

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



The Moments of the Profile in Random Binary Digital Trees

Ramin Kazemi and Saeid Delavar

Department of Statistics, Imam Khomeini International University, Qazvin, Iran

Corresponding author: kazemi@ikiu.ac.ir

Article history:

Received December 2012

Accepted March 2013

Available online April 2013

Abstract

The purpose of this paper is to provide a precise analysis of the t -th moment of the profile in random binary digital trees. We assume that the n input strings are independent and follow a (binary) Bernoulli model. In tries, the main difference with the previous analysis is that we have to deal with an inhomogeneous part in the proper functional equation satisfied by the t -th moment and in digital search trees with an inhomogeneous part in a proper functional-differential equation. We show that t -th moment of the profile ($t \geq 2$) is asymptotically of the same order as the expected value ($t = 1$). These results are derived by methods of analytic combinatorics.

Keywords: Digital trees, Tries, Digital search trees, Profile, The t -th moment.

1. Introduction

Digital trees like tries or digital search trees are important in many computer science applications like data compression, pattern matching or hashing (see [2], [4] and [5]). For example, the popular Lempel-Ziv compression scheme is strongly related to digital search trees [20]. They are two classes of so-called $\log n$ trees and their construction is based on digital keys and not on the order structure of the keys as in the case of binary search trees [6].

Digital trees have been widely studied in the literature (see [16], [19] and the references therein). The motivation of studying the profiles of such trees is multifold. However here we will concentrate on the t -th moment of the profile in these trees.

1.1. Binary tries

Tries are prototype data structures useful for many indexing and retrieval purposes. They were first proposed by de la Briandais [2] in the late 1950s for information processing; Fredkin [10] suggested the

current name as it being part of retrieval. Tries are multiway trees whose nodes are vectors of characters or digits. Due to their simplicity and efficiency, tries found widespread use in diverse applications ranging from document taxonomy to IP addresses lookup, from data compression to dynamic hashing, from partial-match queries to speech recognition, from leader election algorithms to distributed hashing tables (see [12], [14], [16] and [19]).

Tries are natural choice of data structures when the input records involve a notion of digits (or alphabets). They are often used to store such data so that future retrieval can be made efficient. Given a sequence of n strings over the *binary* alphabet $\Sigma = \{0,1\}$, we can construct a trie as follows. If $n = 0$, then the trie is empty. If $n = 1$, then a single (external) node holding the strings is allocated. If $n \geq 2$, then the trie consists of a root (internal) node directing strings to the two subtrees according to the first digit of each string, and strings directed to the same subtree are themselves tries (see [14], [16] and [19] for more details). Unlike other search trees such as digital search trees and binary search trees where records or keys are stored at the internal nodes, the internal nodes in tries are branching nodes used merely to direct records to each subtrees, records being all stored in external nodes that are leaves of such tries [18]. Figure 1 shows a trie built on eight strings s_1, \dots, s_8 (i.e., $s_1 = 0\dots$, $s_2 = 1\dots$, $s_3 = 01\dots$, $s_4 = 11\dots$, etc.) with internal (ovals) and external (squares) nodes.

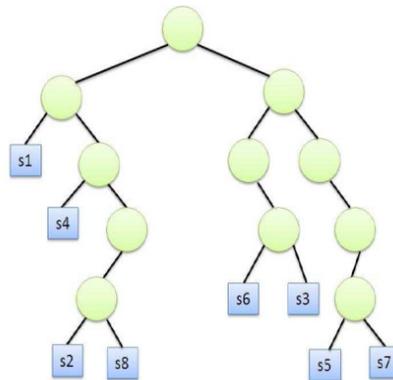


Figure 1: A trie built on eight strings s_1, \dots, s_8 .

Trie provides a model for the analysis of several important algorithms, such as Radix Exchange Sort [14], and Extendible Hashing [10].

1.2. Binary digital search trees

A *binary digital search tree* is built from a sequence of keys where each key is a string of binary digits. It is easiest for implementation that the keys the same length. The first key, or a pointer to it, is stored in the root node; a subsequent key is guided to the left or right subtree as whether its first bit is 0 or 1, respectively. In the subtrees the algorithm is repeated recursively, but if the subtree is at the level j , the j -th bit is used for branching. The process continues in the subtrees until an empty subtree is reached and that is where new key is adjoined. The process is very similar to the growth of binary search trees, only different in the branching method where the bits are used instead of ordinal values of the keys. The nodes containing of strings are called the internal nodes and the remaining nodes which are directly

attached to nodes already existing in the tree are called the external nodes [7], [8]. These trees have introduced by Coffman and Eve in 1970 and also are called *Coffman-Eve sequence trees* [3]. Figure 2 shows a digital search tree built on eight strings s_1, \dots, s_8 .

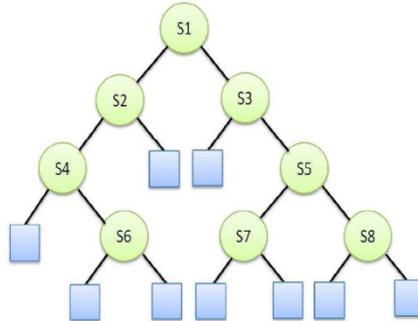


Figure 2: A digital search tree built on eight strings s_1, \dots, s_8 .

