A Note on a Problem Posed by D. E. Knuth on a Satisfiability Recurrence

DEDICATED TO PHILIPPE FLAJOLET 1948-2011

July 4, 2013

Philippe Jacquet
Bell Labs
Alcatel-Lucent
91620 Nozay
France
philippe.jacquet@inria.fr

Charles Knessl[1]
Dept. Math. Stat. & Compt. Sci.
University of Illinois at Chicago
Chicago, Illinois 60607-7045
U.S.A.
knessl@uic.edu

Wojciech Szpankowski[2]
Dept. Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

We resolve a conjecture proposed by D.E. Knuth concerning a recurrence arising in the satisfiability problem. Knuth's recurrence resembles recurrences arising in the analysis of tries, in particular PATRICIA tries, and asymmetric leader election. We solve Knuth's recurrence exactly and asymptotically, using analytic techniques such as the Mellin transform and analytic depoissonization.

1 Introduction

In this note we consider the following recurrence for the sequence $\{T_n\}$:

$$T_n = n + 2\sum_{k=1}^{n-1} \binom{n}{k} p^k q^{n-k} T_k, \quad n \geq 2, \tag{1}$$

with $T_0 = 0$, $T_1 = 1$, $0 < p < 1$ and $p+q = 1$. At the Analysis of Algorithms (AoA) conference in Montreal in 2012, R. Sedgewick [12] reported that D. E. Knuth [10] was interested in its solution, especially for $p = \frac{2}{3}$, but also for general p. Here we compute explicitly, for any p, the exponential generating function of T_n, and then give asymptotic results for $n \to \infty$. The asymptotics involve nearly constant (i.e., small amplitude) periodic functions of $\log_{1/p} n$, and these we explicitly calculate. The analysis is different for $p = \frac{1}{2}$ and $p \neq \frac{1}{2}$.

The motivation for studying recurrences such as (1) arises from analyzing the "satisfiability problem". As defined in [3], satisfiability is NP-complete, and the procedure is exponential in the worst case. The Davis-Putnam procedure [1] is a method for solving the satisfiability problem, and the method determines if a conjugate normal form (CNF) is satisfiable. We do not describe this procedure in detail, referring the reader to [3] for the description of the five steps in the Davis-Putnam procedure. A distribution on the CNF involves two integer

[1]This author was supported by NSA Grant H98230-11-1-0184.

[2]This author was supported by NSF Science & Technology Center Grant CCF-0939370, NSF Grants DMS-0800568, and CCF-0830140, and NSA Grant H98230-11-1-0141. He is also a Visiting Professor at ETI, Gdańsk University of Technology, Poland.

variables, the number of clauses n and the number of variables L. The average time analysis then involves solving the two variable recurrence

$$T_n^{L+1} = anL + 2\sum_{k=1}^{n-1} \binom{n}{k} p^k q^{n-k} T_k^L, \quad n \geq 1, \ L \geq 0, \tag{2}$$

with the boundary conditions $T_n^0 = 0$ and $T_0^L = 0$, and a is a constant. There are also alternate versions of the Davis-Putnam procedure that lead to similar recurrences (see equation (1.1) in [3]).

In [2] and [3] some bounds on the solution were obtained, using the solutions to simpler, one variable recurrences, such as that in (1). The particular recurrence in (1) is discussed in [2], and there the special case $p = \frac{2}{3}$ is further motivated [10]. In [2] it was conjectured that for $n \to \infty$, T_n in (1) is of order n^{β_*}, where $\beta_* = \lceil \frac{\log 2}{\log(1/p)} \rceil$, and this we verify, giving a more precise estimate in Theorem 1. We hope that our detailed analysis of (1) will help in the understanding of more complicated recurrences such as (2), which we plan to analyze in the future.

Knuth's recurrence (1) resembles recurrences arising in the analysis of tries [9, 14], in particular, PATRICIA tries [11, 13] and also asymmetric leader election algorithms [8]. In fact, the Poisson transform of Knuth's recurrence reduces to a certain functional equation often arising in the analysis of algorithms and data structures (cf. [4, 8]), namely,

$$f(z) = f(pz) + f(qz)e^{-pz} + a(z) \tag{3}$$

where $p + q = 1$ and $a(z)$ is a given function. Such a functional equation was studied before in [4, 8]. The point to observe is that $f(qz)$ is multiplied by the coefficient function e^{-pz}. This makes the problem interesting (otherwise a standard approach can be applied; cf. [14]). Second-order asymptotics of (3), and ultimately the recurrence (1), are quite challenging, especially if one strives to compute explicitly all of the constants and periodic functions involved. We believe that the novelty of our work lies in deriving exact and asymptotic solutions to Knuth's recurrence, including the constants and periodic functions. We also suggest a quickly converging numerical procedure to estimate these quantities.

