Digits and beyond

Helmut Prodinger

ABSTRACT: *This is a survey about digits from a personal point of view. Counting the occurrences of digits (and, more generally, subblocks) is discussed in the context of various positional number systems. The methods to achieve this are Delange's elementary method and Flajolet's idea to use the Mellin–Perron summation (or integral) formula.*

Then we move to problems from Theoretical Computer Science (register function of binary trees, number of exchanges in Baxter's odd–questions are also discussed. An open problem from physicists Yekutieli and Mandelbrot can also be treated in that fashion.

Furthermore, we consider representations of numbers where some digits are forbidden. As a representative example, we discuss the Cantor distribution and its moments, asymptotically analyzed by Mellin transforms. Other problems in this context lead to sums involving Bernoulli numbers which can be attacked by analytic Depoissonization.

Very briefly we mention carry propagation, mergesort parameters and jump interpolation search trees.

1 Introduction

This is a survey about digits from a personal point of view, i. e., I will talk about digit expansions but stress those things that are in one way or another related to my own research.

We are interested in the binary expansion, and, more generally in base q–expansion, with an integer $q \ge 2$, and digits $\{0, \dots, q-1\}$. Delange [3] has analyzed the sum–of–digits function $S(n)$ (sometimes, especially for $q = 2$, written as $\nu(n)$), and found in particular the following result for the mean value

$$\frac{1}{m} \sum_{0 \le n < m} S(n) = \frac{q-1}{2} \log_q m + F(\log_q m),$$

with a periodic function $F(x)$ of period 1, which is continuous but nowhere differentiable, and with explicitly known Fourier coefficients. Perhaps even more important than his result, is his *method* which is based on finding an explicit formula for the individual digits, in terms of floor functions, rewriting them as integrals over piecewise constant functions. Forming the mean value just means the union of the intervals of integration, which makes things easier. Then one pulls out the main (logarithmic) term and finds that what remains is a periodic function of $\{\log_q m\}$ (the fractional part of $\log_q m$) which can be brought into some standard form from which one also gets the Fourier coefficients.

In Section 2 we consider such a Delange type analysis for the system with base q but digits $\{d, d+1, \dots, d+q-1\}$, where $d = 0$ means the standard case. The restriction for the parameter d is that 0 must be in the set of digits. We are counting how often a given digit i occurs in the representation of a natural number, more precisely, we consider the mean value of this counting function. This is based on the paper [26] which contains slightly more general results, namely not only counting individual digits i but rather a given (contiguous) subblock w. So, as an example, one knows how often "on average" the subblock 123014 occurs in the base 5 representation of n.

Delange's technique is elementary and nice, however, in more complicated situations, it can become quite messy. Flajolet came up with an alternative method [10]: It is based on the Mellin–Perron summation formula, which allows to write the counting function of interest as a contour integral in the complex plane. Shifting the line of integration produces then the terms in the explicit formula. As it often happens in this type of problems, one gets exact formulæ. The fact that the error term disappears completely is based on

$$0 = \int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \zeta(s) n^s \frac{ds}{s(s+1)}$$

or similar results. In most instances, this technique is somewhat superior to Delange's elementary approach. As an additional reading we recommend Marko Riedel's master's thesis [43].

In Section 3, we describe the technique for Delange's result ($q = 2$), and also for a related number system called the *Gray code*, which is obtained from the binary representation by a series of reflections. Interestingly enough, the sections of Knuth's *The Art of Computer Programming, Vol. 4*, dealing with Gray codes, are already available at the web, see Knuth's homepage http://Sunburn.Stanford.EDU/knuth/.

In Section 4, we discuss a symmetric number system with base q (an even integer) and digits $\{-q/2, \dots, q/2\}$. This redundant system is made unique by some conditions which in the instance $q = 2$ simply say that no two adjacent nonzero digits can occur. This instance $q = 2$ was rediscovered many times; perhaps the earliest reference is [42]. Some historical remarks can be found in [40]. We discuss again the problem of counting digits and subblocks in such a *symmetric signed system*.

Digits occur however in counting problems related to the *Analysis of Algorithms*. In Section 6 we deal with the *Register function* of binary trees (in the literature often appearing as Horton–Strahler numbers; here are some (random) references: [35, 30, 5, 4, 49, 33, 34]).

