

On a Recurrence Arising in Graph Compression

November 12, 2011

Yongwook Choi

J. Craig Venter Institute
Rockville, MD 20850
U.S.A.
ychoi@jcvi.org

Charles Knessl*

Dept. Math. Stat. & Comp. Sci.
University of Illinois at Chicago
Chicago, IL 60607-7045
U.S.A.
knessl@uic.edu

Wojciech Szpankowski†

Dept. Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

In a recently proposed graphical compression algorithm [1], the following tree arose in the course of the analysis. The root contains n balls that are consequently distributed between two subtrees according to a simple rule: In each step, all balls independently move down to the left subtree (say with probability p) or the right subtree (with probability $1-p$). A new node is created as long as there is at least one ball in that node. Furthermore, a nonnegative integer d is given, and at level d or greater one ball is removed from the leftmost node before the balls move down to the next level. These steps are repeated until all balls are removed (i.e., after $n + d$ steps). Observe that when $d = \infty$ the above tree can be modeled as a *trie* that stores n independent sequences generated by a memoryless source with parameter p . Therefore, we coin the name (n, d) -tries for the tree just described, and to which we often refer simply as d -tries. Parameters of such a tree (e.g., path length, depth, size) are determined by an interesting two-dimensional recurrence (in terms of n and d) that – to the best of our knowledge – was not analyzed before. We study it, and show how much parameters of such a d -trie differ from the corresponding parameters of regular tries. We use methods of analytic algorithmics, from Mellin transforms to analytic poissonization.

1 Introduction

In [1] an algorithm was described to compress the *structure* of a graph. The main idea behind the algorithm is quite simple: First, a vertex of a graph, say v_1 , is selected and the number of neighbors of v_1 is stored in a binary string. Then the remaining $n - 1$ vertices are partitioned into two sets: the neighbors of v_1 and the non-neighbors of v_1 . This process continues by selecting randomly a vertex, say v_2 , from the neighbors of v_1 and storing two numbers: the number of neighbors of v_2 among each of the above two sets. Then the remaining $n - 2$ vertices are partitioned into four sets: the neighbors of both v_1 and v_2 , the neighbors of v_1 that are non-neighbors of v_2 , the non-neighbors of v_1 that are neighbors of v_2 , and the non-neighbors of both v_1 and v_2 . This procedure continues until all vertices are processed. In the Erdős-Rényi model, a random graph has any pair of vertices connected by an edge with probability p . It is proved in [1] that for large n our algorithm optimally compresses any graph generated by the Erdős-Rényi model (and, in fact, it works well in practice even for graphs not generated by the Erdős-Rényi model). To establish this asymptotic optimality result, an interesting tree was used in the construction, that we describe next.

*This author was supported by NSF Grant DMS-0800568 and NSA Grant H98230-11-1-0184.

†This work was supported in part by the NSF Science and Technology Center for Science of Information Grant CCF-0939370, NSF Grant CCF-0830140, AFOSR Grant FA8655-11-1-3076, and NSA Grant H98230-11-1-0141.

path length. Recurrence (2) is equivalent to the following boundary condition

$$a(n, 1) = a(n + 1, 0).$$

For $d = \infty$ recurrence (1) becomes a traditional recurrence arising in the analysis of tries [19] whose solutions (exact and asymptotic) are well known. Thus, it is natural to study the difference $\tilde{a}(n, d) := a(n, d) - a(n, \infty)$, and that is our objective. In passing, we should point out that recurrence (2) resembles the one used to analyze another digital search tree, known as a *digital search tree*. In this paper we prove, however, that a (n, d) -trie more closely resembles a trie, rather than a digital search tree.

Our main interest lies in solving recurrence (1) for $d = O(1)$. In fact, for graph compression we only need $d = 0$, and we focus on this case. We shall show that the second term in (1) becomes exponentially small for n large and d fixed. Then we shall approximate the recurrence for the excess quantity $\tilde{a}(n, d)$ by

$$\tilde{a}(n, d) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \tilde{a}(k, d-1)$$

with an appropriate initial condition. The above we can solve asymptotically using Mellin transform technique and depoissonization. In particular, for $f(n) = n$ (that is, for the path length in a d -trie) we prove that the excess quantity $\tilde{a}(n, d)$ becomes asymptotically, as $n \rightarrow \infty$ and $d = O(1)$,

$$\frac{1}{2h \log p} \log^2 n + \frac{d}{h} \log n + \left[-\frac{1}{2h} + \frac{1}{h \log p} \left(\gamma + 1 + \frac{h_2}{2h} + \Psi(\log_p n) \right) \right] \log n$$

where $\Psi(\cdot)$ is the periodic function when $\log p / \log(1-p)$ is rational, and h is the entropy rate.

Digital trees such as tries and digital search trees have been intensively studied for the last thirty years [2, 3, 5, 7, 11, 12, 13, 16, 17, 18, 19]. However, our two-dimensional recurrence seems to be new and harder to analyze. It somewhat resembles the profile recurrences for digital trees, which were studied for tries in [15] and digital search trees in [4], and which are known to also be challenging.

The paper is organized as follows. In the Section 2 we precisely formulate our problem and analyze it for $f(n) = n$. Proofs are presented in Section 3, where we also discuss some asymptotics for $d \rightarrow \infty$.

2 Problem Statement

In this section, we first formulate some recurrences describing (n, d) -tries, then summarize our main results, discuss some extensions, and present numerical results.

2.1 Main Results

Let us consider a (n, d) -trie with n balls and parameter $d \geq 0$. First, we analyze the average path length $b(n, d)$. It satisfies the following recurrence equations

$$b(n+1, 0) = n + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [b(k, 0) + b(n-k, k)], \quad \text{for } n \geq 2, \quad (3)$$

and

$$b(n, d) = n + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [b(k, d-1) + b(n-k, k+d-1)], \quad \text{for } n \geq 2, d \geq 1. \quad (4)$$

Recurrence (3) follows from the fact that starting with $n+1$ balls in the root node, and removing one ball, we are left with n balls passing through the root node. The root contributes n since each time a ball moves down it adds 1 to the path length. Those n balls move down to the left or the right subtrees. Let us assume k balls move down to the left subtree (the other $n-k$ balls must move down to the right subtree); this occurs with probability $\binom{n}{k} p^k q^{n-k}$. At level one, one ball is removed from those k balls in the root of the left subtree. This contributes $b(k, 0)$. There will be no removal from $n-k$ balls in the right subtree until all k balls in the left subtree are removed. This contributes $b(n-k, k)$. Similarly, for $d > 0$ we arrive at recurrence (4).

Here $0 < p < 1$ and $q = 1 - p$, and we also use the boundary conditions

$$b(0, d) = b(1, d) = 0, \quad d \geq 0; \quad b(2, 0) = 0. \quad (5)$$

By setting $d = 1$ in (4) and comparing the result to (3) we can replace (3) by the simpler boundary condition

$$b(n, 1) = b(n+1, 0), \quad \text{for } n \geq 0. \quad (6)$$

We are primarily interested in estimating $b(n, 0)$ for large n .