The paper is organized as follows. In Section 2 we state our main results and discuss some of the basic asymptotics of T_n. In Section 3 we give the asymptotics more precisely, characterizing explicitly the periodic fluctuations of T_n. In Section 4 we develop an efficient numerical approach for evaluating some parts of the asymptotic formula for T_n.

2 Main Results

We summarize below our main results. We focus on the large n asymptotics, but our analysis (cf. (17)) also leads to an exact expression for the generating function $T(z)$ of T_n.

Theorem 1 (i) *Assume $p \neq \frac{1}{2}$. Then*

$$T_n = Cn^{-\frac{\log 2}{\log p}} \left[1 + P(\log n) + O(n^{-1})\right] + \frac{1}{1-2p} n \tag{4}$$

where C is constant that can be expressed as

$$C = \frac{2p}{2p-1}\frac{1}{\log(1/p)}\sum_{k=-\infty}^{\infty}(2p)^k\int_1^{1/p} y^{\frac{\log 2}{\log p}} e^{-qp^k y}\left[\prod_{j=1}^{\infty}\left(1-e^{-qp^{k-j}y}\right)\right] dy \qquad (5)$$

(see also (51)), and $P(x)$ is a periodic function of period $-\log p$ and small amplitude that can be explicitly expressed as

$$CP(\log n) := \sum_{\ell=-\infty,\ell\neq 0}^{\infty} c_\ell e^{2\pi i\ell \log_{1/p}(n)} \qquad (6)$$

where

$$c_\ell = \frac{2p}{2p-1}\int_0^1 e^{-2\pi i x\ell}\sum_{k=-\infty}^{\infty}(2p)^{k-x}e^{-qp^{k-x}}\left[\prod_{j=1}^{\infty}\left(1-e^{-qp^{k-x-j}}\right)\right] dx$$

for integer ℓ.
(ii) *For $p=\frac{1}{2}$ we have*

$$T_n = \frac{n\log n}{\log 2} + nP_*(\log n) + O(1)\ , \qquad (7)$$

where $P_*(x)$ is a periodic function of period $\log 2$*, explicitly shown in (49) and (50), which is constant to four decimal places, with $P_*(\log n) \approx .6295$.*

In the Table 1 we present the values of the constant $C = C(p)$ for $p \neq 0.5$. The reader is referred to Sections 3 and 4 for a more detailed discussion of this constant. Furthermore, we

p	$C(p)$
0.1	–0.123
0.2	–0.459
0.3	–1.292
1/3	–1.822
0.4	–3.947
0.6	4.837
2/3	2.161
0.7	1.344
0.8	0.102
0.9	3.687×10^{-6}

Table 1: The constant $C(p)$ versus p for some $p \neq 0.5$.

point out that the amplitude of the periodic function $CP(\log n)$ is very small, unless p itself is small. Let $x = \log_{1/p} n$, and define amplitude as $A = \max(x) - \min(x)$ where $\max(x)$ and $\min(x)$ are the maximum and minimum of P over the period $0 \leq x < 1$. For example for $p = 2/3$ we find that

$$\max(x) = .1542 \times 10^{-10}\ldots \quad \text{for} \quad x = 0.52,$$
$$\min(x) = -.1548 \times 10^{-10}\ldots \quad \text{for} \quad x = 0.02$$

so the amplitude is of order 10^{-10}.

Note that for $0 < p < \frac{1}{2}$ we have $T_n = O(n)$ and the term $n/(1-2p)$ dominates the asymptotics in (4), while for $\frac{1}{2} < p < 1$ we have $T_n = O(n^{\frac{\log 2}{\log(1/p)}})$. This means that for $0 < p < \frac{1}{2}$ both the term n and the sum in the right side of (1) are asymptotically important, while if $\frac{1}{2} < p < 1$ the sum dominates.

