^1 \leq \xi_1 n + \xi_2$. This implies that upper and lower bound for u_n^1 might be established through Theorem 7 and 8, since for the lower bound we assume $a_n = \xi_0 2^{-n}$ while for the upper bound we set $a_n = \xi_1\ n 2^{-n} + \xi_2 2^{-n}$, and these fall into our recurrence (4.6). The accuracy of our evaluation depends, in fact, on a good approximation of u_n^1 for small values of n, say, $n \leq N$. In fact, we assume that we know $u_0^1 = u_1^1 = 0$ and $u_2^1, \ldots, u_N^1$. Then

Lemma 9. For $n > N$ the following holds

$$\xi_0 \leq u_n^1 \leq \xi_1 n + \xi_2 \qquad (4.37)$$

with $\xi_0 = u_{N+1}^1$, $\xi_1 = [(N+1)\ln^2]^{-1}$, $\xi_2 = \xi_0 - 1/\ln 2$.

Proof. The proof uses induction applied to recurrence (4.3), and is left to the reader.

□

Let us now define two sequences $\underline{x}_n$ and $\overline{x}_n$ as

$$\underline{x}_0 = \underline{x}_1 = 0,\ \ \underline{x}_2 = U_2^{(2)}/4\ , \ldots,\ \underline{x}_N = U_N^{(2)}/4$$
$$(2^n - 2)\underline{x}_n = \xi_0 + \sum_{k=1}^{n-1} \binom{n}{k} \underline{x}_k \quad n > N \tag{4.38}$$

and

$$\overline{x}_0 = \overline{x}_1 = 0,\ \ \overline{x}_2 = U_2^{(2)}/4\ , \ldots,\ \overline{x}_N = U_N^{(2)}/4$$
$$(2^n - 2)\overline{x}_n = \xi_2 + \xi_1\, n + \sum_{k=1}^{n-1} \binom{n}{k} \overline{x}_k \quad n > N. \tag{4.39}$$

Note that by Lemma 9 $4\underline{x}_n \le U_n^{(2)} \le 4\overline{x}_n$. The asymptotic approximations for (4.38) and (4.39) are available by Theorem 7 and 8 with $a_n = \xi_0 2^{-n}$ and $a_n = \xi_1 n 2^{-n} + \xi_2 2^{-n}$ respectively. Hence

Theorem 10. For large n the following holds

$$\underline{x}_n = 0.5\xi_0[\theta + 0.5] + \frac{1}{\ln 2} \sum_{r=2}^{N} \frac{\zeta(r) G_r}{r} + O(n^{-1}) \tag{4.40a}$$

with

$$G_r = \underline{x}_r - 2^{-r}[\underline{x}_k + \xi_0 + \sum_{i=1}^{r} \binom{r}{k} \underline{x}_i], \quad r = 1, 2, \ldots, N$$

and

$$\overline{x}_n = \xi_1 + 0.5\xi_2(\theta + 0.5) + \frac{1}{\ln 2} \sum_{r=2}^{N} \frac{\zeta(r) g_r}{r} + O(n^{-1}) \tag{4.40b}$$

with

$$g_r = \overline{x}_r - 2^{-r}[\overline{x}_r + \xi_1 r + \xi_2 + \sum_{i=1}^{N} \binom{r}{i} \overline{x}_i] \quad r = 1, 2, \ldots, N$$

□

Note that by Theorem 10 we have proved that $U_n^{(2)} = O(1)$. Let $U_n^{(2)} = 4\beta$, and $\underline{\beta}$, $\overline{\beta}$ be the lower and the upper bound for β, that is, $\underline{x}_n = \underline{\beta}$ and $\overline{x}_n = \overline{\beta}$. The accuracy of β evaluation depends on N. Table 2 contains $\underline{\beta}$ and $\overline{\beta}$ for $2 \le N \le 6$.

Table 2.