If we let $d \rightarrow \infty$ in (4) and assume that $b(n, d)$ tends to a limit $b(n, \infty)$, then (4) becomes

$$b(n, \infty) = n + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [b(k, \infty) + b(n-k, \infty)], \quad (7)$$

with $b(0, \infty) = b(1, \infty) = 0$. This is the same as the recurrence for the mean path length in a trie, which was analyzed, for example, in [12, 19]. One form of the solution is given by the alternating sum

$$b(n, \infty) = \sum_{\ell=2}^n (-1)^\ell \binom{n}{\ell} \frac{\ell}{1 - p^\ell - q^\ell}, \quad (8)$$

and an alternate form is given by the integral [19]

$$b(n, \infty) = \frac{n!}{(2\pi i)^2} \oint \left[\int_{Br} z^{-s} \frac{\Gamma(s+1)}{1 - p^{-s} - q^{-s}} ds \right] \frac{e^z}{z^{n+1}} dz, \quad (9)$$

where $\Gamma(\cdot)$ is the Gamma function, Br is a vertical Bromwich contour on which $-2 < \Re(s) < -1$ and the z -integral is over a small loop about $z = 0$.

The asymptotic expansion of (9) as $n \rightarrow \infty$ may be obtained by a combination of singularity analysis and depoissonization arguments (see [7, 8, 19]) and we obtain

$$b(n, \infty) = \frac{1}{h} n \log n + \frac{1}{h} \left[\gamma + \frac{h_2}{2h} + \Phi(\log_p n) \right] n + o(n), \quad (10)$$

where $h = -p \log p - q \log q$, $h_2 = p \log^2 p + q \log^2 q$, γ is the Euler constant, and $\Phi(x)$ is the periodic function

$$\Phi(x) = \sum_{k=-\infty, k \neq 0}^{\infty} \Gamma \left(-\frac{2k\pi i r}{\log p} \right) e^{2k\pi r i x}, \quad (11)$$

provided that $\log p / \log q = r/s$ is rational, with r and s being integers with $\gcd(r, s) = 1$. If $\log p / \log q$ is irrational, then the term with Φ is absent from the $O(n)$ term of (10). We shall later use (10) to analyze the behavior of $b(n, d)$ for $n \rightarrow \infty$ and a fixed d .

Next we set

$$b(n, d) = b(n, \infty) + \tilde{b}(n, d) \quad (12)$$

so that $\tilde{b}(n, d) = b(n, d) - b(n, \infty)$ measures how the path lengths in the d -trie differs from those in a trie. From (4) and (7), we then obtain

$$\tilde{b}(n, d) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [\tilde{b}(k, d-1) + \tilde{b}(n-k, k+d-1)], \quad \text{for } n \geq 2, d \geq 1, \quad (13)$$

which unlike (4) is a homogeneous recurrence. Then from (6) and (12) we have the boundary condition

$$\tilde{b}(n+1, 0) - \tilde{b}(n, 1) = b(n, \infty) - b(n+1, \infty). \quad (14)$$

From (5) and (7) we also have $\tilde{b}(0, d) = \tilde{b}(1, d) = 0$ for $d \geq 0$, and $\tilde{b}(2, 0) = 0$.

We further define $b_*(n, d)$ to be the solution of

$$b_*(n, d) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} b_*(k, d-1), \quad \text{for } n \geq 2, d \geq 1, \quad (15)$$

and

$$b_*(n+1, 0) - b_*(n, 1) = b(n, \infty) - b(n+1, \infty). \quad (16)$$

Note that (15) differs from (13) in that the former neglects the term involving $\tilde{b}(n-k, k+d-1)$. We will show that this term in (13) is asymptotically negligible for $n \rightarrow \infty$ with $d = O(1)$, so that $\tilde{b}(n, d) \sim b_*(n, d)$. The recurrence (15) is much easier to solve by transform methods [7, 19] than is (13).

We summarize our main result below. In Section 3 we establish Theorem 1 along with some other exact and asymptotic results for (3)-(6) and (13)-(16).

Theorem 1 *For $n \rightarrow \infty$ and $d = O(1)$ we have $\tilde{b}(n, d) = O(\log^2 n)$. More precisely*

$$\tilde{b}(n, d) = \frac{1}{2h \log p} \log^2 n + \frac{d}{h} \log n + \left[-\frac{1}{2h} + \frac{1}{h \log p} \left(\gamma + 1 + \frac{h_2}{2h} + \Psi(\log_p n) \right) \right] \log n + O(1), \quad (17)$$

where $\Psi(\cdot)$ is the periodic function

$$\Psi(x) = \sum_{k=-\infty, k \neq 0}^{\infty} \left[1 + \frac{2k\pi ir}{\log p} \right] \Gamma \left(-\frac{2k\pi ir}{\log p} \right) e^{2k\pi ir x} \quad (18)$$

and $\log p / \log q = r/t$ is rational, as in (11). If $\log p / \log q$ is irrational, the term involving Ψ in (17) is absent.

We see that $b(n, d) - b(n, \infty) = O(\log^2 n)$, which shows that the (n, d) -tries studied in [1] are in some sense more similar to tries than to digital search trees (DST). In [1], it was shown that $b(n, 0)$ was bounded above by average path lengths in tries and below by average path lengths in DST's. It was also conjectured that $b(n, d) - b(n, \infty)$ is $O(n)$, but our result shows that this difference is in fact much smaller.

2.2 Related Recurrence Equations

The method presented in the next section, allow us to analyze a class of recurrences of the type (3) with inhomogeneous terms other than n . For example, suppose we define $a(n, d)$ by

$$a(n, d) = f(n) + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [a(k, d-1) + a(n-k, k+d-1)] \quad (19)$$

where $f(n)$ is a given sequence that grows algebraically or logarithmically for $n \rightarrow \infty$. The boundary condition is again of the type (3), or equivalently,

$$a(n, 1) = a(n+1, 0), \quad (20)$$

and we have $a(0, d) = a(1, d) = 0$. Also, let $a(n, \infty)$ satisfy (19) with the second argument of $a(\cdot, \cdot)$ replaced by infinity. This recurrence can be solved by generating functions and Mellin transforms, and we can then establish that $a(n, d) - a(n, \infty) \equiv \tilde{a}(n, d)$, will satisfy

$$\tilde{a}(n, d) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} [\tilde{a}(k, d-1) + \tilde{a}(n-k, k+d-1)] \quad (21)$$

and

$$\tilde{a}(n+1, 0) - \tilde{a}(n, 1) = a(n, \infty) - a(n+1, \infty). \quad (22)$$

The asymptotic behavior of $\tilde{a}(n, d)$ for $d = O(1)$ and $n \rightarrow \infty$ can be obtained in a manner completely analogous to the case $f(n) = n$, discussed in the next section.

For example, the case

$$f(n) = \lceil \log(n+1) \rceil$$

arose in analyzing the compression algorithm in [1]. In [1] it was shown that $a(n, \infty)$ has the asymptotic form

$$a(n, \infty) = \frac{n}{h} A_*(-1) + o(n), \quad n \rightarrow \infty \quad (23)$$

where

$$A_*(-1) = \sum_{\ell=2}^{\infty} \frac{\lceil \log(\ell+1) \rceil}{\ell(\ell-1)}$$

if $\log p / \log q$ is irrational. If $\log p / \log q = r/s$ is rational, the constant $A_*(-1)$ in (23) must be replaced by the oscillatory function

$$A_*(-1) + \sum_{k=-\infty, k \neq 0}^{\infty} A_* \left(-1 + \frac{2k\pi ir}{\log p} \right) e^{2k\pi ir \log_p n} \quad (24)$$

where

$$A_*(s) = \sum_{n \geq 2} \frac{\lceil \log(n+1) \rceil}{n!} \Gamma(n+s).$$

By analyzing (21) and (22) for $n \rightarrow \infty$ we can show that the difference $a(n, d) - a(n, \infty)$ is $O(\log n)$, and more precisely

$$\tilde{a}(n, 0) = a(n, 0) - a(n, \infty) \sim \frac{A_*(-1)}{h \log p} \log n.$$

Again if $\log p / \log q$ is rational we must replace $A_*(-1)$ by the Fourier series in (24).