3 Proof of Main Results

We now give a proof of the theorem using analytic techniques such as the Mellin transform [5] and depoissonization [7] (see also [6, 14]). In the following we will set

$$\beta = -\frac{\log 2}{\log p}.$$

We will first establish a rough estimate. We shall prove that there exists an α such that for all integer n: $T_n < e^{\alpha n}$. Let N and α be large enough such that for all integer $n \leq \frac{1}{2}e^{\alpha n}$ and for all $n \leq N$:

$$T_n \leq e^{\alpha n} \tag{8}$$

$$pe^{\alpha} + q \leq \frac{e^{\alpha}}{2^{1/N}} \tag{9}$$

where we recall that $q = 1-p$. We show by recursion that $T_n \leq e^{\alpha n}$ for $n > N$. Indeed assuming that the hypothesis is true up to $n-1$, from (1) we have

$$\begin{aligned} T_n &= n + 2\sum_{k=1}^{n-1} \binom{n}{k} p^k q^{n-k} T_k \\ &\leq n + 2\sum_{k=1}^{n-1} \binom{n}{k} p^k q^{n-k} e^{\alpha k} \\ &= n + 2(pe^{\alpha} + q)^n. \end{aligned}$$

Since $n \leq \frac{1}{2}e^{\alpha n}$ and

$$(pe^{\alpha} + q)^n < \frac{1}{2^{n/N}} e^{\alpha n} < \frac{1}{2} e^{\alpha n},$$

we get $T_n \leq e^{\alpha n}$.

We define the Poisson transform as $T(z) = \sum_n T_n \frac{z^n}{n!} e^{-z}$ which exists for all complex z since the series converges due to the estimate $T_n \leq e^{n\alpha}$. We find from (1) that

$$T(z) = z + 2T(pz) - 2e^{-qz}T(pz). \tag{10}$$

We can prove that for $|\arg(z)| < \frac{\pi}{2} - \varepsilon$ and any fixed $\varepsilon > 0$, the following is true

$$|T(z)| \leq (|z| + |z|^{\beta})B \tag{11}$$

for some $B > 0$. Considering the sequence

$$B_k = \max_{|z| \leq p^{-k}} \left\{ \frac{|T(z)|}{|z| + |z|^{\beta}} \right\}$$

we can show, using similar arguments as in [7] or Theorem 10.5 of [14], that

$$B_k \leq (1 + e^{-qp^{-k}}) B_{k-1} \tag{12}$$

and thus the sequence $\{B_k\}$ is uniformly bounded.

On the other side, when $\arg(z) > \frac{\pi}{2} - \varepsilon$ then we can prove that there exists $\alpha' < 1$ such that $|T(z)e^z| \leq e^{\alpha'|z|}$ as described in [7]. Then by analytic depoissonization (for more details see [7]) we conclude that

$$T_n = T(n)\left(1 + O(n^{-1})\right).$$

Thus we need to establish asymptotic behavior of $T(z)$ for $z \to \infty$ in a cone around the real axis. We rewrite equation (10) as

$$T(z) = z + 2(1 - e^{-qz})T(pz) \tag{13}$$

and let

$$f(z) = \prod_{m>0} (1 - e^{-qp^{-m}z}). \tag{14}$$

The function $f(z)$ satisfies

$$f(pz) = (1 - e^{-qz})f(z).$$

We have $f(z) \to 1$ when $z \to +\infty$, in fact $f(z) = 1 - O(e^{-qz/p})$. Also $f(z) \to 0$ when $z \to 0$, and in fact from (14) we have $f(z) = o(z^M)$ for all $M > 0$. Let

$$\tau(z) = f(z)T(z)\ . \tag{15}$$

Then from (13) we obtain a new equation for $\tau(z)$

$$\tau(z) = zf(z) + 2\tau(pz) \tag{16}$$

that can be solved to give

$$\tau(z) = \sum_{m \geq 0} 2^m p^m z f(p^m z).$$

We thus have the solution of (13) in the form

$$T(z) = z + \sum_{m \geq 0} (2p)^{m+1} z \prod_{i=0}^{m} (1 - e^{-qp^i z}). \tag{17}$$

The Mellin transform

$$f^*(s) = \int_0^\infty f(z) z^{s-1} dz \tag{18}$$

of $f(z)$ is defined for $\Re(s) < 0$ while the Mellin transform of $zf'(z)$, that is $-sf^*(s)$, is defined for all complex s. We also have

$$\lim_{s \to 0} sf^*(s) = -\int_0^\infty f'(z)dz = -f(\infty) = -1\ . \tag{19}$$