N	$\underline{\beta}$	$\overline{\beta}$
2	0.46574	0.49869
3	0.48020	0.49031
4	0.48479	0.48824
5	0.486411	0.48766
6	0.48701	0.48748

We prove in the next subsection that $\beta = 0.487385$, which confirms the above approximations. In fact, the method established here can be used to solve the recurrence (4.6) in the case when Theorem 7 and 8 are not applicable, that is, when the additive term is not of the form $2^n a_n$. For example, if the additive term in (4.6) is $lg\ n$, then using our approach, we can prove that $0.4997 \le x_n \le 0.5001$.

Exact asymptotic analysis of $U_n^{(2)}$

Although we have obtained above very tight bounds for $U_n^{(2)}$ it is interesting to see if exact asymptotic analysis of $U_n^{(2)}$ is possible. This interest follows not only from "pure mathematical whim", but such a solution enables to extend the analysis of (4.6) to the case when the additive term is any sequence of numbers, not particularly $2^n a_n$. This finds many applications in practice.

An asymptotic solution of $U_n^{(2)}$ depends on finding an appropriate approximation on the last term of (4.29), that is

$$V_n^{(1)} \stackrel{def}{=} \frac{1}{n+1} \sum_{k=2}^{n} \binom{n+1}{k} \frac{V_k}{2^{k-1}-1} \qquad (4.41a)$$

where V_k is given by (4.28), i.e.

$$V_k \stackrel{def}{=} \sum_{j=2}^{k} \binom{k}{j} \frac{B_j}{2^{j-1}-1} 2^{1-j} \frac{B_{k+1-j}(1/2) - B_{k+1-j}}{k+1-j} \qquad (4.41b)$$

To apply Theorem 8 we must express V_n in terms of Bernoulli polynomials $B_n(x)$ as (4.16) suggests. Note that developing the denominator $(2^{j-1}-1)$ in (4.41b) in a geometric series we obtain

$$V_n = \sum_{\lambda=2}^{\infty} \sum_{j=2}^{n} \begin{bmatrix} n \\ j \end{bmatrix} B_j \, 2^{-\lambda(j-1)} \, \frac{B_{n+1-j}(\tfrac{1}{2}) - B_{n+1-j}}{n+1-j} \tag{4.42}$$

We prove

Lemma 10. Let $q = \tfrac{1}{2}$ and define

$$T_\lambda^n = \sum_{j=0}^{n} \begin{bmatrix} n \\ j \end{bmatrix} B_j \, q^{\lambda(j-1)} \, \frac{B_{n+1-j}(q) - B_{n+1-j}}{n+1-j} \tag{4.43}$$

Then

$$T_\lambda^n = \sum_{l=1}^{2^{\lambda-1}-1} B_n(lq^l) + B_n \tag{4.44}$$

Proof. Let $T_\lambda(z)$ be the exponential generating function for T_λ^n. Then multiplying both sides of (4.43) by $z^k/k!$ one finds

$$T_\lambda(z) = \frac{z}{e^z - 1} \, \frac{e^{z/2} - 1}{e^{z2^{-\lambda}} - 1} \tag{4.45}$$

(The easiest way to show (4.45) is by using so called generalized Bernoulli polynomials as defined in [15]). Now use the fact that $(e^{z/2} - 1) = (e^{z/4} - 1)(e^{z/4} + 1) = (e^{z/8} - 1)(e^{z/8} + 1)(e^{z/4} + 1)$ and so on. Finally we obtain ($q = \tfrac{1}{2}$)

$$T_\lambda(z) = \frac{z}{e^z - 1} \prod_{k=2}^{\lambda} (1 + e^{zq^k}) \tag{4.46}$$

For example, for $\lambda = 3$

$$T_3(z) = \frac{z}{e^z - 1} \, \frac{e^{zq} - 1}{e^{zq^3} - 1} = \frac{z}{e^z - 1} (e^{zq^3} + 1)(e^{zq^2} + 1) = \frac{ze^{z(q^2+q^3)}}{e^z - 1} + \frac{ze^{zq^2}}{e^z - 1} + \frac{ze^{zq^3}}{e^z - 1} + \frac{z}{e^z - 1}$$

Hence by finding explicit formula for the product in (4.46) we obtain terms as above. But each of this term is the generating function for a Bernoulli polynomial. For example, the first term in the above is the generating function for $B_n(q^2 + q^3)$. This proves, after some additional algebra, (4.44).