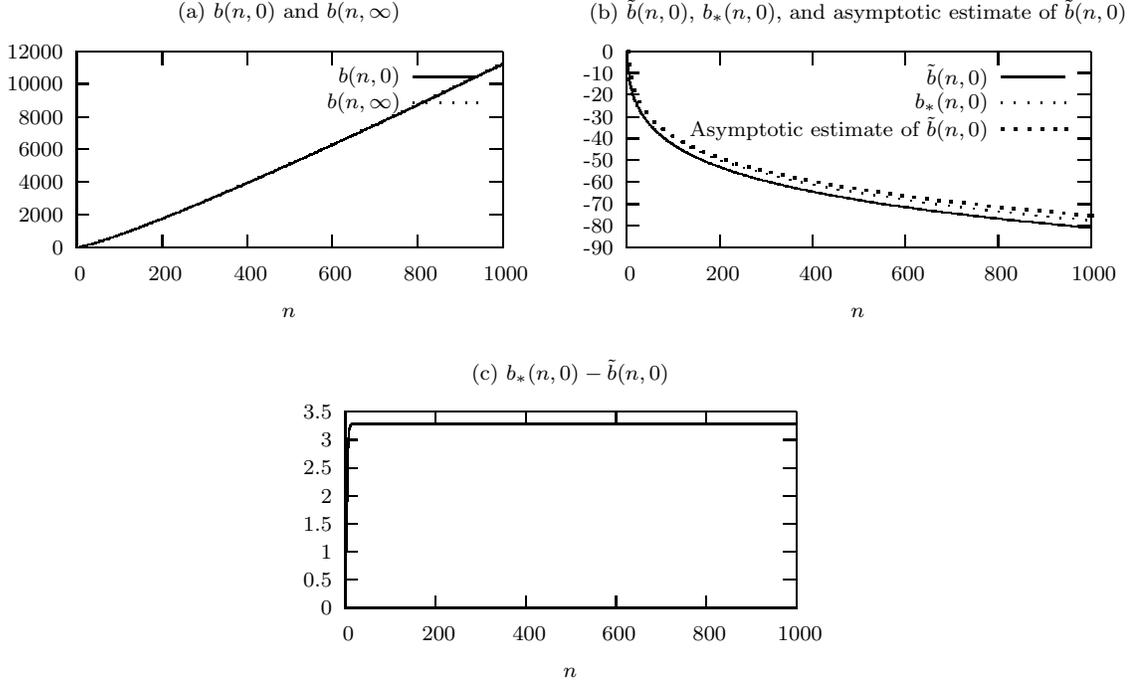


Figure 2: Numerical values with $p = 0.5$.

2.3 Numerical Data

To confirm our results, we numerically computed some of the quantities discussed above for n up to 1000, with $p = 0.5$. In Figure 2(a), we plot the values of $b(n, 0)$ and $b(n, \infty)$. In Figure 2(b), we plot the values of $\tilde{b}(n, 0)$ (defined in (12)), $b_*(n, 0)$ (defined by (15) and (16)), and our asymptotic estimate of $\tilde{b}(n, 0)$ shown in Theorem 1. We computed this asymptotic estimate up to the $\log n$ term without the periodic function $\Psi(\cdot)$, that is,

$$\frac{1}{2h \log p} \log^2 n + \frac{d}{h} \log n + \left[-\frac{1}{2h} + \frac{1}{h \log p} \left(\gamma + 1 + \frac{h_2}{2h} \right) \right] \log n.$$

Finally, we plot $b_*(n, 0) - \tilde{b}(n, 0)$ in Figure 2(c), which confirms that $\tilde{b}(n, 0) \sim b_*(n, 0)$, and suggests that the difference is $O(1)$ for $n \rightarrow \infty$.

3 Analysis

We first discuss some exact solutions of recurrence (4) for small values of n and arbitrary d , to prove our Theorem 1, and finally provide solutions of (4) for other ranges of (n, d) , where $d \rightarrow \infty$.

3.1 Some Exact Solutions

We first consider (4) for small values of n and arbitrary d . Using (5) we rewrite (4) as

$$b(n, d) = n + \sum_{k=2}^n \binom{n}{k} p^k q^{n-k} b(k, d-1) + \sum_{k=0}^{n-2} \binom{n}{k} p^k q^{n-k} b(n-k, k+d-1). \quad (25)$$

When $n = 2$, (25) yields $b(2, d) = 2 + (p^2 + q^2)b(2, d - 1)$ and since $b(2, 0) = 0$ we have

$$b(2, d) = \frac{1}{pq} + \left(2 - \frac{1}{pq}\right) (p^2 + q^2)^{d-1} \quad \text{for } d \geq 1. \quad (26)$$

Note that $b(2, \infty) = (pq)^{-1}$ by (8). Setting $n = 3$ in (25) then yields

$$b(3, d) = 3 + (p^3 + q^3)b(3, d - 1) + 3p^2qb(2, d - 1) + 3pq^2b(2, d). \quad (27)$$

Using (26) to evaluate the right side of (27) and noting that $b(3, 0) = b(2, 1) = 2$ by (6), we solve the difference equation (27) with respect to d to ultimately obtain

$$b(3, d) = \frac{2}{pq} + \frac{3}{pq}(2pq^2 - 1)(p^2 + q^2)^d + \left(2 + \frac{1}{pq} - 6q\right) (p^3 + q^3)^d \quad \text{for } d \geq 0. \quad (28)$$

We can then continue solving $b(n, d)$ for increasing n , and it is clear that $b(n, d)$ will have the form

$$b(n, d) = b(n, \infty) + \sum_{J=2}^n (p^J + q^J)^d B(n, J), \quad (29)$$

where $b(n, \infty)$ is the trie path length in (8) and (9). It follows that $b(n, d) - b(n, \infty) = O\left[(p^2 + q^2)^d\right]$ for n fixed and $d \rightarrow \infty$. We can characterize the double sequence $B(n, J)$ by using (29) in (25) and equating coefficients of $(p^J + q^J)^d$. For $J \geq 2$ this leads to

$$B(n, J) = \frac{1}{p^J + q^J} \sum_{k=J}^n \binom{n}{k} \left(p^k q^{n-k} + q^k [p(p^J + q^J)]^{n-k}\right) B(k, J). \quad (30)$$

From (26) we have

$$B(2, 2) = \left(2 - \frac{1}{pq}\right) \frac{1}{p^2 + q^2} = -\frac{1}{pq}$$

and from (28)

$$B(3, 2) = \frac{3}{pq}(2pq^2 - 1), \quad B(3, 3) = \frac{1}{pq}(2pq + 1 - 6pq^2).$$

From (6) and (29) we find that

$$b(n + 1, \infty) + \sum_{J=2}^{n+1} B(n + 1, J) = b(n, \infty) + \sum_{J=2}^n B(n, J)(p^J + q^J). \quad (31)$$

For example, to compute $B(4, J)$ we would first set $n = 3$ in (31) and since we already know $B(3, 3)$ and $B(3, 2)$ we have an expression for the sum $B(4, 2) + B(4, 3) + B(4, 4)$. Then setting $J = 4$ and $n = 4$ in (30) leads simply to $B(4, 4) = B(4, 4)$, which is automatically satisfied. Setting $J = 2$ and $n = 4$ expresses $B(4, 2)$ in terms of $B(2, 2)$ and $B(3, 2)$, which we computed already, and setting $J = 3$ and $n = 4$ expresses $B(4, 3)$ in terms of $B(3, 3)$. Then $B(4, 4)$ follows from (31) with $n = 3$. We can thus get the $B(n, J)$ recursively, but it does not seem possible to obtain an explicit analytic expression for this double sequence.