A main distinction between the algorithm for digital search trees and that for tries is that all the nodes of the digital tree hold keys, whereas in tries the keys reside only in leaves. Digital search tree provides a model for the analysis of several important algorithms, such as the Lempel-Ziv parsing algorithm [15], and Conflict Resolution [17].

1.3. Biased random binary digital trees

The *profile* of a digital tree is a parameter that represents the number of nodes (either internal or external) with the same distance to the root. It is a function of the number of strings stored in a tree and the distance from the root. Several, if not all, digital tree parameters such as height, size, depth, shortest path, and fill-up level can be expressed in terms of the (external and internal) profiles [7], [18].

Throughout the paper, we write $B_{n,k}^T$ to denote the number of external nodes (leaves) at distance k from the root; the number of internal nodes at distance k from the root is denoted by $I_{n,k}^T$ in binary tries. For simplicity, we will refer to $B_{n,k}^T$ as the external profile and $I_{n,k}^T$ the internal profile. Also we write $B_{n,k}^D$ and $I_{n,k}^D$ for binary digital search trees.

The standard data model for digital trees is the Bernoulli probability distribution (infinitely long independent keys of independent bits). The probability model should ideally be unbiased. In practice this unbiased is not guaranteed [1].

In this paper we study the profiles of a binary digital tree built over n binary strings generated by a Bernoulli model. More precisely, we assume that the input is a sequence of n independent and identically distributed random variables, each being composed of an infinite sequence of Bernoulli random variables with mean p , where $0 < p < 1$ is the probability of a "1" and $q = 1 - p$ is the probability of a "0". The corresponding digital tree constructed from these n bit-strings is called a random digital tree. We consider the Bernoulli model with unequal probabilities of symbols (biased model). More precisely, we assume that $p < q$.

The purpose of this paper is to provide a precise analysis of the t -th moment of the profile in digital trees. In tries, the main difference with the previous analysis is that we have to deal with an inhomogeneous part in the proper *functional equation* satisfied by the t -th moment. Also in digital search trees we have to deal with an inhomogeneous part in a proper *functional-differential equation* satisfied by the t -th moment. However, we show that the t -th moment is asymptotically of the same order as the expected value in two cases. These results are derived by methods of analytic combinatorics such as generating functions, Mellin transform, Poissonization, the saddle point method and singularity analysis.

2. Results on binary tries

The probability generating function of $B_{n,k}^T$, $P_{n,k}^T(u) = E(u^{B_{n,k}^T})$, satisfies the following recurrence relation [18]

$$P_{n,k}^T(u) = \sum_{j=0}^n \binom{n}{j} p^j q^{n-j} P_{j,k-1}^T(u) P_{n-j,k-1}^T(u), \quad n \geq 2, k \geq 1 \tag{1}$$

with initial conditions $P_{0,k}^T(u) = 1(k \geq 0)$, $P_{1,0}^T(u) = u$, $P_{1,k}^T(u) = 1(k \geq 1)$ and $P_{n,0}^T(u) = 1(n \geq 1)$. From (1) one gets directly a recurrence relation for the exponential generating functions

$$M_k(x, u) = \sum_{n \geq 0} P_{n,k}^T(u) \frac{x^n}{n!}$$

of the form

$$M_k(x, u) = M_{k-1}(px, u)M_{k-1}(qx, u) + (P_{1,k}^T(u) - P_{1,k-1}^T(u))x, k \geq 1,$$

with initial conditions $M_0(x, u) = e^x + x(u - 1)$. With the help of the initial conditions of $P_{1,k}^T(u)$ it follows that

$$M_1(x, u) = e^x + (e^{px}qx + e^{qx}px - x)(u - 1) + pqx^2(u - 1)^2$$

and

$$M_k(x, u) = M_{k-1}(px, u)M_{k-1}(qx, u), \quad k \geq 2. \tag{2}$$

The corresponding probability generating functions for the internal profile, $\bar{P}_{n,k}^T(u) = E(u^{I_{n,k}^T})$ satisfy the same recurrence (1) but with initial conditions $\bar{P}_{n,0}^T(u) = u(n \geq 2)$, $\bar{P}_{n,k}^T(u) = 1$ for $n \leq 1$ and $k \geq 0$. Similarly to the external profile the exponential generating function

$$\bar{M}_k(x, u) = \sum_{n \geq 0} \bar{P}_{n,k}^T(u) \frac{x^n}{n!}$$

satisfy

$$\overline{M}_k(x, u) = \overline{M}_{k-1}(px, u)\overline{M}_{k-1}(qx, u), \quad k \geq 1, \tag{3}$$

with initial condition $\overline{M}_0(x, u) = ue^x - (1 + x)(u - 1)$.