For z real and positive, our previous estimates show that $\tau(z) = O(z + z^\beta)$ when $z \to \infty$ and $\tau(z) = o(z^M)$ for all $M > 0$ when $z \to 0$. The Mellin transform $\tau^*(s)$ of the function $\tau(z)$ is defined for $\Re(s) < -\max\{1, \beta\}$ and satisfies

$$\tau^*(s) = \frac{f^*(s+1)}{1 - 2p^{-s}}. \tag{20}$$

Via the inverse Mellin transform [5] we find that

$$\tau(z) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{f^*(s+1)}{1 - 2p^{-s}} z^{-s} ds. \tag{21}$$

Let us assume that $p \neq \frac{1}{2}$. The function $\frac{f^*(s+1)}{1-2p^{-s}}$ has simple poles:
(i) at $s = -1$, from $f^*(s+1)$,
(ii) at $s = s_k$, from $(1 - 2p^{-s})^{-1}$, where

$$s_k = -\beta + \frac{2ik\pi}{\log p}, \qquad k \in \mathbb{Z}.$$

Thus

$$\tau(z) = \frac{z}{1 - 2p} + \sum_{k \in Z} \frac{f^*(1 + s_k)}{\log p} z^{-s_k} + o(z^{-M}) \tag{22}$$

for any arbitrary $M > 0$. The factor $\frac{1}{1-2p}$ is a consequence of the residue of $-\frac{f^*(s+1)}{1-2p^{-s}}$ at $s = -1$, and (19). Notice also that

$$T(z) = \frac{\tau(z)}{f(z)} = \tau(z)(1 + O(e^{-qz/p}))$$

since $f(z)$ converges to 1 exponentially fast.

Finally, when $p = \frac{1}{2}$ the singularity at $s = s_0 = -1$ becomes a double pole and then

$$\tau(z) = \frac{z \log z}{\log 2} + \frac{z}{2} + \frac{(sf^*)'(0)}{\log 2} z + \sum_{k \in \mathbb{Z}^*} \frac{f^*(1 + s_k)}{-\log 2} z^{-s_k} + o(z^{-M})\ . \tag{23}$$

Here $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ denotes the set of all integers except $k = 0$.

4 Periodic Oscillations

In this section we shall obtain explicitly the periodic functions that appear in (4) and (7) within Theorem 1. We shall also obtain alternate expressions for the Poisson transform $T(z)$, from which the large z behavior is easily obtained.

4.1 Case $p \neq \frac{1}{2}$

First consider $p \neq \frac{1}{2}$ and set

$$T(z) = \frac{z}{1 - 2p} + \tilde{T}(z). \tag{24}$$

Then from (10) we obtain for $\tilde{T}(z)$ the functional equation

$$\tilde{T}(z) = 2(1 - e^{-qz})\tilde{T}(pz) - \frac{2pz}{1-2p}e^{-qz}. \tag{25}$$

Unlike (10) the non-homogeneous term in (25) decays exponentially for $z \to \infty$. If we furthermore set

$$\tilde{T}(z) = \left[\prod_{\ell=0}^{\infty}\left(1 - e^{-qp^{\ell}z}\right)\right] S(z). \tag{26}$$

we find that S satisfies

$$S(z) = 2S(pz) - \frac{2pz}{1-2p}e^{-qz}\left[\prod_{\ell=0}^{\infty}\left(1 - e^{-qp^{\ell}z}\right)^{-1}\right]. \tag{27}$$

Solving (27) by iteration leads to

$$S(z) = -\frac{2pz}{1-2p}\sum_{m=0}^{\infty}(2p)^{m}e^{-qp^{m}z}\left[\prod_{\ell=m}^{\infty}\left(1 - e^{-qp^{\ell}z}\right)^{-1}\right]. \tag{28}$$

Inverting the transform in (24) leads to

$$T_n = \frac{n}{1-2p} + \tilde{T}_n. \tag{29}$$

where

$$\tilde{T}_n = \frac{n!}{2\pi i}\oint \frac{e^{z}}{z^{n+1}}\tilde{T}(z)dz. \tag{30}$$