We can transform (30) into another equation by introducing the generating function

$$\mathcal{F}_J(z) = \sum_{n=0}^{\infty} B(n, J) \frac{z^n}{n!} = \sum_{n=J}^{\infty} B(n, J) \frac{z^n}{n!}. \quad (32)$$

Using (32) in (30) leads to the functional equation

$$\mathcal{F}_J(z) = \frac{1}{p^J + q^J} \left(\mathcal{F}_J(pz)e^{qz} + \mathcal{F}_J(qz)e^{p(p^J+q^J)z} \right) \quad \text{for } J \geq 2. \quad (33)$$

Then if $\mathcal{F}_J(z) = e^z \mathcal{G}_J(z)$ we obtain

$$(p^J + q^J) \mathcal{G}_J(z) = \mathcal{G}_J(pz) + \mathcal{G}_J(qz)e^{p(p^J+q^J-1)z}. \quad (34)$$

Again this appears difficult to solve explicitly (however, see [10]).

We can take the analysis somewhat further in the *symmetric case* where $p = q = 1/2$, as then (34) simplifies to

$$\mathcal{G}_J(z)2^{1-J} = \mathcal{G}_J\left(\frac{z}{2}\right) \left(1 + \exp\left[(2^{1-J} - 1)\frac{z}{2} \right] \right). \quad (35)$$

Then setting $\mathcal{G}_J(z) = z^J \mathcal{H}_J(z)$ and noting that by (32) $\mathcal{F}_J(z)$ and $\mathcal{G}_J(z)$ are $O(z^J)$ as $z \rightarrow 0$, we see that $\mathcal{H}_J(0)$ will be non-zero and finite. Thus (35) becomes

$$\mathcal{H}_J(z) = \mathcal{H}_J\left(\frac{z}{2}\right) \frac{1}{2} \left(1 + \exp\left[(2^{1-J} - 1)\frac{z}{2} \right] \right) \quad (36)$$

which can be solved by iteration to get the infinite product

$$\mathcal{H}_J(z) = \mathcal{H}_J(0) \prod_{L=0}^{\infty} \left(\frac{1 + e^{(2^{1-J}-1)2^{-L-1}z}}{2} \right). \quad (37)$$

Then inverting (32) we obtain for $B(n, J)$ the representation

$$B(n, J) = \mathcal{H}_J(0) \frac{n!}{2\pi i} \oint \frac{e^z}{z^{n+1-J}} \prod_{L=0}^{\infty} \left(\frac{1 + e^{(2^{1-J}-1)2^{-L-1}z}}{2} \right) dz. \quad (38)$$

Thus the double sequence $B(n, J)$ is known up to the single sequence $\mathcal{H}_J(0) = B(J, J)/J!$. To determine $B(J, J)$ we must still use (31). Thus putting (37) in (31) will lead to a single variable recurrence for $B(J, J)$, and we note that in the symmetric case $B(2, 2) = -4$ and $B(3, 3) = 3$.

Next we return to *general* p, q and estimate $B(n, 2)$ in (30) for $n \rightarrow \infty$. Let us set $C(n) = B(n, 2)$ and we recall that, by (29),

$$\tilde{b}(n, d) \sim C(n)(p^2 + q^2)^d; \quad d \rightarrow \infty, \quad n = O(1). \quad (39)$$

While we mainly want to estimate $\tilde{b}(n, d)$ for $n \rightarrow \infty$ and $d = O(1)$, it is useful to try to understand the full asymptotic structure of $\tilde{b}(n, d)$, for n and/or d large.

We thus examine how (39) behaves when n also becomes large. Setting $J = 2$ in (30) leads to

$$(p^2 + q^2)C(n) = \sum_{k=2}^n \binom{n}{k} p^k q^{n-k} C(k) + \sum_{k=0}^{n-2} \binom{n}{k} p^k q^{n-k} (p^2 + q^2)^k C(n-k) \quad (40)$$

for $n \geq 3$ with $C(2) = (2 - p^{-1}q^{-1})/(p^2 + q^2) = -p^{-1}q^{-1}$.

We argue intuitively that $C(n)$ will behave algebraically for $n \rightarrow \infty$ (we shall prove this fact shortly). Then we use the fact that the “kernel” in (40) behaves

$$\binom{n}{k} p^k q^{n-k} \rightarrow \delta(k - np), \quad n \rightarrow \infty$$

where $\delta(\cdot)$ is the delta function. Then for algebraically or logarithmically varying smooth $f(k)$ (for $k \rightarrow \infty$) we have (see [6, 9] for rigorous proofs)

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} f(k) = f(np) + O(nf''(np)), \quad n \rightarrow \infty. \quad (41)$$

Then the term involving $(p^2 + q^2)^k C(n - k)$ will lead to terms that are exponentially smaller than those arising from $C(k)$, and (40) may be replaced by the asymptotic relation

$$C(n)(p^2 + q^2) \sim C(np), \quad n \rightarrow \infty. \quad (42)$$

A general solution to (42) has the form

$$C(n) = n^\nu \bar{C}(n) \quad (43)$$

where $\bar{C}(np) = \bar{C}(n)$ and $p^\nu = p^2 + q^2$ so that

$$\nu = \frac{\log(p^2 + q^2)}{\log p} > 0. \quad (44)$$

Thus $\bar{C}(\cdot)$ is a periodic function of $\log_p n$ of period 1, which we can write as the Fourier series

$$\bar{C}(n) = c^{(0)}(p) + \sum_{\ell=-\infty, \ell \neq 0}^{\infty} c^{(\ell)}(p) e^{2\pi i \ell \log_p n}. \quad (45)$$

It again appears difficult to identify explicitly the Fourier coefficients $c^{(\ell)}(p)$, but we can do this in the symmetric case $p = q = 1/2$. Then we set $\sum_{n=0}^{\infty} C(n) z^n / n! = \mathcal{F}_2(z)$ as in (32) and from (38) obtain

$$C(n) = \frac{-2n!}{2\pi i} \oint \frac{e^z}{z^{n-1}} \prod_{\ell=1}^{\infty} \left(\frac{1 + e^{-z2^{-\ell-1}}}{2} \right) dz. \quad (46)$$