Let $\mu_{n,k,t}^T = E[(B_{n,k}^T)^t]$ be the t -th moment of the external profile in binary tries. By taking the derivative of order t with respect to u and setting $u = 1$ from $M_k(x, u)$ we obtain for the exponential generating function

$$B_k^{(t)}(x) = \sum_{n \geq 0} \mu_{n,k,t}^T \frac{x^n}{n!}$$

the following functional recurrence

$$\begin{aligned} B_k^{(t)}(x) &= e^{qx} B_{k-1}^{(t)}(px) + e^{px} B_{k-1}^{(t)}(qx) \\ &+ \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \lambda(m, n) B_{k-1}^{(m)}(px) B_{k-1}^{(n)}(qx), \quad k \geq 2 \end{aligned} \tag{4}$$

with initial conditions $B_0^{(t)}(x) = x$ and

$$B_1^{(t)}(x) = px e^{px} + qx e^{qx} + \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \gamma(m, n) p q x^2 - x,$$

where $\gamma(m, n) \in \mathbb{Z}$.

Let $\mathbf{B}_k^{(t)}(x) = e^{-x} B_k^{(t)}(x)$ be the Poisson transform $B_k^{(t)}(x)$. Thus recurrence (4) translates into

$$\begin{aligned} \mathbf{B}_k^{(t)}(x) &= \mathbf{B}_{k-1}^{(t)}(px) + \mathbf{B}_{k-1}^{(t)}(qx) \\ &+ \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \lambda(m, n) \mathbf{B}_{k-1}^{(m)}(px) \mathbf{B}_{k-1}^{(n)}(qx), \quad k \geq 2 \end{aligned} \tag{5}$$

where $\lambda(m, n) \in \mathbb{Z}$, since $p + q = 1$. If $t = 1$, then

$$\mathbf{B}_k^{(1)}(x) = \mathbf{B}_{k-1}^{(1)}(px) + \mathbf{B}_{k-1}^{(1)}(qx).$$

This case analyzed by Park, Hwang, Nicodeme and Szpankowski [18]. If $t = 2$, then $\lambda(1,1) = 2$. If $t = 3$, then $\lambda(1,2) = \lambda(2,1) = 3$ and another $\lambda(m, n) = 0$ and so on.

Lemma 1. [18] Suppose $x = r e^{i\theta}$, where $r \geq 0$ and $|\theta| \leq \pi$. Then

$$|B_k^{(1)}(x)| \leq B_k^{(1)}(r) e^{-cr\theta^2}$$

uniformly for all $k = k(n) \geq 1$ and some constant $c > 0$.

Lemma 2. Suppose $x = r e^{i\theta}$, where $r \geq 0$ and $|\theta| \leq \pi$. Then

$$|B_k^{(t)}(x)| \leq B_k^{(t)}(r) e^{-cr\theta^2} \tag{6}$$

for some constant c independent of r, k and θ .

Proof. By iterating the equation (5) we can represent $\mathbf{B}_k^{(t)}(x)$ by

$$\begin{aligned} \mathbf{B}_k^{(t)}(x) &= \sum_{\ell=0}^k \binom{k-1}{\ell} \mathbf{B}_1^{(t)}(p^\ell q^{k-1-\ell}x) \\ &+ \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \varphi(m, n) \mathbf{W}_{k,m,n}(x) \\ &+ \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \varphi(m, n) \sum_{j=1}^{k-2} \sum_{\ell=0}^j \binom{j}{\ell} \mathbf{W}_{k-j,m,n}(p^\ell q^{j-\ell}x), \end{aligned} \tag{7}$$

where $\mathbf{W}_{k,m,n}(x) = \mathbf{B}_{k-1}^{(m)}(px)\mathbf{B}_{k-1}^{(n)}(qx)$ for $k \geq 2$ and again $\varphi(m, n) \in \mathbb{Z}$. Now we can prove (6) by induction and Lemma 1.

It is clear that $\mathbf{B}_k^{(t)}(x)$ can be represented as (see equation (5))

$$\mathbf{B}_k^{(t)}(x) = \sum_{i=1}^{2^k} x^i f_{i,k}^{(t)}(x),$$

where $f_{i,k}^{(t)}(x)$ can be represented as a finite linear combinations of functions of the form $\exp\{-\sum_{\ell_1} \sum_{\ell_2} p^{\ell_1} q^{\ell_2} x\}$ with $0 \leq \ell_1 + \ell_2 \leq k$. The proof of our result for the t -th moment of the profile relies on a precise analysis of the recurrence (5) for $\mathbf{B}_k^{(t)}(x)$ that is based on the Mellin transform and analytic de-Poissonisation. Let $Mellin\{\mathbf{B}_k^{(t)}(x); s\}$ denote the Mellin transform [11]

$$Mellin\{\mathbf{B}_k^{(t)}(x); s\} = \int_0^\infty \mathbf{B}_k^{(t)}(x)x^{s-1}dx$$

with $\Re(s) > -2$ for $k \geq 1$. Thus

$$Mellin\{\mathbf{B}_k^{(t)}(x); s\} = \sum_{i=1}^{2^k} \Gamma(s+i)F_{i,k}^{(t)}(s), \quad k \geq 2,$$

where $F_{i,k}^{(t)}(s)$ can be represented as a finite linear combinations of functions of the form $\{\sum_{\ell_1} \sum_{\ell_2} p^{\ell_1} q^{\ell_2}\}^{-s}$ and

$$Mellin\{\mathbf{B}_1^{(t)}(x); s\} = \Gamma(s+1) \left(p^{-s} + q^{-s} + (s+1) \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \lambda(m, n)pq - 1 \right).$$