We expand $\tilde{T}_n$ for $n \to \infty$ by using a depoissonization argument, first expanding $\tilde{T}(z)$ as $z \to \infty$. Using (28) in (26) we let

$$m = \left\lfloor \frac{\log z}{\log(1/p)} \right\rfloor + k = \frac{\log z}{\log(1/p)} - \omega(z) + k. \tag{31}$$

where

$$\omega(z) = \left\langle \frac{\log z}{\log(1/p)} \right\rangle \tag{32}$$

and $\langle\cdot\rangle$ denotes the fractional part, so that $0 \leq \omega < 1$. It follows that

$$p^{m}z = p^{k}p^{-\omega}, \qquad 2^{m} = 2^{k-\omega(z)}z^{\frac{\log 2}{\log(1/p)}} \tag{33}$$

and then

$$\tilde{T}(z) = \frac{2p}{2p-1}z^{\frac{\log 2}{\log(1/p)}}\sum_{k=-\lfloor \log_{1/p}(z)\rfloor}^{\infty}(2p)^{k-\omega(z)}\exp(-qp^{k-\omega})R_k(z), \tag{34}$$

where

$$\begin{aligned} R_k(z) &= \prod_{\ell=0}^{m-1}\left(1 - e^{-qp^{\ell}z}\right) = \prod_{j=1}^{m}\left[1 - \exp(-qp^{k-\omega}p^{-j})\right] \\ &= \prod_{j=1}^{\infty}\left[1 - \exp(-qp^{k-\omega}p^{-j})\right](1 + O(\exp(-qz))\,. \end{aligned} \tag{35}$$

Note that the error in the asymptotic relation in (35) is exponentially small as $z \to \infty$. Then the leading term for $\tilde{T}(z)$ is obtained by replacing the lower limit on the sum in (34) by $k = -\infty$, and using (35). Thus we have obtained the leading term for $\tilde{T}(z)$ as $z \to +\infty$ for z real, up to an exponentially small error. In view of the appearance of the fractional part ω, it would appear that there may be a lack of smoothness in the asymptotic approximation. But this is in fact not the case as in the approximation we can drop the fractional part, thus replacing $\omega(z)$ by $\log_{1/p}(z)$, and then $p^{-\omega} = z$. Then the leading term is invariant under the mapping $z \to pz$ (just shift the summation index $k \to k-1$) and is thus an infinitely smooth function, for z real and positive. We argue that the asymptotic approximation will also hold in a sector in the complex $z-$plane containing the real axis (see the estimates in Section 2).

Then, by depoissonization,

$$T_n - \frac{n}{1-2p} = n^{\frac{\log 2}{\log(1/p)}} P_1(n) \times (1 + O(\frac{1}{n})) \tag{36}$$

where

$$P_1(n) = \frac{2p}{2p-1} \sum_{k=-\infty}^{\infty} (2p)^{k-\omega(n)} e^{-qp^{k-\omega(n)}} \prod_{j=1}^{\infty} \left(1 - e^{-qp^{k-\omega(n)-j}}\right) \tag{37}$$

and $\omega(n)$ is obtained by replacing z by n in (32). Again, we can replace $p^{-\omega(n)}$ by n. While the error in (37) is of the form $1 + O(n^{-1})$, from the depoissonization, we can improve the estimate by using the identity

$$\frac{n!}{2\pi i} \oint \frac{e^z}{z^{n+1}} z^{\beta} dz = \frac{\Gamma(n+1)}{\Gamma(n+1-\beta)}, \qquad \beta = \frac{\log 2}{\log(1/p)} \tag{38}$$

and the fact that the error term in (35) is exponentially small. We use (39) in the numerical studies in Section 5, while the results in Theorem 1 replace the right-hand side of (39) by n^{β}, since $n \to \infty$.

We have thus identified explicitly the periodic function $P(\cdot)$ in (4). Writing (38) as the Fourier series

$$P_1(n) = \sum_{\ell=-\infty}^{\infty} c_\ell e^{2\pi i \ell \log_{1/p}(n)} \tag{39}$$

we have

$$\begin{aligned} c_\ell &= \frac{2p}{2p-1} \int_0^1 e^{-2\pi i x \ell} \sum_{k=-\infty}^{\infty} (2p)^{k-x} e^{-qp^{k-x}} \left[\prod_{j=1}^{\infty} \left(1 - e^{-qp^{k-x-j}}\right)\right] dx \qquad (40) \\ &= \frac{2p}{2p-1} \frac{1}{\log(1/p)} \int_1^{1/p} \exp\left(\frac{2\pi i \ell \log(y)}{\log(p)}\right) y^{\frac{\log 2}{\log p}} \sum_{k=-\infty}^{\infty} (2p)^k e^{-qp^k y} \left[\prod_{j=1}^{\infty} \left(1 - e^{-qp^{k-j} y}\right)\right] dy. \end{aligned}$$