The question about the average value of the register function was solved independently by Flajolet–Raoult–Vuillemin [14] and Kemp [25], compare also [32]. The quantity $v_2(n)$, defined to be k if $n = 2^k(2i+1)$, plays a crucial role here. It is related to the sum–of–digits function $S(n)$, since $\sum_{k\le n} v_2(k) = n - S(n)$. Consequently one can, after some manipulations, invoke Delange's result. This was the route taken by Flajolet–Raoult–Vuillemin. Another approach (Kemp) is, after approximating the explicit quantities using real analysis, to consider Mellin transforms, which leads to Dirichlet series of the type $\sum_{n\ge1} v_2(n)/n^s$. Perhaps the method that is most advanced [12], also uses Mellin transforms, but *on the level of generating functions*, in order to find the local behaviour of a generating function closely related to the desired average values. Once this is established, *singularity analysis of generating functions* (transfer theorems), developed in [11], provides the asymptotics of the, say, average value of the register function of binary trees of size n.

Section 7 is in a sense the companion of Section 6. The analysis of Batcher's odd–even method (see Knuth [29]) was originally left open and eventually solved by Sedgewick [44]. What was the binary expansion and the quantity $v_2(n)$ for the register problems is now the *Gray code* representation and the function $\vartheta(k)$ which is 1 for $k = 2^m(4i+1)$, and -1 for $k = 2^m(4i+3)$. Flajolet and Ramshaw [13] use a result analogous to Delange's but for the Gray code representation (also included in Section 3), and proceed along the lines of [14]. The more advanced technique (singularity analysis of generating functions) is also described here and was worked out a long time ago for the proceedings of a French summer school in Ile de Ré [38].

The next Section 8 deals with an open problem which is related to the register function of binary trees. It was posed by Yekutieli and Mandelbrot [49] and solved in [39]. The register function can be extended by attaching a value to each internal node in the binary tree (which is the register function of the subtree having that node as the root), the value at the root being the register function of the tree. The question was to count the number of internal nodes having register function exactly 1 less than the register function of the entire tree. It was observed empirically that the expected value of this parameter is asymptotically a periodic function of $\log_4 n$ if all trees of size n (n internal nodes) are considered to be equally likely. Here, the following arithmetic function $\psi(m)$ plays a role: $\psi(m) = -k$ if $m = 2^i(4k+1)$ and $\psi(m) = k+1$ if $m = 2^i(4k+3)$. Otherwise, the same comments about the methodology as for the previous sections apply here.

The Section 9 deals with sets of numbers which have representations avoiding digits from a subset of *forbidden digits*. The most common case is $q = 3$ with forbidden digit 1. If one considers numbers of the form $\sum_{k\ge 1} a_k 3^{-k}$ where $a_k \in \{0,2\}$, one obtains the classical Cantor set. One question is about the moments of the Cantor distribution (strings of the form $\sum_{k\ge 1} a_k 3^{-k}$ are produced at random, where "random" means that each digit might be 0 or 2, with probability 1/2). Grabner and I could satisfactorily solve this problem, which originated in [31], using the method of Mellin transforms. Another problem, suggested in [23], was eventually solved by A. Knopfmacher and myself [27]: If one draws n random numbers, according to the Cantor distribution, what is the expected value of their *minimum*? Using exponential generating functions to solve the relevant difference equations, one can relate it to "the hardest asymptotic nut" from [28]. These days, one has alternative methods to attack the sums coming out in the analysis à la "asymptotic nut;" these are techniques known as Rice's method [15] which are again intimately related to Mellin transforms, and (analytic) Depoissonization, which was floating around for years in the literature, an early example being perhaps [45]. Recently it has been *polished* and systematized by Jacquet and Szpankowski [24].

The last sections are very brief and mention *carry propagation* in positional number systems, mergesort, and jump interpolation search trees.

2 A Delange type analysis

We describe Delange's technique by considering the number system with base q and digits $\{d, d+1, \ldots, d+q-1\}$ where $1-q \le d \le 0$. We are going to study $B_i(n)$, the number of occurrences of digit $i \ne 0$ in the representation of n. This is not a real restriction since the number of occurrences of digit 0 can be obtained as the difference of the length of the representation and the number of the other digits. A full treatment, even for subblocks instead of just digits, is in [26].