Also we can rewrite $Mellin\{\mathbf{B}_k^{(t)}(x); s\}$ as $\Gamma(s+1)F_k^{(t)}(s)$ where

$$F_k^{(t)}(s) = F_{1,k}^{(t)}(s) + \sum_{i=1}^{2^k} (s+1) \cdots (s+i)F_{i+1,k}^{(t)}(s).$$

For a real number α with $(\log \frac{1}{p})^{-1} < \alpha < (\log \frac{1}{q})^{-1}$, let

$$\rho = \rho(\alpha) = \frac{1}{\log(p/q)} \log \left(\frac{1 - \alpha \log(1/p)}{\alpha \log(1/q) - 1} \right). \tag{8}$$

Equivalently, α and ρ satisfy the equation

$$\alpha = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho} \log \frac{1}{p} + q^{-\rho} \log \frac{1}{q}}.$$

Furthermore, we set

$$\beta(\rho) = \frac{p^{-\rho} q^{-\rho} \log(p/q)^2}{(p^{-\rho} + q^{-\rho})^2}, \tag{9}$$

and we also use the abbreviations

$$\alpha_0 = \frac{2}{\log \frac{1}{p} + \log \frac{1}{q}}, \quad \alpha_1 = \frac{1}{\log \frac{1}{p}}, \quad \alpha_2 = \frac{p^2 + q^2}{p^2 \log \frac{1}{p} + q^2 \log \frac{1}{q}}.$$

It is natural to choose $\rho = \rho_{n,k}$ as the saddle point of the function

$$T(s)^k n^{-s} = e^{k \log T(s) - s \log n}$$

which satisfies the equation

$$\frac{k}{\log n} = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho} \log \frac{1}{p} + q^{-\rho} \log \frac{1}{q}},$$

that is $\rho = \rho_{n,k} = \rho(k/\log n)$. Note also that on the line $\Re(s) = \rho$ there will be infinitely many saddle points

$$s_j = \rho + \frac{2\pi i j}{\log \frac{p}{q}}, \quad (j \in \mathbb{Z})$$

since $|T(s_j)| = T(\rho)$. Consequently, the behavior of $T(s)^k z^{-s}$ around $s = s_j$ is almost the same as that of $T(s)^k z^{-s}$ around $s = \rho$. This phenomenon gives a periodic leading factor in the asymptotics of $B_k^{(t)}(n) (\overline{B}_k^{(t)}(n))$ and then also in that of $\mu_{n,k,t}^T (\overline{\mu}_{n,k,t}^T)$.

Lemma 3. Let $T(s) = p^{-s} + q^{-s}$ and

$$f(s) = \frac{F_k^{(t)}(s)}{T(s)^k}.$$

Suppose that $\alpha_1 + \varepsilon \leq k/\log n \leq \alpha_2 - \varepsilon$ (for some $\varepsilon > 0$). Then

$$\mathbf{B}_k^{(t)}(n) = H(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) \right),$$

where

$$H(\rho, x) = \sum_{j \in \mathbb{Z}} f(\rho + it_j) \Gamma(\rho + it_j + 1) e^{-2j\pi i x} \tag{10}$$

is a non-zero periodic function with period 1 in x and $t_j = 2j\pi/\log(p/q)$.

Furthermore, for $|\theta| \leq \pi/2 - \varepsilon$ (for some $\varepsilon > 0$), we have

$$\begin{aligned} \mathbf{B}_k^{(t)}(ne^{i\theta}) &= \frac{T(\rho_{n,k})^k}{\sqrt{2\pi\beta(\rho_{n,k})k}} \sum_{|j| \leq j_0} f(\rho_{n,k} + it_j) \Gamma(\rho_{n,k} + it_j) (ne^{i\theta})^{-\rho_{n,k} - it_j} p^{-ikt_j} \\ &\times \left(1 + O\left(\frac{1}{\sqrt{k}}\right)\right), \end{aligned} \tag{11}$$

where $j_0 = \sqrt{\log n} \log(p/q) / (2\pi)$.

Proof. By the inverse Mellin transform

$$\begin{aligned} \mathbf{B}_k^{(t)}(x) &= \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} \Gamma(s + 1) F_k^{(t)}(s) x^{-s} ds \\ &= \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} \Gamma(s + 1) f(s) T(s)^k x^{-s} ds, \end{aligned} \tag{12}$$

with $\rho > -2$ where

$$f(s) = \frac{F_k^{(t)}(s)}{T(s)^k}.$$

The proof is completed quite identical to that of Lemma 5 in [7] for the new function $f(s)$.

Theorem 1. Let $\mu_{n,k,t}^T$ denote the t -th moment of the external profile in binary tries with $0 < p < q < 1$. Let k and n be positive integers. Then the following assertions hold:

1. If $\alpha_1 + \varepsilon \leq \frac{k}{\log n} \leq \alpha_2 - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\mu_{n,k,t}^T = H(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O(k^{-1/2})\right),$$

where $H(\rho, x)$ is as (10).