To obtain the large n behavior of the integral in (46) we first expand the integral for $z \rightarrow \infty$ and apply a depoissonization argument. Setting $\ell = \log_2 z + J$ we have $2^\ell = 2^J z$ and

$$\begin{aligned} \prod_{\ell=1}^{\infty} \left(\frac{1 + e^{-z2^{-\ell-1}}}{2} \right) &= \exp \left[\sum_{\ell=1}^{\infty} \log \left(\frac{1 + e^{-z2^{-\ell-1}}}{2} \right) \right] \\ &= \exp \left[\sum_{J=1-\log_2 z}^{\infty} \log \left(\frac{1 + e^{-2^{-J-1}}}{2} \right) \right] \\ &= \exp \left[\sum_{J=0}^{\infty} \log \left(\frac{1 + e^{-2^{-J-1}}}{2} \right) + \sum_{J=1}^{\log_2 z - 1} \log \left(\frac{1 + e^{-2^{-J-1}}}{2} \right) \right] \\ &\sim \exp \left[\log \left(\frac{1}{2} \right) (\log_2 z - 1) + \sum_{J=0}^{\infty} \log \left(\frac{1 + e^{-2^{-J-1}}}{2} \right) + \sum_{J=1}^{\infty} (1 + e^{-2^{-J-1}}) \right] \\ &= \frac{2}{z} K_* \end{aligned}$$

where

$$K_* = \prod_{J=0}^{\infty} \left(\frac{1 + e^{-2^{-J-1}}}{2} \right) \prod_{J=1}^{\infty} \exp \left(1 + e^{-2^{J-1}} \right) = 1.$$

Thus $C(n) \sim -4n!/(n-1)! = -4n$ as $n \rightarrow \infty$. This shows that $c^{(0)}(1/2) = -4$ and a more careful calculation can be used to identify the other Fourier coefficients $c^{(\ell)}(1/2)$ in (45) (then we would set $\ell = \lfloor \log_2 z \rfloor + J = \log_2 z + J - \{\log_2 z\}$ so that $2^\ell = 2^J z 2^{-\{\log_2 z\}}$). We omit the details.

In Table 1 we consider various values of p and estimate $\bar{C}(n) \approx c^{(0)}(p)$ numerically, by computing $C(n)n^{-\nu}$ from (40), for large n . This shows that as a function of p , $|c^{(0)}(p)|$ is minimal when p is between 0.6 and 0.7, and becomes large as either $p \rightarrow 0$ or $p \rightarrow 1$. For $p \rightarrow 0$ the oscillatory terms in (45) become more numerically significant. Table 1 indicates this when $p = 0.1$, by giving a range of values of $C(n)n^{-\nu}$.

Table 1: Values of the zeroth Fourier coefficient.

p	$C(n)n^{-\nu} _{n \rightarrow \infty} \approx c^{(0)}(p)$
0.5	-4
0.4	-5.664
0.3	-9.728
0.25	-14.03
0.2	-22.5
0.1	-98 to -105
0.6	-3.331
0.7	-3.276
0.75	-3.479
0.8	-3.903
0.9	-6.423

To justify the approximation in (42) we first inductively show that for all n

$$C(n) \leq An^{\nu+\epsilon} \tag{47}$$

for all $\epsilon > 0$ and $A > 0$. By isolating the terms in the sums in (40) with $k = n$ and $k = 0$ we obtain, for $n > 2$,

$$C(n) = \frac{1}{p^2 + q^2 - p^n - q^n} \left[\sum_{k=2}^{n-1} \binom{n}{k} p^k q^{n-k} C(k) + \sum_{k=1}^{n-2} \binom{n}{k} p^k q^{n-k} (p^2 + q^2)^k C(n-k) \right]. \tag{48}$$

Assuming inductively that (47) holds for $C(k)$ for $k = 1, 2, \dots, n-1$ we then have

$$\begin{aligned} \sum_{k=2}^{n-1} \binom{n}{k} p^k q^{n-k} C(k) &\leq \sum_{k=2}^{n-1} \binom{n}{k} p^k q^{n-k} A k^{\nu+\epsilon} \\ &\leq A(np)^{\nu+\epsilon}. \end{aligned}$$

Using a similar estimate for the second sum in (48) we are led to

$$\begin{aligned} C(n) &\leq \frac{A}{p^2 + q^2 - p^n - q^n} \left[(np)^{\nu+\epsilon} + n^{\nu+\epsilon} (p(p^2 + q^2) + q)^n \right] \\ &= An^{\nu+\epsilon} \left[\frac{p^2 + q^2}{p^2 + q^2 - p^n - q^n} p^\epsilon + \frac{(p(p^2 + q^2) + q)^n}{p^2 + q^2 - p^n - q^n} \right], \end{aligned} \quad (49)$$

as $C(n-k) \leq A(n-k)^{\nu+\epsilon} \leq An^{\nu+\epsilon}$ and $p^\nu = p^2 + q^2$. Since $p(p^2 + q^2) + q < p + q = 1$, the second term in (49) is asymptotically negligible for n large and (47) follows by induction.

We have thus obtained some exact expressions for $b(n, d)$ for small values of n , a general asymptotic result for $d \rightarrow \infty$ with $n = O(1)$, and then examined how this result behaves when n also becomes large. However, this cannot be used to infer the behavior of $b(n, d)$ for $n \rightarrow \infty$ with $d = O(1)$, which we examine next.

3.2 Main Asymptotic Result for $b(n, d)$

We first give an intuitive derivation of the asymptotics of $b(n, d)$ for fixed $d \geq 0$ and $n \rightarrow \infty$, and in particular of $b(n, 0)$. Starting from (13) we again argue that the second sum is negligible for $n \rightarrow \infty$ and that the first is asymptotic to $\tilde{b}(np, d-1)$ so that (13) becomes

$$\tilde{b}(n, d) \sim \tilde{b}(np, d-1), \quad n \rightarrow \infty \quad (50)$$

and, in particular,

$$\tilde{b}(n, 1) \sim \tilde{b}(np, 0), \quad n \rightarrow \infty \quad (51)$$

which when added to (14) leads to

$$\tilde{b}(n+1, 0) - \tilde{b}(np, 0) \sim b(n, \infty) - b(n+1, \infty). \quad (52)$$

The right side of (52) may be estimated from (10) or by (9). Using (9) we can show that term by term differentiating of the asymptotic series in (10) is permissible, and thus (52) becomes, for $n \rightarrow \infty$,

$$\tilde{b}(n+1, 0) - \tilde{b}(np, 0) = -\frac{1}{h} \log n - \frac{1}{h} \left(\gamma + 1 + \frac{h_2}{2h} \right) - \frac{1}{h} \psi(\log_p n) + o(1), \quad (53)$$

where $\psi(\cdot)$ is the periodic function

$$\psi(x) = \sum_{k=-\infty, k \neq 0}^{\infty} \left[1 + \frac{2k\pi ir}{\log p} \right] \Gamma \left(-\frac{2k\pi ir}{\log p} \right) e^{2k\pi irx}, \quad (54)$$

where we note that, in view of (11), $\psi(x) = \Phi(x) + (\log p)^{-1} \Phi'(x)$.