First, one needs an explicit formula: If n is written as $\ldots a_2a_1a_0$ in the (q,d)–system, then $a_k = i$ iff

$$\left\lfloor \frac{n}{q^{k+1}} + 1 - \frac{i}{q} - \frac{d}{q(q-1)} + \frac{1}{q} \right\rfloor - \left\lfloor \frac{n}{q^{k+1}} + 1 - \frac{i}{q} - \frac{d}{q(q-1)} \right\rfloor = 1.$$

Consequently,

$$B_i(n) = \sum_{k\geq 1} \left(\left\lfloor \frac{n}{q^{k}} + 1 - \frac{i}{q} - \frac{d}{q(q-1)} + \frac{1}{q} \right\rfloor - \left\lfloor \frac{n}{q^{k}} + 1 - \frac{i}{q} - \frac{d}{q(q-1)} \right\rfloor \right).$$

Delange's simple but extremely useful observation is that one can write the floor functions as integrals. We are considering an average of the function $B_i(n)$:

$$\begin{aligned}\frac{1}{m}\sum_{n=0}^{m-1} B_i(n) &= \frac{1}{m}\sum_{n=0}^{m-1}\sum_{k\geq 1} \left(\left\lfloor \tfrac{n}{q^{k}} + 1 - \tfrac{i}{q} - \tfrac{d}{q(q-1)} + \tfrac{1}{q} \right\rfloor - \left\lfloor \tfrac{n}{q^{k}} + 1 - \tfrac{i}{q} - \tfrac{d}{q(q-1)} \right\rfloor \right)\\ &= \frac{1}{m}\int_0^m \sum_{k\geq 1} \left(\left\lfloor \tfrac{t}{q^{k}} + 1 - \tfrac{i}{q} - \tfrac{d}{q(q-1)} + \tfrac{1}{q} \right\rfloor - \left\lfloor \tfrac{t}{q^{k}} + 1 - \tfrac{i}{q} - \tfrac{d}{q(q-1)} \right\rfloor \right) dt.\end{aligned}$$

It can be readily checked that nonzero contributions are only possible for $k \leq l+1$, with $l = \lfloor \log_q m - \log_q(1+\frac{d}{q-1})\rfloor$. After a simple change of variable in the integrals, it turns out that one should study functions like

$$g_\beta(x) = \int_0^x \left(\left\lfloor u + \beta + \tfrac{1}{q} \right\rfloor - \lfloor u+\beta \rfloor - \tfrac{1}{q} \right) du, \qquad h_\beta(x) = \sum_{k\geq 0} q^{-k} g_\beta(xq^k)$$

and

$$H_\beta(x) = \frac{1-\gamma-\{-\gamma+x\}}{q} + q^{1-\gamma-\{-\gamma+x\}} h_\beta(q^{\{-\gamma+x\}-1+\gamma}),$$

with $\{x\} = x - \lfloor x \rfloor$ (the fractional part of x) and $\gamma = \log_q(1+\frac{d}{q-1})$.

Eventually one finds that

$$\frac{1}{m}\sum_{n=0}^{m-1} B_i(n) = \frac{\log_q m}{q} + H_\beta(\log_q m),$$

where β depends on the digit i in a simple way: $\beta = 1 - \frac{i}{q} - \frac{d}{q(q-1)} - \frac{1}{q}$.

Since the function $H_\beta(x)$ is periodic with period 1, it is natural to study its *Fourier series* expansion. The result is: $H_\beta(x) = \sum_{k\in\mathbb{Z}} h_k e^{2\pi i k x}$, with

$$\begin{aligned} h_0 &= \log_q \frac{\Gamma(1-\{\beta+\frac{1}{q}\})}{\Gamma(1-\{\beta\})} - \frac{1}{q}\Big(\frac{1}{2} + \frac{1}{\log q}\Big),\\ h_k &= \frac{\zeta(\chi_k, 1-\{\beta+\frac{1}{q}\}) - \zeta(\chi_k, 1-\{\beta\})}{\log q \cdot \chi_k(1+\chi_k)}, \quad k \neq 0, \end{aligned}$$

with $\chi_k = 2\pi i k/\log q$ and the Hurwitz' zeta function $\zeta(z,a)$.