2. If $k = \alpha_2(\log n + \xi \sqrt{\alpha_2 \beta(-2) \log n})$, where $\xi = o((\log n)^{\frac{1}{6}})$, then

$$\mu_{n,k,t}^T = -n^2 F_k^{(t)}(-2) \Phi(\xi) \left(1 + O\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right),$$

where Φ is the normal distribution function .

3. If $\frac{k}{\log n} \geq \alpha_2 + \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\mu_{n,k,t}^T = -n^2 F_k^{(t)}(-2)(1 + \mathcal{O}(n^{-\eta}))$$

for some $\eta > 0$.

Proof. The asymptotics for $\mu_{n,k,t}^T$ accomplish by the analytical de-Poissonization which requires to compute the Cauchy integral

$$\mu_{n,k,t}^T = \frac{n!}{2\pi i} \int_{|x|=n} e^x \mathbf{B}_k^{(t)}(x) \frac{dx}{x^{n+1}}.$$

Lemma 3 and Lemma 2 complete the range $\alpha_1 + \varepsilon \leq \frac{k}{\log n} \leq \alpha_2 - \varepsilon$. For range $\frac{k}{\log n} \geq \alpha_2 + \varepsilon$ we can be handled in a similar way. First one has to shift the path of integration to the line $\Re(s) = \rho < -2$. The integrand has a polar singularity at $s = -2$, thus we have

$$\begin{aligned} \mathbf{B}_k^{(t)}(x) &= -f(-2)x^2 T(-2)^k \\ &+ \frac{1}{2\pi} \int_{-\infty}^{\infty} x^{-s} \Gamma(\rho + it + 1) f(\rho + it) T(\rho + it)^k dt. \end{aligned}$$

Thus the leading term of $\mathbf{B}_k^{(t)}(x)$ is $-x^{-2} F_k^{(t)}(-2)$. It is not difficult to show that $f(-2) < 0$, since

$$T(s)^k = \sum_{j=0}^k \binom{k}{j} (p^j q^{k-j})^{-s}.$$

By de-Poissonisation this term is proposed as leading term of $\mu_{n,k,t}^T$ if $x = n$. For the remaining integral we set $x = ne^{i\theta}$ and obtain

$$\int_{-\infty}^{\infty} x^{-s} \Gamma(\rho + it + 1) f(\rho + it) T(\rho + it)^k dt = \mathcal{O}\left(\frac{n^{-\rho T(\rho)^k}}{\sqrt{k}}\right)$$

uniformly for $|\theta| \leq \theta_0$ ($\theta_0 < \pi/2$) and $-M \leq \rho \leq -2 - \varepsilon$ ($M > 2$). If we choose ρ sufficiently close to -2 we obtain

$$\mathbf{B}_k^{(t)}(ne^{i\theta}) = -n^2 F_k^{(t)}(-2) e^{2i\theta} (1 + \mathcal{O}(n^{-\eta}))$$

for some $\eta > 0$. Hence from Lemma 2 and analytic de-Poissonisation we get the result in this range.

For another range we move the line of integration to the saddle point

$$\Re(s) = \rho = -2 - \frac{\xi}{\sqrt{\alpha_2 \beta(-2) \log n}} + \mathcal{O}\left(\frac{\xi^2}{\log n}\right)$$

and proof is completed just similar to [18].

Theorem 2. Let $\bar{\mu}_{n,k,t}^T$ denote the t -th moment of the internal profile in binary tries with $0 < p < q < 1$. Let k and n be positive integers. Then the following assertions hold:

1. If $\alpha_1 + \varepsilon \leq \frac{k}{\log n} \leq \alpha_0 - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\bar{\mu}_{n,k,t}^T = 2^k - H(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right),$$

where $H(\rho, x)$ is a non-zero periodic function with period 1 in x .

2. If $k = \alpha_0(\log n + \xi\sqrt{\alpha_0\beta(0)\log n})$, where $\xi = o((\log n)^{\frac{1}{6}})$, then

$$\bar{\mu}_{n,k,t}^T = 2^k \Phi(-\xi) \left(1 + \mathcal{O}\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right).$$

3. If $\alpha_0 + \varepsilon \leq \frac{k}{\log n} \leq \alpha_2 - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\bar{\mu}_{n,k,t}^T = H(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right).$$

4. If $k = \alpha_2(\log n + \xi\sqrt{\alpha_2\beta(-2)\log n})$, where $\xi = o((\log n)^{\frac{1}{6}})$, then

$$\bar{\mu}_{n,k,t}^T = -\frac{1}{4pq} \Phi(\xi) n^2 F_k^{(t)}(-2) (p^2 + q^2) \left(1 + \mathcal{O}\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right).$$

5. If $\frac{k}{\log n} \geq \alpha_2 + \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\bar{\mu}_{n,k,t}^T = -\frac{1}{4pq} n^2 F_k^{(t)}(-2) (1 + \mathcal{O}(n^{-\eta}))$$

for some $\eta > 0$.

Proof. The analysis uses exactly the same methods of proof of Theorem 1. The Poisson transform $\bar{\mathbf{B}}_k^{(t)}(x)$ is given by

$$\bar{\mathbf{B}}_k^{(t)}(x) = 2^k - \frac{1}{2\pi i} \int_{\rho-i\infty}^{\rho+i\infty} (s+1)\Gamma(s)T(s)^k x^{-s} ds,$$

where $\rho > 0$. There is a second phase transition due to the additional polar singularity at $\rho = 0$.