Here we set $y = p^{-x}$ in the first integral in (41). Comparing (37) to (4) we see that

$$C[1 + P(\log n)] = P_1(n) = c_0 + \sum_{\ell=-\infty, \ell \neq 0}^{\infty} c_\ell e^{2\pi i \ell \log_{1/p}(n)}, \tag{41}$$

so that $c_0 = C$ and $CP(\log n) = {}_1(n) - c_0$. In particular the zeroth Fourier coefficient is

$$c_0 = \frac{2p}{2p-1}\frac{1}{\log(1/p)}\sum_{k=-\infty}^{\infty}(2p)^k\int_1^{1/p} y^{\frac{\log 2}{\log p}} e^{-qp^k y}\left[\prod_{j=1}^{\infty}\left(1-e^{-qp^{k-j}y}\right)\right]dy. \tag{42}$$

For $p = 2/3$, numerical evaluation of the right side of (43) yields $c_0 \approx 2.1608$.

We note that the series in (43) is rapidly convergent, at both $k = \pm\infty$. For $k \to -\infty$ we have double exponential decay due to the factor $e^{-qp^k y}$ in the integrand. For $k \to +\infty$, the product in (43) behaves as (with $A = qy$)

$$\begin{aligned}\prod_{j=1}^{\infty}\left(1-e^{-Ap^{k-j}}\right) &= \exp\left[\sum_{\ell=1}^{\infty}\log\left(1-e^{-Ap^{-\ell}}\right)+\sum_{\ell=1-k}^{0}\log\left(1-e^{-Ap^{-\ell}}\right)\right] \\ &= \prod_{\ell=1}^{\infty}\left(1-e^{-Ap^{-\ell}}\right)\exp\left[\sum_{\ell=0}^{k-1}\log\left(\frac{1-e^{-Ap^{\ell}}}{Ap^{\ell}}\right)+\log(Ap^{\ell})\right] \\ &= \prod_{\ell=1}^{\infty}\left(1-e^{-Ap^{-\ell}}\right)\prod_{\ell=0}^{\infty}\left(\frac{1-e^{-Ap^{\ell}}}{Ap^{\ell}}\right)A^k p^{k(k-1)/2}(1+O(p^k)).\end{aligned} \tag{43}$$

Hence the product is roughly $O(p^{k^2/2})$, corresponding to Gaussian decay as $k \to +\infty$.

4.2 Case $p = \frac{1}{2}$

Next we consider $p = \frac{1}{2}$. Setting

$$T(z) = zT_*(z) \tag{44}$$

in (10), with $p = \frac{1}{2}$, leads to

$$T_*(z) = 1 + (1-e^{-z/2})T_*\left(\frac{z}{2}\right). \tag{45}$$

Solving (46) by iteration yields

$$T_*(z) = \sum_{m=0}^{\infty}\prod_{\ell=1}^{m}\left(1-e^{-z2^{\ell-m-1}}\right). \tag{46}$$

To expand (47) for $z \to \infty$ we again set $m = \lfloor\log_2 z\rfloor + k = \log_2 z - \omega + k$, as in (31). Then (47) becomes

$$\begin{aligned}T_*(z) &= \sum_{k=-\lfloor\log_2 z\rfloor}^{\infty}\left[\prod_{\ell=1}^{m}\left(1-e^{-2^{\ell-1+\omega-k}}\right)\right] \\ &= \sum_{k=0}^{\infty}\prod_{\ell=1}^{m}\left(1-e^{-2^{\ell-1+\omega-k}}\right)+\sum_{k=1}^{\lfloor\log_2 z\rfloor}\left[\prod_{\ell=1}^{m}\left(1-e^{-2^{\ell-1+\omega+k}}\right)-1+1\right] \\ &= \lfloor\log_2 z\rfloor+\sum_{k=0}^{\infty}\prod_{\ell=1}^{\infty}\left(1-e^{-2^{\ell-1+\omega-k}}\right)+\sum_{k=1}^{\infty}\left[\prod_{\ell=1}^{\infty}\left(1-e^{-2^{\ell-1+\omega+k}}\right)-1\right]+O(e^{-z^{\beta}}).\end{aligned} \tag{47}$$