Now (53) suggests that $\tilde{b}(n, 0)$ admits an asymptotic expansion of the form

$$\tilde{b}(n, 0) = A \log^2 n + B \log n + O(1), \quad n \rightarrow \infty \quad (55)$$

and then

$$\tilde{b}(n+1, 0) - \tilde{b}(np, 0) = -2A(\log p) \log n - A \log^2 p - B \log p + o(1). \quad (56)$$

Comparing (53) to (56) we conclude that $A = (2h \log p)^{-1}$ and then

$$B = -\frac{1}{2h} + \frac{1}{h \log p} \left[\gamma + 1 + \frac{h_2}{2h} + \psi(\log_p n) \right]. \quad (57)$$

We have thus formally derived the result in Theorem 1 for $\tilde{b}(n, 0)$. For any fixed $d > 0$ we can extend this argument by asymptotically solving (50) by an expansion of the form

$$\tilde{b}(n, d) = A(d) \log^2 n + B(d) \log n + O(1) \quad (58)$$

to find from (50) that $A(d) = A(d-1)$ and $B(d) = B(d-1) + 2 \log p A(d-1)$. Then using (58) in (52) or (53) we find that $A(d) = A(0) = (2h \log p)^{-1}$ and $B(d) - B(d-1) = 2 \log p A(d-1) = h^{-1}$ so that $B(d) = B(0) + h^{-1}d$, where $B(0) = B$ is as in (57).

We proceed to provide a rigorous derivation of the theorem. Using arguments completely analogous to (47)–(49), we can inductively establish the bound

$$\tilde{b}(n, d) \leq A_0 n^{\nu+\epsilon} (p^2 + q^2)^d; \quad n \geq 2, d \geq 0 \quad (59)$$

where again ν is given by (44). Using the bound in (59) we thus estimate the second sum (13) by

$$\begin{aligned} \sum_{k=0}^{n-2} \binom{n}{k} p^k q^{n-k} \tilde{b}(n-k, k+d-1) &\leq A_0 \sum_{k=0}^n (n-k)^{\nu+\epsilon} (p^2 + q^2)^{k+d-1} p^k q^{n-k} \binom{n}{k} \\ &\leq A_0 n^{\nu+\epsilon} (p^2 + q^2)^{d-1} \sum_{k=0}^n \binom{n}{k} [p(p^2 + q^2)]^k q^{n-k} \\ &= A_0 n^{\nu+\epsilon} (p^2 + q^2)^{d-1} [q + p(p^2 + q^2)]^n \end{aligned}$$

which is $o(\tilde{b}(n, d))$ (by an exponentially small factor). It follows from comparing (13) and (15) that $\tilde{b}(n, d) \sim b_*(n, d)$ for $n \rightarrow \infty$. We proceed to analyze (15), with (16), and thus re-establish Theorem 1.

Introducing the exponential generating function

$$B_d^*(z) = \sum_{n=2}^{\infty} b_*(n, d) \frac{z^n}{n!} = e^z A_d(z), \quad (60)$$

where $b_*(n, d)$ is defined from (15), we find that

$$B_d^*(z) = B_{d-1}^*(pz) e^{qz} \quad (61)$$

or, since $A_d(z) = B_d^*(z) e^{-z}$,

$$A_d(z) = A_{d-1}(pz). \quad (62)$$

This can be solved by iteration to yield

$$A_d(z) = A_0(p^d z). \quad (63)$$

Then setting

$$\mathcal{G}_*(z) = \sum_{n=2}^{\infty} b(n, \infty) \frac{z^n}{n!} \quad (64)$$

and noting that

$$\sum_{n=1}^{\infty} b_*(n+1, 0) \frac{z^n}{n!} = \frac{d}{dz} B_0^*(z), \quad (65)$$

(16) leads to

$$\frac{d}{dz} B_0^*(z) - B_1^*(z) = \mathcal{G}_*(z) - \mathcal{G}'_*(z). \quad (66)$$

If $\mathcal{G}_*(z) = e^z \tilde{\mathcal{G}}(z)$, from the integral representation in (9) we conclude that the Mellin transform of $\tilde{\mathcal{G}}(z)$ is

$$\int_0^{\infty} \tilde{\mathcal{G}}(z) z^{s-1} dz = \frac{\Gamma(s+1)}{1-p^{-s}-q^{-s}}, \quad (67)$$

Using (61), (63), and the definitions of $A_d(\cdot)$ and $\tilde{\mathcal{G}}(\cdot)$, (66) becomes

$$A'_0(z) + A_0(z) - A_0(pz) = -\tilde{\mathcal{G}}'(z). \quad (68)$$

We introduce the Mellin transform of $A_0(z)$

$$\mathcal{M}(s) = \int_0^{\infty} A_0(z) z^{s-1} dz \quad (69)$$

and use (69) to obtain the functional equation

$$-(s-1)\mathcal{M}(s-1) + (1-p^{-s})\mathcal{M}(s) = \frac{(s-1)\Gamma(s)}{1-p^{1-s}-q^{1-s}}. \quad (70)$$

Next we set

$$\mathcal{M}(s) = \Gamma(s)\mathcal{N}(s) \quad (71)$$

with which (70) becomes

$$-\mathcal{N}(s-1) + (1-p^{-s})\mathcal{N}(s) = \frac{s-1}{1-p^{1-s}-q^{1-s}}. \quad (72)$$

To solve (72) we let

$$\mathcal{N}(s) = \prod_{k=0}^{\infty} \left[\frac{1-p^{k+2}}{1-p^{k-s}} \right] \mathcal{N}_1(s) \quad (73)$$

and then (72) becomes

$$\mathcal{N}_1(s) - \mathcal{N}_1(s-1) = \prod_{k=1}^{\infty} \left[\frac{1-p^{k-s}}{1-p^{k+1}} \right] \frac{s-1}{1-p^{1-s}-q^{1-s}}. \quad (74)$$

Now, for $s \rightarrow -\infty$ the right side of (74) behaves as $(s-1) \prod_{k=1}^{\infty} (1-p^{k+1})^{-1}$, with an exponentially small error. Letting

$$\mathcal{N}_1(s) = \frac{s(s-1)}{2} \prod_{k=1}^{\infty} \left(\frac{1}{1-p^{k+1}} \right) + \mathcal{N}_2(s) \quad (75)$$

the equation for $\mathcal{N}_2(\cdot)$ becomes

$$\mathcal{N}_2(s) - \mathcal{N}_2(s-1) = \frac{s-1}{\prod_{k=1}^{\infty} (1-p^{k+1})} \left[\frac{1}{1-p^{1-s}-q^{1-s}} \prod_{k=1}^{\infty} (1-p^{k-s}) - 1 \right] \quad (76)$$

whose right hand side is, unlike that of (74), exponentially small for $s \rightarrow -\infty$. The solution to (76) is

$$\mathcal{N}_2(s) = \mathcal{N}_2(-\infty) + \sum_{i=0}^{\infty} \left[\frac{\prod_{k=1}^{\infty} (1 - p^{k-s+i})}{1 - p^{1+i-s} - q^{1+i-s}} - 1 \right] \frac{s - 1 - i}{\prod_{k=1}^{\infty} (1 - p^{k+1})}. \quad (77)$$

From (60) we see that $A_d(z) = O(z^2)$ as $z \rightarrow 0$ so that $\mathcal{M}(s)$ in (69) must be analytic at $s = -1$. From (71) we then conclude that $\mathcal{N}(-1) = 0$. From (73) we have $\mathcal{N}_1(-1) = 0$ and from (75) and (77) we thus obtain an expression for $\mathcal{N}_2(-\infty)$:

$$\mathcal{N}_2(-\infty) \prod_{k=1}^{\infty} (1 - p^{k+1}) + 1 - \sum_{i=0}^{\infty} (i+2) \left[\frac{\prod_{k=1}^{\infty} (1 - p^{k+i+1})}{1 - p^{2+i} - q^{2+i}} - 1 \right] = 0. \quad (78)$$