Remark 1. The results stated in Theorem 1 and Theorem 2 are not optimal. For example, the range $\alpha_1 + \varepsilon \leq \frac{k}{\log n} \leq \alpha_2 - \varepsilon$ can be weakened to

$$\alpha_1(\log n - \log \log \log n - K_n) \leq k \leq \alpha_2(\log n - K_n \sqrt{\log n}),$$

where K_n is any sequence tending to infinity [6, 18]. Thus the t -th moment of the internal profile is asymptotically equivalent to 2^k for $k \leq \alpha_0(\log n - K_n \sqrt{\log n})$. Also when $k \geq \alpha_2(\log n + K_n \sqrt{\log n})$, then $\bar{\mu}_{n,k,t}^T \sim T(-2)\mu_{n,k,t}^T/pq$ as it was for the expected value [18].

Remark 2. For $t = 1$,

$$F_k^{(1)}(s) = (p^{-s} + q^{-s} - 1)(p^{-s} + q^{-s})^{k-1},$$

$$f(s) = \frac{p^{-s} + q^{-s} - 1}{p^{-s} + q^{-s}},$$

$$\text{Mellin}\{\mathbf{B}_1^{(1)}(x); s\} = \Gamma(s + 1)(p^{-s} + q^{-s} - 1),$$

$$\text{Mellin}\{\mathbf{B}_k^{(1)}(x); s\} = \Gamma(s + 1)(p^{-s} + q^{-s} - 1)(p^{-s} + q^{-s})^{k-1}.$$

We point out that if we set $\alpha = k/\log n$, then we can rewrite

$$(p^{-\rho} + q^{-\rho})^k n^{-\rho} = n^{\alpha \log(p^{-\rho} + q^{-\rho}) - \rho}.$$

Thus, for $\alpha_0 < \alpha < \alpha_2$ the behavior of $\mu_{n,k,t}^T$ and $\bar{\mu}_{n,k,t}^T$ is governed by a power of n depending on the ratio $\alpha = k/\log n$. The maximum exponent is obtained for

$$\alpha = \frac{1}{h} = \frac{1}{p \log \frac{1}{p} + q \log \frac{1}{q}},$$

where $h = p \log \frac{1}{p} + q \log \frac{1}{q}$ denotes the entropy of the Bernoulli source. Actually, the t -th moment of the number of nodes at level $k = \frac{1}{h} \log n$ is of order $n/\sqrt{\log n}$.

3. Results on digital search trees

Let $B_{n,k}^D$ denote the external profile in digital search trees generated by a Bernoulli model with parameter $q > p$. It is clear that $B_{0,0}^D = 1$ and $B_{n,0}^D = 0 (n \geq 1)$ and the probability generating function of the external profile, $P_{n,k}^D(u) = E(u^{B_{n,k}^D})$, satisfies the following recurrence relation [8]

$$P_{n,k}^D(u) = \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} P_{j,k-1}^D(u) P_{n-1-j,k-1}^D(u) \tag{13}$$

with initial conditions $P_{0,k}^D(u) = 1$ for $k \geq 1$, $P_{0,0}^D(u) = u$, $P_{n,0}^D(u) = 1$ for $n \geq 1$.

The exponential generating function

$$N_k(x, u) = \sum_{n \geq 0} P_{n,k}^D(u) \frac{x^n}{n!}$$

fulfills the following functional recurrence

$$\frac{\partial}{\partial x} N_k(x, u) = N_{k-1}(px, u) N_{k-1}(qx, u), \quad k \geq 1 \tag{14}$$

with initial conditions $N_k(0, u) = 1(k \geq 1)$ and $N_0(x, u) = 1 + u(e^x - 1)$. Let $\mu_{n,k,t}^D = E([B_{n,k}^D]^t)$ be the t -th moment of the external profile. By taking the derivative of order t with respect to u and setting $u = 1$ we obtain for the exponential generating function

$$E_k^{(t)}(x) = \sum_{n \geq 0} \mu_{n,k,t}^D \frac{x^n}{n!}$$

the following functional recurrence

$$E_k^{(t)}(x) = e^{qx} E_{k-1}^{(t)}(px) + e^{px} E_{k-1}^{(t)}(qx) + \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \lambda(m, n) E_{k-1}^{(m)}(px) E_{k-1}^{(n)}(qx), \tag{15}$$

where $\lambda(m, n) \in \mathbb{Z}$ and $E_0^{(t)}(x) = e^x - 1$.

For example,

- a) if $t = 1$, then $\lambda(1,1) = 0$. This case analyzed by Drmota and Szpankowski [7, 8],
- b) if $t = 2$, then $\lambda(1,1) = 2$. This case analyzed by Kazemi and Vahidi-Asl [13],
- c) if $t = 3$, then $\lambda(1,2) = \lambda(2,1) = 3$ and another $\lambda(m, n) = 0$ and so on.