The computation of the Fourier coefficients is relatively straight forward; one has to use results like

$$\int_1^\infty \frac{\lfloor u \rfloor\, du}{u^{s+1}} = \sum_{k\ge 1} k \int_k^{k+1} \frac{du}{u^{s+1}} = \frac{1}{s}\sum_{k\ge1} k\Big(\frac{1}{k^s} - \frac{1}{(k+1)^s}\Big) = \frac{1}{s}\sum_{k\ge 1}\frac{1}{k^s} = \frac{1}{s}\zeta(s).$$

The computations in the case of counting subblocks are more complicated, and in the end an additional error term of order $1/m$ occurs.

3 The Mellin–Perron summation formula

This section is based on [10]. Let $f(x)$ be a function defined over $[0,+\infty)$. Its Mellin transform $f^*(s) = \mathcal{M}[f(x);s]$ is defined by $f^*(s) = \int_0^\infty f(x)x^{s-1}dx$. By linearity and the rescaling property we have

$$F(x) = \sum_k \lambda_k f(\mu_k x) \quad \Longrightarrow \quad F^*(s) = \Big(\sum_k \lambda_k \mu_k^{-s}\Big) f^*(s). \tag{1}$$

The condition is for s to belong to a 'fundamental strip' defined by the property that the integral giving $f^*(s)$ and the sum $\sum_k \lambda_k\mu_k^{-s}$ are both absolutely convergent. Similar to the Laplace transform there is an inversion theorem (cf. [7]). When applied to (1), it provides

$$\sum_k \lambda_k f(\mu_k x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Big(\sum_k \lambda_k\mu_k^{-s}\Big) f^*(s)x^{-s}ds, \tag{2}$$

with c in the fundamental strip.

Formula (2) could be called Mellin's summation formula. It is especially useful when the integral can be computed by *residues*, and in that case each residue contributes a term in an asymptotic expansion of $F(x)$.

This formula lends itself to various number theoretic applications, most notably proofs of the prime number theorem. Introduce the step function $H_0(x)$ defined by

$$H_0(x) = \begin{cases} 1 & \text{if } x \in [0,1], \\ 0 & \text{if } x > 1, \end{cases}$$

together with the functions $H_m(x) = (1-x)^m H_0(x)$. In the interesting case where $\mu_k \equiv k$, we obtain from (2), formulæ of the Perron type that provide integral representations for the iterated summations of arithmetic functions in terms of their Dirichlet generating function.

Let $c > 0$ lie in the half–plane of absolute convergence of $\sum_k \lambda_k k^{-s}$. Then for any $m \ge 1$, we have

$$\frac{1}{m!}\sum_{1\le k<n}\lambda_k\Big(1-\frac{k}{n}\Big)^m = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Big(\sum_{k\ge1}\frac{\lambda_k}{k^s}\Big)n^s\frac{ds}{s(s+1)\cdots(s+m)}. \tag{3}$$

For $m = 0$,

$$\sum_{1\le k<n}\lambda_k + \frac{\lambda_n}{2} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Big(\sum_{k\ge1}\frac{\lambda_k}{k^s}\Big)n^s\,\frac{ds}{s}.$$

Formula (3) is obtained from (2) by setting $x \equiv n^{-1}$, $f(x) \equiv H_m(x)$, and observing that $H_m^*(s) = m!(s(s+1)\cdots(s+m))^{-1}$. For $m = 0$ the formula has to be modified slightly by taking a principal value for the sum, since $H_0(x)$ is discontinuous at $x = 1$. For instance, if we use $\lambda_k \equiv 1$ and $m = 1$, we get

$$\frac{n-1}{2} = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s) n^s \frac{ds}{s(s+1)}.$$

Shifting the line of integration to the left and taking residues into account we obtain

$$0 = \int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \zeta(s) n^s \frac{ds}{s(s+1)}. \tag{4}$$

Identity (4) is the basis for the existence of several *exact* rather than plainly *asymptotic* summation formulæ. We use this Mellin–Perron technique to derive an alternative proof of Delange's theorem: The sum–of–digits function $S(n)$ satisfies

$$S(n) = \frac{1}{2} n \log_2 n + n F_0(\log_2 n),$$

where $F_0(u)$ can be represented by the Fourier series $F_0(u) = \sum_{k\in\mathbb{Z}} f_k e^{2\pi i k u}$ and

$$\begin{aligned} f_0 &= \frac{\log_2 \pi}{2} - \frac{1}{2\log 2} - \frac{3}{4}, \\ f_k &= -\frac{1}{\log 2}\frac{\zeta(\chi_k)}{\chi_k(\chi_k+1)} \quad \text{for} \quad \chi_k = \frac{2\pi i k}{\log 2}, \quad k \neq 0. \end{aligned}$$