It follows by depoissonization of (45) that

$$T_n = n\log_2 n + nQ_*(n) + O(1), \tag{48}$$

where

$$Q_*(n) \quad = \quad -\omega(n) + \sum_{k=0}^{\infty}\prod_{\ell=1}^{\infty}\left(1 - e^{-2^{\ell-1+\omega(n)-k}}\right) + \sum_{k=1}^{\infty}\left[\prod_{\ell=1}^{\infty}\left(1 - e^{-2^{\ell-1+\omega(n)+k}}\right) - 1\right].$$

The periodic function in (50) may once again be written as the Fourier series

$$Q_*(n) = \sum_{\ell=-\infty}^{\infty} c_\ell^* e^{2\pi i\ell\log_2(n)}, \tag{49}$$

and we have identified thus $P_*(\log n)$ in (7). In (49) $\omega(n) = \langle\log_2 n\rangle$ but we may drop the fractional part $\langle\cdot\rangle$, as if $\omega(n)$ is replaced by $\log_2 n$ it is easy to verify that $Q_*(n) = Q_*(n/2)$. Thus the expression below (49) is the same whether $\omega = \langle\log_2 n\rangle$ or $\omega = \log_2 n$. Numerical studies show that the right side of (50) fluctuates between .629494 and .629513, so this function is nearly constant, and may be approximated to four significant figures by the zeroth Fourier coefficient c_0^* in (50).

5 Further Numerical Analysis

Here we develop a semi-numerical semi-analytic method for evaluating the Fourier coefficients of the periodic functions that appear in Theorem 1. This will require that we compute the first few T_n numerically and use them to evaluate a rapidly converging series.

Using the analysis in Sections 2 and 3, we can extract the Mellin transform of $f(z)$ from numerical analysis. For $p = \frac{2}{3}$, for the problem originally proposed by D.E. Knuth [10], we have the leading term (see also (43))

$$C = \frac{f^*(s_0+1)}{\log p} \approx 2.1608643975035492760 6532 \tag{50}$$

where we recall that $s_0 = -\beta = \log 2/\log p$. The $k = 1$ term in (22)

$$-\frac{f^*(s_0 + 1 + \frac{2i\pi}{\log p})}{\log p} \approx -7.66\times 10^{-11} + 9.84\times 10^{-12} i. \tag{51}$$

The problem with the analysis in Section 2 is that it is somewhat difficult to extract the Mellin transform of $f(z)$. Below we propose an easier indirect method in the case β is not an integer (see also [8]).

Let k be an integer. For all analytic functions $F(z)$ in a complex neighborhood of 0, we define $\coprod_k F(z)$ as the Taylor polynomial of degree k:

$$\coprod\nolimits_k F(z) = \sum_{j=0}^{k} F^{(j)}(0)\frac{z^j}{j!} \tag{52}$$

with $F^{(j)}(z)$ the j-th derivative of function $F(z)$. We also define $\coprod^k F(z) = F(z) - \coprod_k F(z)$.

Let $b = \lfloor \beta \rfloor$ and recall that $T_0 = 0$ and $T_1 = 1$. The function $\coprod^b T(z)$ is $O(z^{b+1})$ when $z \to 0$ and is $O(z^\beta)$ when $z \to \infty$. Thus the Mellin transform $T^*(s)$ of $\coprod^b T(z)$ exists for $-b-1 < \Re(s) < -\beta$.

We set $\theta(z) = T(pz)e^{-qz}$. Its Mellin transform $\theta^*(s)$ exists for $\Re(s) > -1$ and satisfies (cf. also [8])

$$\theta^*(s) = \sum_{n \geq 1} p^n T_n \frac{\Gamma(s+n)}{n!} \ . \tag{53}$$

Since additive polynomial terms only shift the fundamental strip of the Mellin transform, from (13) we arrive at

$$T^*(s) = -2 \frac{\theta^*(s)}{1 - 2p^{-s}} \ . \tag{54}$$

Therefore the leading term in $T(z)$ is equal to

$$\sum_{k \in \mathbb{Z}} 2 \frac{\theta^*(s_k)}{\log p} z^{-s_k} \tag{55}$$

where we recall $s_k = -\beta + \frac{2ik\pi}{\log p}$. Thus for all $k \in \mathbb{Z}$

$$f^*(1 + s_k) = 2 \sum_{n=1}^{\infty} T_n p^n \frac{\Gamma(n + s_k)}{n!} \ . \tag{56}$$

Expression (57) is an implicit formula for the $f^*(1+s_k)$. The series converges geometrically in view of the factor p^n (and the algebraic growth T_n), but the T_n must be calculated from (1).