Let $\Delta_k^{(t)}(x) = e^{-x} E_k^{(t)}(x)$ be the Poisson transform $E_k^{(t)}(x)$. It is clear that

$$(\Delta_k^{(t)}(z))' = e^{-z} E_k^{(t)}(z) - \Delta_k^{(t)}(z).$$

Thus recurrence (15) translates into

$$\Delta_k^{(t)}(x) + \Delta_k^{(t)}(x) = \Delta_{k-1}^{(t)}(px) + \Delta_{k-1}^{(t)}(qx) + \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \lambda(m, n) \Delta_{k-1}^{(m)}(px) \Delta_{k-1}^{(n)}(qx), k \geq 1 \tag{16}$$

with initial conditions $\Delta_k^{(t)}(0) = 0(k \geq 1)$ and $\Delta_0^{(t)}(x) = 1 - e^{-x}$, since $p + q = 1$.

It is easy to show that

$$\Delta_k^{(t)}(x) = \sum_{\ell_1} \sum_{\ell_2} \alpha(\ell_1, \ell_2) \exp\{-\sum_i \sum_j \gamma(i, j) p^{\ell_i} q^{\ell_j} x\}$$

with $\ell_i, \ell_j \geq 0$ and $\alpha(\ell_1, \ell_2), \gamma(i, j) \in \mathbb{Z}$. We can express

$$Mellin\{\Delta_k^{(t)}(x); s\} = \int_0^\infty \Delta_k^{(t)}(x) x^{s-1} dx = -\Gamma(s) F_k^{(t)}(s),$$

where

$$F_k^{(t)}(s) = \sum_{\ell_1} \sum_{\ell_2} \alpha(\ell_1, \ell_2) \{\sum_i \sum_j \gamma(i, j)^{-s} p^{-\ell_i s} q^{-\ell_j s}\}.$$

Thus, $F_k^{(t)}(s)$ can be considered an entire function. Then (16) translates into

$$F_{k+1}^{(t)}(s) - F_{k+1}^{(t)}(s - 1) = T(s)F_k^{(t)}(s) + H_k^{(t)}(s), k \geq 0, F_0^{(t)}(s) = 1, \tag{17}$$

where for $p < q$,

$$\begin{aligned} H_k^{(t)}(s) &= \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \frac{\lambda(m,n)}{\Gamma(s)} \int_0^\infty \Delta_k^{(m)}(px)\Delta_k^{(n)}(qx)x^{s-1}dx \\ &\leq p^{-s} \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \frac{\lambda(m,n)}{\Gamma(s)} \int_0^\infty \Delta_k^{(m)}(x)\Delta_k^{(n)}(x)x^{s-1}dx \\ &= p^{-s} \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \frac{\lambda(m,n)}{2\pi i \Gamma(s)} \int_{c-i\infty}^{c+i\infty} Mellin\{\Delta_k^{(m)}(x); y\}Mellin\{\Delta_k^{(n)}(x); s - y\}dy. \end{aligned}$$

With the same consideration of [8], $Mellin\{\Delta_k^{(t)}(x); s\} \leq C\Gamma(s)T(s)^k$ for some constant C . Thus

$$\begin{aligned} |H_k^{(t)}(s)| &\leq C' p^{-s} \sum_{m=1}^{t-1} \sum_{n=1}^{t-1} \frac{\lambda(m,n)}{2\pi i \Gamma(s)} \\ &\quad \cdot \int_{c-i\infty}^{c+i\infty} |\Gamma(y - 1)||\Gamma(s - y - 1)| \\ &\quad \times (T(\Re(y) - 1)T(\Re(s - y) - 1))^k dy \\ &\leq C(s, p)T(c - 1)^{2k}, \Re(y) = c = \Re(s - y) \\ &= C(s, p)T(\Re(s)/2 - 1)^{2k}. \end{aligned}$$

By the above discussion $H_k^{(t)}(s) = O(T(\Re(s))/2 - 1)^{2k}$. Just similar to [13], it turns out the inhomogeneous part in (17) is relatively small. This suggests that the asymptotics of the t -th moment should be of the same order of magnitude as for the expected value.

Theorem 3. Let $\mu_{n,k,t}^D$ denote the t -th moment of the external profile in digital search trees with $0 < p < q < 1$. If n and k are positive integers with $\frac{1}{\log \frac{1}{p}} + \varepsilon \leq \frac{k}{\log n} \leq \frac{1}{\log \frac{1}{q}} - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\mu_{n,k,t}^D = L(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O(k^{-1/2})\right), \tag{18}$$

where $\rho_{n,k} = \rho(k/\log n)$ and $L(\rho, x)$ is a non-zero periodic function with period 1 in x and given by

$$L(\rho, x) = \sum_{j \in \mathbb{Z}} f(\rho + it_j)\Gamma(\rho + it_j)e^{-2j\pi ix}, t_j = 2j\pi/\log(p/q).$$

Theorem 4. Let $\bar{\mu}_{n,k,t}^D$ denote the t -th moment of the internal profile in digital search trees with $0 < p < q < 1$. Let k and n be positive integers such that $k/\log n$ satisfies $(\log \frac{1}{p})^{-1} < k/\log n < (\log \frac{1}{q})^{-1}$. Then the following assertions hold:

1. If $\frac{1}{\log \frac{1}{p}} + \varepsilon \leq \frac{k}{\log n} \leq \alpha_0 - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\bar{\mu}_{n,k,t}^D = 2^k - \bar{L}(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right).$$

2. If $k = \alpha_0(\log n + \xi\sqrt{\alpha_0\beta(0)\log n})$, where $\xi = o((\log n)^{\frac{1}{6}})$, then

$$\bar{\mu}_{n,k,t}^D = 2^k \Phi(-\xi) \left(1 + \mathcal{O}\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right)$$

where Φ is the normal distribution function.