We have thus obtained the final expression for $\mathcal{M}(s)$ in (71) as

$$\mathcal{M}(s) = \frac{\Gamma(s)}{\prod_{L=0}^{\infty} (1 - p^{L-s})} \left(\frac{s(s-1)}{2} + \beta + \sum_{i=0}^{\infty} (s-i-1) \left[\frac{\prod_{k=1}^{\infty} (1 - p^{k-s+i})}{1 - p^{1+i-s} - q^{1+i-s}} - 1 \right] \right), \quad (79)$$

where

$$\beta = \mathcal{N}_2(-\infty) \prod_{k=1}^{\infty} (1 - p^{k+1})$$

can be computed from (78). Inverting the transforms in (60) and (69) we obtain

$$b_*(n, d) = \frac{n!}{2\pi i} \oint \frac{e^z}{z^{n+1}} \left[\frac{1}{2\pi i} \int_{Br} (p^d z)^{-s} \mathcal{M}(s) ds \right] dz. \quad (80)$$

The final step is to expand $b_*(n, d) (\sim \tilde{b}(n, d))$ for $n \rightarrow \infty$ with d fixed. The integral over z can be asymptotically evaluated by a standard depoissonization argument, which corresponds to replacing z by n in the inner s -integral. The function $\mathcal{M}(s)$ in (79) has a triple pole at $s = 0$, and there are other double poles on the imaginary s -axis if $1 - p^{1-s} - q^{1-s}$ has zeros there, which occurs only if $\log p / \log q$ is rational, say r/t where r and t are integers. First we compute the contribution from $s = 0$. Using the expansion $\Gamma(s) = [1 - \gamma s + O(s^2)]/s$ as $s \rightarrow 0$, with γ being the Euler constant, (79) becomes

$$\begin{aligned} \mathcal{M}(s) &= \frac{1}{s} [1 - \gamma s + O(s^2)] (1 - p^{-s})^{-1} \prod_{L=1}^{\infty} (1 - p^{L-s})^{-1} \\ &\times \left(\frac{s-1}{1 - p^{1-s} - q^{1-s}} \prod_{k=1}^{\infty} (1 - p^{k-s}) - (s-1) + \frac{s(s-1)}{2} + \beta \right. \\ &\left. + \sum_{i=1}^{\infty} (s-i-1) \left[\frac{\prod_{k=1}^{\infty} (1 - p^{k-s+i})}{1 - p^{1+i-s} - q^{1+i-s}} - 1 \right] \right). \end{aligned} \quad (81)$$

Now

$$1 - p^{-s} = s \log p - \frac{1}{2} s^2 (\log p)^2 + O(s^3)$$

and

$$1 - p^{1-s} - q^{1-s} = -hs - \frac{h_2}{2} s^2 + O(s^3).$$

Also, using the expression in (78) to compute $\beta + 1$ the expansion of (81) for $s \rightarrow 0$ becomes

$$\begin{aligned}\mathcal{M}(s) &= \frac{1}{s^3} \frac{1 - \gamma s}{\log p} \left[1 + \frac{s}{2} \log p + O(s^2) \right] \left\{ \frac{1 - s}{h} \left[1 - \frac{h_2}{2h} s + O(s^2) \right] + O(s^2) \right\} \\ &= \frac{1}{s^3} \frac{1}{h \log p} + \frac{1}{s^2} \left[-\frac{\gamma}{h \log p} - \frac{1}{h \log p} \left(1 + \frac{h_2}{2h} \right) + \frac{1}{2h} \right] + O\left(\frac{1}{s}\right).\end{aligned}\quad (82)$$

It follows that the integrand $p^{-ds} z^{-s} \mathcal{M}(s)$ in (80) has the residue

$$\text{Res}_{s=0} \left\{ p^{-ds} z^{-s} \mathcal{M}(s) \right\} = \frac{1}{2} \frac{\log^2 z}{h \log p} + \frac{d}{h} \log z + \log z \left[\frac{1}{\log p} \left(\frac{\gamma + 1}{h} + \frac{h_2}{2h^2} \right) - \frac{1}{2h} \right] + O(1) \quad (83)$$

where the $O(1)$ refers to terms that are $O(1)$ for $z \rightarrow \infty$, and these can be evaluated by explicitly computing the $O(s^{-1})$ term(s) in (82). Then the expansion of $\tilde{b}(n, d) \sim b_*(n, d)$ follows by setting $z = n$ in (83), and we have thus regained the formula in (17). If $\log p / \log q$ is rational we must also compute the contribution from the double poles along the imaginary axis at such points $p^{-s} = q^{-s} = 1$ and $p^{1-s} + q^{1-s} = 1$. These poles lead to the oscillatory terms in (17), as can be seen by computing their residues from (79).

We have thus established (17) rigorously, though the intuitive derivation in (50)–(58) is much simpler, and more revealing of the basic asymptotic structure of the equations (13) and (14).

3.3 Other Asymptotic Ranges

Here we briefly discuss $\tilde{b}(n, d)$ when n and d are simultaneously large, and try to identify what ranges of n and d lead to different asymptotic expansions. We recall that (39) applies for n fixed and $d \rightarrow \infty$, while (17) applies for d fixed and $n \rightarrow \infty$. We confine ourselves here to an intuitive discussion.

The form of the expansion in (39) (with $C(n)$ given by (43) and (45)) suggests that an important scale is $n, d \rightarrow \infty$ with $d - \log_{1/p}(n) = O(1)$. Note that then the algebraic growth of n^ν as $n \rightarrow \infty$ is balanced by the geometric decay of $(p^2 + q^2)^d$ in (39). We introduce the new variable ξ with

$$d = \frac{\log n}{\log(1/p)} + \xi, \quad \xi = O(1) \quad (84)$$

with

$$\tilde{b}(n, d) = \mathcal{B}(n, \xi) = \mathcal{B}(n, d - \log_{1/p}(n)), \quad (85)$$

and we note that

$$\tilde{b}(np, d - 1) = \mathcal{B}(np, \xi). \quad (86)$$

We again argue that for $n \rightarrow \infty$ the second sum in (13) is negligible and approximate (13) by

$$\tilde{b}(n, d) = \tilde{b}(np, d - 1) + O[n\tilde{b}''(np, d - 1)], \quad (87)$$

as in (41). In view of (85) and (86) a general asymptotic solution of (87) is any function that satisfies $\mathcal{B}(n, \xi) = \mathcal{B}(np, \xi)$ which we can write as a Fourier series, with

$$\mathcal{B}(n, \xi) = \mathcal{B}_0(\xi) + \sum_{\ell=-\infty, \ell \neq 0}^{\infty} e^{2\pi i \ell \log_p(n)} \mathcal{B}_\ell(\xi). \quad (88)$$

Thus (88) gives an approximation to $\tilde{b}(n, d)$ for $n, d \rightarrow \infty$ with $\xi = O(1)$, but we cannot explicitly determine the Fourier coefficients $\mathcal{B}_\ell(\xi)$, which are now functions of ξ . If we require $\mathcal{B}(n, \xi)$ to asymptotically match to (39), we would equate the large n behavior (39) to the expansion of $\mathcal{B}(n, \xi)$ for $\xi \rightarrow +\infty$, and this yields

$$\mathcal{B}_0(\xi) \sim c^{(0)} e^{\xi \log(p^2+q^2)}, \quad \xi \rightarrow +\infty, \quad (89)$$

and a similar matching condition can be obtained for $\mathcal{B}_\ell(\xi)$ for $\ell \neq 0$, by comparing (88) and (43) with (45). Thus (89) shows that $\mathcal{B}_0(\xi)$ will decay exponentially for $\xi \rightarrow +\infty$.