□

Comparing now (4.42) and (4.43) we find that

$$V_n = \sum_{\lambda=2}^{\infty} \left[T_\lambda^n - 2^\lambda \frac{B_{n+1}(½) - B_{n+1}}{n+1} + \frac{B_n(½) - B_n}{2} \right] \tag{4.47}$$

Because of (4.44), (4.41a) and Theorem 8 the asymptotic approximation of $V_n^{(1)}$ is easy to obtain, if one evaluates the following

$$r_n(q) = \frac{1}{n+1} \sum_{k=0}^{n} \binom{n+1}{k} \frac{B_{k+1}(q) - B_{k+1}}{(k+1)(2^{k-1} - 1)} \tag{4.48}$$

Then $U_n^{(2)} = 4\beta$, where by (4.29) and the above

$$\beta = 2[R_{n,0}(½) + R_{n,0}(1)] - 8r_n(½) +$$
$$2 \sum_{\lambda=2}^{\infty} [R_{n,0}(\lambda 2^{-\lambda}) + ½ R_{n,0}(1) - 2^\lambda r_n(½) + ½ R_{n,0}(½)] \tag{4.49}$$

The appropriate approximation of $R_{n,0}(q)$ is given in Theorem 8. We need only asymptotic approximation for $r_n(½)$.

Lemma 11. For large n the following holds

$$r_n(½) = \frac{1}{8}(lgn + \frac{\gamma}{\ln 2} - ½) + \frac{1}{\ln 2} \int_0^{½} \zeta'(1-t)dt + ½ f_0(n) + O(n^{-1}) \tag{4.50a}$$

where

$$\int_0^{½} \zeta'(1-t)dt = \frac{\gamma}{8} - \frac{\ln 2\pi}{4} + \sum_{k=2}^{\infty} \frac{\zeta(k) 2^{-k-1}}{k(k+1)} \tag{4.50b}$$

Proof. We know that [1], [15]

$$\frac{B_{k+1}(q) - B_{k+1}}{n+1} = \int_0^q B_k(t)dt$$

hence

$$r_n(q) = \int_0^q \frac{1}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k(t)}{2^{k-1}-1}\, dt = \int_0^q R_{n,0}(t)dt$$

Therefore, Theorem 8 (4.17a) might be used. Noting that [4] $\zeta'(1-t) = \ln\Gamma(1-t) - ½\ln 2\pi$ and [4]

$$\ln\Gamma(1-t) = \gamma t + \sum_{k=2}^{\infty} \frac{\zeta(k)t^k}{k}$$

we finally obtain (4.50).

□

Using (4.49), (4.50) and (4.17) we easily figure out the constant β. In fact, all coefficients at $lg\ n$ disappear, and the first two terms of (4.49) give

$$2[R_{n,0}(½) + R_{n,0}(1)] - 8r_n(½) = \theta + 0.5 - \frac{8}{\ln 2} \sum_{k=2}^{\infty} \frac{\zeta(k)2^{-k-1}}{k(k+1)}$$

while the sum in (4.49) is

$$\frac{2}{\ln 2} \sum_{k=2}^{\infty} \frac{\zeta(k)2^{-k}}{k} \sum_{\lambda=2}^{\infty} \left\{ 2^{-k(\lambda-1)} \sum_{i=1}^{2^{\lambda-1}} i^k - \frac{2^{\lambda-1}}{k+1} + ½ \right\}$$

This gives β as established in (4.21c). The Proposition 2(ii) follows from the above, (4.35) and (4.18) if one notes that $\mathrm{var}U_n = u_n^{\underline{2}} + u_n^{\underline{1}} - (u_n^{\underline{1}})^2$. The fluctuating function $F_2(n)$ in (4.21) is equal to $F_2(n) = 12f_0(n) + 4F_0(n)$, where $f_0(n)$ and $F_0(n)$ are given by (4.17d) and (4.34).