3. If $\alpha_0 + \varepsilon \leq \frac{k}{\log n} \leq \frac{1}{\log \frac{1}{q}} - \varepsilon$ (for some $\varepsilon > 0$), then uniformly

$$\bar{\mu}_{n,k,t}^D = \bar{L}(\rho_{n,k}, \log_{p/q} p^k n) \frac{T(\rho_{n,k})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right)$$

with the same function $\bar{L}(\rho, x)$ as in Theorem 3.

We state a very short outline of the proofs: First we have to obtain an explicit solution for $F_k^{(t)}(s)(\bar{F}_k^{(t)}(s))$ of (17) which is the most difficult part. Then we have to find proper asymptotics for $F_k^{(t)}(s)(\bar{F}_k^{(t)}(s))$ and to invert then Mellin transform of $\Delta_k^{(t)}(x)(\bar{\Delta}_k^{(t)}(x))$. This leads us to an infinite number of saddle points (cf. also [8]). The final step will be to show that the Poisson transform of $\Delta_k^{(t)}(n)(\bar{\Delta}_k^{(t)}(n))$ is asymptotically equal to $\mu_{n,k,t}^D(\bar{\mu}_{n,k,t}^D)$. The reader is referred to [8] and [13] for a detailed discussion of the above mentioned tools that belong to analytic combinatorics.

4. ACKNOWLEDGEMENTS

The authors thank the anonymous referees for their valuable comments.

5. References

- [1] R. Aguech, N. Lasmar and H. M. Mahmoud, Distribution of inter-node distances in digital trees, International Conference on Analysis of Algorithms, DMTCS proc. AD, pp-1-10 (2005).
- [2] R. de la Briandais, File searching using variable length keys, Proceedings of the AFIPS Spring Joint Computer Conference. AFIPS Press, Reston, Va., pp-295-298 (1959).
- [3] J. E. Coffman and T. Eve, File structures using hashing functions, Communications of the ACM, 13, pp-427-432 (1970).
- [4] L. Devroye, Universal asymptotics for random tries and PATRICIA trees, Algorithmica, 42, pp-11-29 (2005).

- [5] L. Devroye and H. K. Hwang, Width and mode of the profile for some random trees of logarithmic height, *Ann. App. Probab.*, 16, pp-886-918 **(2006)**.
- [6] M. Drmota, *Random Trees, An Interplay Between Combinatorics and Probability*. Springer, Wien-New York, **(2009)**.
- [7] M. Drmota and W. Szpankowski, The expected profile of digital search trees, *J. Comb. Theory. A*, 118, pp-1939-1965 **(2011)**.
- [8] M. Drmota and W. Szpankowski, Un(expected) behavior of digital search tree profile, *SIAM*, pp-130-138 **(2009)**.
- [9] R. Fagin, J. Nievergelt, N. Pippenger and H. Strong, Extendible hashing fast access method for dynamic files, *ACM Transactions on Database Systems*, 4, pp-315-344 **(1979)**.
- [10] E. Fredkin, Trie memory, *Communications of the ACM*, 3, pp- 490-499 **(1960)**.
- [11] Flajolet, P. and Sedgewick, R, *Analytic combinatorics*. Cambridge University Press, Cambridge, **(2008)**.
- [12] Gusfield, D. *Algorithms on Strings, trees, and sequences*. Cambridge University Press, Cambridge, **(1997)**.
- [13] R. Kazemi and M. Q. Vahidi-Asl, The variance of the profile in digital search trees, *Discrete. Math. Theor. C*, 13:3, pp-21-38 **(2011)**.
- [14] D. E. Knuth, *The Art of Computer Programming, Volume III: Sorting and Searching*, 2nd edition, Addison Wesley, Reading, MA, **(1998)**.
- [15] G. Louchard and W. Szpankowski, Average profile and limiting distribution for a phrase Size in the Lempel-Ziv parsing algorithm, *IEEE Trans. Inform. Theory*, 41, pp-478-488 **(1995)**.
- [16] H. Mahmoud, *Evolution of random search trees*, John Wiley & Sons Inc., New York, **(1992)**.
- [17] P. Mathys and P. Flajolet, Q-ary collision resolution algorithms in random-access systems with free and blocked channel access, *IEEE Trans. Inform. Theory*, 31, pp-217-243 **(1985)**.
- [18] G. Park, H. K. Hwang, P. Nicodeme and W. Szpankowski, Profile of tries, *SIAM Journal on Computing*, 38, pp-1821-1880 **(2009)**.
- [19] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, John Wiley, New York, **(2001)**.
- [20] J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory*, 24(5), pp-530-536 **(1978)**.