Next we examine $\tilde{b}(n, d)$ for $d = O(\log n)$ by defining ω from

$$d = \omega \log n, \quad 0 < \omega < \frac{1}{\log(1/p)} \quad (90)$$

and then set

$$\tilde{b}(n, d) = \log^2(n) \mathcal{F}(\omega). \quad (91)$$

Then we approximate (13) again by $\tilde{b}(n, d) \sim \tilde{b}(np, d-1)$ which in view of (91) becomes

$$\log^2(n) \mathcal{F}(\omega) \sim \log^2(np) \mathcal{F}\left(\frac{d-1}{\log(np)}\right) \sim (\log n + \log p)^2 \mathcal{F}\left(\omega - \frac{1}{\log n} - \frac{\omega \log p}{\log n} + O(\log^{-2} n)\right). \quad (92)$$

From (92) we obtain the following limiting ODE:

$$0 = -\mathcal{F}'(\omega)(1 + \omega \log p) + 2 \log p \mathcal{F}(\omega). \quad (93)$$

The solution to (93) is

$$\mathcal{F}(\omega) = (1 + \omega \log p)^2 \mathcal{F}_* \quad (94)$$

where \mathcal{F}_* is a constant. For $\omega \rightarrow 0$, the expansion in (91) behaves as $\mathcal{F}_* \log^2(n)$ and if we match the ω -scale result to the $d = O(1)$ result in Theorem 1, we conclude that

$$\mathcal{F}_* = \frac{1}{2h}.$$

Finally, by asymptotically matching (91) as $\omega \rightarrow [\log(1/p)]^{-1}$ to the approximation in (85) and (88), for $\xi \rightarrow -\infty$, we conclude that

$$\mathcal{B}_0(\xi) \sim \frac{1}{2h} \log^2(p) \xi^2, \quad \xi \rightarrow -\infty. \quad (95)$$

Note that ξ and ω are related by

$$1 + \omega \log p = \frac{\log p}{\log n} \xi \quad (96)$$

so that when $0 < \omega < [\log(1/p)]^{-1}$ we have $\xi < 0$.

To summarize the formal results in this subsection, our analysis suggests that the asymptotics of $\tilde{b}(n, d)$ are different for the three cases:

(i) $n = O(1), d \rightarrow \infty$ (where (39) holds),

(ii) $\xi = d - \log_{1/p}(n) = O(1)$ where (88) holds, and

(iii) $d = O(\log n)$ where $\tilde{b}(n, d) \sim (2h)^{-1}(1 + \omega \log p)^2 \log^2 n$ with $d = \omega \log n$ and $0 < \omega < [\log(1/p)]^{-1}$.

The result in Theorem 1 appears to be a limiting case of the $d = O(\log n)$ expansion, when it is expanded for $\omega \rightarrow 0$. However, Theorem 1 also gives the second term ($O(\log n)$) in the asymptotic series for $d = O(1)$.

We have only given the asymptotic behaviors of $\mathcal{B}_0(\xi)$ as $\xi \rightarrow \pm\infty$ (cf. (89) and (95)). To get a more explicit expression for $\tilde{b}(n, d) \sim \mathcal{B}(n, \xi)$ in (88) we again argue that $\tilde{b}(n, d) \sim b_*(n, d)$ holds for $\xi = O(1)$ (in fact this relation fails only for $n = O(1)$ and $d \rightarrow \infty$). If instead of defining ξ from (84) we let

$$d = [\log_{1/p}(n)] + \xi' = \log_{1/p}(n) + \xi' - \{\log_{1/p}(n)\}, \quad (97)$$

where $\{\cdot\}$ denotes the fractional part, then

$$p^d n = p^{\xi'} \exp[-d \log(1/p) \{\log_{1/p}(n)\}]$$

and for $n \rightarrow \infty$ with $\xi, \xi' = O(1)$ the limiting form of (80) is

$$\frac{1}{2\pi i} \int_{Br} p^{-s\xi'} \mathcal{M}(s) p^{sd\{\log_{1/p}(n)\}} ds \quad (98)$$

with $\mathcal{M}(\cdot)$ as in (79). We therefore conjecture that the right side of (88) is given explicitly by (98), with ξ in (88) replaced by ξ' in (97).

References

- [1] Y. Choi and W. Szpankowski, Compression of graphical structures, *IEEE Transaction on Information Theory*, 58, 2012; also *IEEE International Symposium on Information Theory*, Seoul, 364–368, 2009.
- [2] L. Devroye, A Study of Trie-Like Structures Under the Density Model, *Annals of Applied Probability*, 2, 402–434, 1992.
- [3] M. Drmota, *Random Trees*, Springer, New York, 2009.
- [4] M. Drmota and W. Szpankowski, The Expected Profile of Digital Search Trees, *J. Combin. Theory, Ser. A*, 118, 1939–1965, 2011.
- [5] P. Flajolet, X. Gourdon, and P. Dumas, Mellin Transforms and Asymptotics: Harmonic sums, *Theoretical Computer Science*, 144, 3–58, 1995.
- [6] P. Flajolet, Singularity Analysis and Asymptotics of Bernoulli Sums, *Theoretical Computer Science*, 215, pp. 371–381, 1999.
- [7] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2008.
- [8] P. Jacquet, and W. Szpankowski, Analytical Depoissonization and Its Applications, *Theoretical Computer Science*, 201, 1–62, 1998.
- [9] P. Jacquet and W. Szpankowski, Entropy Computations via Analytic Depoissonization, *IEEE Trans. Information Theory*, 45, 1072–1081, 1999.
- [10] S. Janson and W. Szpankowski, Analysis of an Asymmetric leader Election Algorithm, *Electronic J. of Combinatorics*, 4, R17, 1997.
- [11] C. Knessl, and W. Szpankowski, Asymptotic Behavior of the Height in a Digital Search Tree and the Longest Phrase of the Lempel-Ziv Scheme, *SIAM J. Computing*, 30, 923–964, 2000.
- [12] D. Knuth, *The Art of Computer Programming. Sorting and Searching*, Vol. 3, Second Edition, Addison-Wesley, Reading, MA, 1998.
- [13] G. Louchard, Exact and Asymptotic Distributions in Digital and Binary Search Trees, *RAIRO Theoretical Inform. Applications*, 21, 479–495, 1987.
- [14] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons Inc., New York, 1992.
- [15] G. Park, H.K. Hwang, P. Nicodeme, and W. Szpankowski, Profile of Tries, *SIAM J. Computing*, 8, 1821–1880, 2009.
- [16] B. Pittel, Asymptotic Growth of a Class of Random Trees, *Annals of Probability*, 18, 414–427, 1985.
- [17] B. Pittel, Path in a Random Digital Tree: Limiting Distributions, *Advances in Applied Probability*, 18, 139–155, 1986.

- [18] W. Szpankowski, A Characterization of Digital Search Trees From the Successful Search Viewpoint, *Theoretical Computer Science*, 85, 117–134, 1991.
- [19] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.