For $p = \frac{2}{3}$, by numerically evaluating (57) (using (1) to numerically compute the T_n), we get the $k = 0$ term as

$$C = \frac{f^*(s_0 + 1)}{\log p} \approx 2.1608643975035492 7606532$$

and the $k = 1$ term is

$$\frac{f^*(s_0 + 1 + \frac{2i\pi}{\log p})}{\log p} \approx -7.66 \times 10^{-11} + 9.84 \times 10^{-12} i.$$

For $p = \frac{1}{3}$ these quantities are respectively approximately -1.8219 and $-2.662 \times 10^{-4} + 1.853 \times 10^{-4} i$.

Omitting the periodic terms the only asymptotic terms of $T(z)$ are Cz^β and $\frac{z}{1-2p}$. Therefore, in view of (39), T_n and

$$C \frac{\Gamma(n+1)}{\Gamma(n+1-\beta)} + \frac{n}{1-2p}$$

converge exponentially fast (still omitting the periodic terms). Table 2 illustrates this for $p = 2/3$, where the exact T_n are computed from the recurrence (1).

Acknowledgment

We thank D. E. Knuth for several comments regarding this manuscript.

n	T_n	$C\frac{\Gamma(n+1)}{\Gamma(n+1-\beta)} + \frac{n}{1-2p}$
1	1.00000000000000000000	–2.30174306232304985900
10	74.19014780949336492932	73.91603109473686910765
20	291.00460566996026651879	291.00361735959358501362
30	619.38760608108754323680	619.38760213254548663704
40	1046.04525284496893107557	1046.04525440741658728759
60	2164.10852244470855065774	2164.10852363157796067744
80	3603.01754773643508459982	3603.01754785633937946726
100	5336.44933450906552767752	5336.44933395400537596930

Table 2: Exact vs Asymptotic Valus of T_n for $p = 2/3$.

References

[1] M. Davis, G. Logemann, and D. Loveland, A Machine Program for Theorem Proving, *Communications ACM*, 5, 394–397, 1962.

[2] A. Goldberg, On the Complexity of the Satisfiability Problem, Courant Comp. Sci. report No. 16, New York University, 1979.

[3] A. Goldberg, P. Purdom and C. Brown, Average Time Analyses of Simplified Davis-Putnam Procedures, *Information Processing Letters*, 15(2), 72–75, 1982.

[4] G. Fayolle, P. Flajolet, and M. Hofri, On a Functional Equation Arising in the Analysis of a Protocol for a Multi-Access Broadcast Channel, *Adv. Applied Probability*, 18, 441–472, 1986.

[5] P. Flajolet, X. Gourdon, and P. Dumas, Mellin Transforms and Asymptotics: Harmonic sums, *Theoretical Computer Science*, 144, 3–58, 1995.

[6] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.

[7] P. Jacquet, and W. Szpankowski, Analytical Depoissonization and its Applications, *Theoretical Computer Science*, 201, 1–62, 1998.

[8] S. Janson, and W. Szpankowski, Analysis of an Asymmetric Leader Election Algorithm, *Electronic J. Combinatorics*, 4, R17, 1997.

[9] D. E. Knuth, *The Art of Computer Programming. Sorting and Searching*, Vol. 3, Second Edition, Addison-Wesley, Reading, MA, 1998.

[10] D. E. Knuth, personal communication.

[11] B. Rais, P. Jacquet, and W. Szpankowski, Limiting Distribution for the Depth in Patricia Tries, *SIAM J. Discrete Mathematics*, 6, 197–213, 1993.

[12] R. Sedgewick, personal communication.

[13] W. Szpankowski, Patricia Tries Again Revisited, *J. ACM*, 37, 691–711, 1990.

[14] W. Szpankowski, *Analysis of Algorithms on Sequences*, John Wiley, New York, 2001.