Finally note that the above approach might be used to extend the solution of our general recurrence (4.6) to the case when the additive term is a_n without the coefficient 2^n at a_n.

4.3. Higher moments of the unsuccessful search

Let

$$a_n = m[u_n^{\underline{m-1}} + \sum_{k=1}^{m} (-1)^k (m-k) u_n^{\underline{m-k}}] \tag{4.51}$$

Then by Property 2 the m-th factorial moment of the unsuccessful search satisfies the recurrence $u_n^m = 2^n a_n - 2a_n + \sum_{k=1}^{n} \left[\begin{matrix} n \\ k \end{matrix} \right] u_k^m$, and by linearity we can split u_n^m into two parts, namely $u_n^m = \overline{U}_n^{(m)} = \underline{U}_n^{(m)}$ where the first component of the above is associated with $2^n a_n$, while the second with $2a_n$. We prove that $\overline{U}_n^m = lg^m n + O(lg^{m-1} n)$ and $\underline{U}_n^{(m)} = O(1)$.

To solve for $\overline{U}_n^{(m)}$ we apply the same approach as in the case $m = 2$. Since the additive term is $2^n a_n$ we may use Theorem 7, and after some algebra one proves that

$$\overline{U}_n^{(m)} = \frac{2m!}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k}{(2^{k-1}-1)^m} + \frac{1}{n+1} \sum_{l=1}^{m-1} \chi_l \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k}{(2^{k-1}-1)^l} \quad (4.52)$$

where $\chi_l, l = 1,2,...,m-1$ are some constants. Using the Mellin transform as in Section 4.2 for (4.30) we prove that

$$R_n^{(m)} \overset{def}{=} \frac{1}{n+1} \sum_{k=2}^{n} \left[\begin{matrix} n+1 \\ k \end{matrix} \right] \frac{B_k}{(2^{k-1}-1)^m} = \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{\zeta(z)\Gamma(z)n^{-z}}{(2^{-z}-1)^m} dz + O(n^{-1})$$

and finding the residue of the function under the integral we show that

$$R_n^{(m)} = \frac{1}{2m!} lg^m n + O(lg^{m-1} n) \quad (4.53)$$

Hence by (4.52) and (4.53) we prove that $\overline{U}_n^{(m)} = lg^m n + O(lg^{m-1} n)$.

To prove $\underline{U}_n^{(m)} = O(1)$ note that by (4.51)–(4.53) we can find such constants ξ_i, $i = 0, 1, \ldots, m$ that $a_n \le \sum_{r=0}^{m} \xi_r \left[\begin{matrix} n \\ r \end{matrix} \right] \overset{def}{=} \overline{a}_n$. Let us solve now (4.6) with the additive term $2^n(2^{-n}\overline{a}_n)$. This is possible since $\overline{a}_n$ is of form (4.13). Direct applications of (4.17a)–(4.17b) shows that $\underline{U}_n^{(m)} = O(1)$ (for more details see proof of Theorem 10). This together with (4.53) establishes Proposition 2(iii).

5. FINAL REMARKS

In this paper a family of Patricia tries with n records was studied from the successful and unsuccessful point of view. We proved that the m-th moment of successful and unsuccessful searches are of order of magnitude $\ln^m n$. We gave a detailed analysis of the first two moments of S_n and U_n for large value of n. For example, we showed the the variances of S_n and U_n for binary symmetric Patricia trie are 1.000 and 0.879 respectively. These results have been achieved through extensive applications of two types of recurrences for which the solutions and asymptotic approximations were obtained elsewhere by the author. However, in deriving the variance of the unsuccessful search we had to generalize the solution of the second recurrence equation to a wider class of additive terms. This finds a number of new applications in computer science and telecommunication, e.g. in the average complexity of an algorithm generating exponentially distributed variates [8], in the performance evaluation of conflict resolution algorithm in a broadcast communication environment [9], [21], in the analysis of the refined lexicographical sorting [16] and extendible hashing [5], [6], and so on.

There are still some open problems related to an analysis of the Patricia trie. First of all, it is interesting to determine an asymptotic distribution for the successful and unsuccessful search. It is reasonable to predict that the successful search in the *asymmetric* Patricia is normally distributed with the parameters established in Proposition 1 (i)-(ii), for large n (see [12]). The symmetric case needs some additional work. Moreover, the analysis of unsuccessful search should be extended to asymmetric Patricia tries. In this paper, we have not considered at all the height of the Patricia, which is another open problem. Note that if H_n denotes the height of the Patricia with n records, then H_n is bounded in probability (on the contrary to the regular tries), that is, $Pr\{lg_V n \le H_n \le n\} = 1$. A simple upper and lower bound for the average of the height is possible. Indeed, using the above and noting that the height in the Patricia is not greater than the height in a regular trie, one shows immediately that for the binary case $lgn < EH_n \le 2lgn$ (for the height of a trie see [6] [17]). It seems that EH_n is closer to the upper bound than the lower one. However, a detailed analysis is necessary.

REFERENCES

[1] Abramowitz, M., Stegun, I., Handbook of mathematical functions, Dover (1964).

[2] Aho, A., Hopcroft, J. and Ullman, J., Data structures and algorithms, Addison-Wesley (1983).

[3] Berndt, B., Ramanojan's notebooks, Springer-Verlag (1985).

[4] Frdelyi, A., Higher transcendental functions, McGraw-Hill Book (1953).

[5] Fagin R., Nievergelt, J., Pippenger, N. and Strong H., Extendible hashing: A fast access method for dynamic files, *ACM TODS* 4 (1979), 315–344.

[6] Flajolet, Ph., On the performance evaluation of extendible hashing and trie searching, *Acta Informatica* 20 (1983), 345–369.

[7] Flajolet, Ph. and Sedgewick, R., Digital search trees revisited, *SIAM J. Comput.*, 15 (1986), 748–767.

[8] Flajolet, Ph, Saheb, N., The complexity of generating an exponentially distributed variate, to appear in *Journal of Algorithms* (1986).

[9] Gallager R., Conflict resolution in a random access broadcast networks, *Proc. AFOSR Workshop in Communication Theory and Applications* (1978), 74–76.

[10] Gonnet G., Handbook of algorithms and data structures, Addison-Wesley (1984).

[11] Henrici, P., Applied and computational complex analysis, John Wiley & Sons, New York (1977).

[12] Jacquet, Ph. and Regnier, M., Limiting distributions for trie parameters, INRIA Technical Report (1985).

[13] Kirschenhofer, P. and Prodinger, H., Some further results on digital trees, *ICALP 86*, to appear.

[14] Knuth, D. The art of computer programming, sorting and searching, Addison-Wesley (1973).

[15] Nörlund, N.E., Mémoire sur les polynomes de Bernoulli, *Acta Mathematica*, 43 (1923), 124–196.

[16] Paige, R., Tarjan, R., Three efficient algorithms based on partition refinements, preprint.

[17] Pittel, B., Paths in a random digital tree: Limiting distributions, *Adv. Appl. Probab.*, 18 (1986), 139–155.

[18] Sedgewick, R., Algorithms, Addison-Wesley (1983).

[19] Szpankowski, W., Analysis of a recurrence equation arising in stack-type algorithms for collision-detecting channels, *Proc. Intern. Seminar on Computer Networking and Performance Evaluation*, Tokyo 1985, pp. 399-412.

[20] Szpankowski, W., Solution of a linear recurrence equation arising in the analysis of some algorithms, to appear in *SIAM J. Algebraic and Discrete Methods.*

[21] Szpankowski, W., An analysis of a contention resolution algorithm - Another approach, submitted to *Acta Informatica.*

[22] Szpankowski, W. Some results on V-ary asymmetric tries, Purdue University, CSD TR-582, March (1986) (submitted to a journal).

[23] Riordan, J., Combinatorial identities, John Wiley & Sons (1968).

[24] Whittaker, E. and Watson, G., A course of modern analysis, Cambridge Press (1935).