Let $v_2(k)$ be the exponent of 2 in the prime decomposition of k and $\nu(k)$ be the number of 1-digits in the binary representation of k. We have $\nu(k) - \nu(k-1) = 1 - v_2(k)$, so that $S(n)$ resembles a double summation of $v_2(k)$. Furthermore, it is well-known (and also rederived in Section 6) that

$$\sum_{k\geq 1} \frac{v_2(k)}{k^s} = \frac{\zeta(s)}{2^s - 1}.$$

Thus, from (3), with $\lambda_k = v_2(k)$ and $m = 1$, we get the basic integral representation

$$S(n) = \frac{n(n-1)}{2} - \frac{n}{2\pi i}\int_{2-i\infty}^{2+i\infty} \frac{\zeta(s)}{2^s-1} n^s \frac{ds}{s(s+1)}.$$

The integrand has a simple pole at $s = 1$ (originationg from the ζ–function), a double pole at $s = 0$ and simple poles at $s = \chi_k$. Shifting the line of integration to $\Re(s) = -\frac{1}{4}$ and taking residues into account we get

$$S(n) = \frac{1}{2} n \log_2 n + n F_0(\log_2 n) - nR(n),$$

where the Fourier series (akin to F_0) $\frac{1}{\log 2}\sum_{k\in\mathbb{Z}} \frac{\zeta(\chi_k)}{\chi_k(\chi_k+1)} n^{\chi_k}$ occurs as the sum of residues of the integrand at the imaginary poles $s = \chi_k$. The remainder term is

$$R(n) = \frac{1}{2\pi i}\int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \frac{\zeta(s)}{2^s-1} n^s \frac{ds}{s(s+1)}, \tag{5}$$

so that there only remains to prove that $R(n) \equiv 0$ when n is an integer. The integral converges since $|\zeta(-\frac{1}{4}+it)| \ll |t|^{3/4}$ (cf. [48]). Using the expansion $\frac{1}{2^s-1} = -1-2^s-2^{2s}-2^{3s}-\cdots$ in (5), which is legitimate since now $\Re(s) < 0$, we find that $R(n)$ is a sum of terms of the form

$$\frac{1}{2\pi i}\int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \zeta(s)(2^k n)^s \frac{ds}{s(s+1)},$$

and each of these terms is 0 by virtue of (4).

It is clear from the discussion above that an exact formula for a sum–of–digits function is obtained each time a similar Dirichlet generating function can be introduced. Let us illustrate this point by the integral representation for the sum–of–digits function associated to *Gray code* representations. The Gray code representation of the integers starts like $0,\ 1,\ 11,\ 10,\ 110,\ 111,\ 101,\ 100,\ 1100,\ 1101,\ldots$; its characteristic is that the representations of n and $n+1$ differ in exactly one binary position and it is constructed in a simple manner by reflections based on powers of two (for a definition, see, e. g., [13]). Let $\gamma(k)$ be the number of 1-digits in the Gray code representation of k, and $\delta_k = \gamma(k) - \gamma(k-1)$. It is easy to see that $\delta_{2k} = \delta_k$, and the pattern for odd values is $\delta_{2k+1} = (-1)^k$. Thus the Dirichlet generating function $\delta(s)$ of $\{\delta_k\}$ is

$$\delta(s) = \frac{2^s L(s)}{2^s-1} \quad \text{with} \quad L(s) = \sum_{k\geq 0} \frac{(-1)^k}{(2k+1)^s}.$$

Thus by (2.4), the summatory function $G(n) = \sum_{k<n}\gamma(k)$ admits the integral representation

$$G(n) = \frac{n}{2\pi i}\int_{2-i\infty}^{2+i\infty} \frac{2^s L(s)}{2^s-1} n^s \frac{ds}{s(s+1)}.$$

The summatory function $G(n)$ of sum–of–digits function of Gray–code satisfies

$$G(n) = \frac{1}{2} n \log_2 n + nF_1(\log_2 n),$$

where $F_1(x)$ is representable by the Fourier series

$$F_1(x) = 2\log_2\Gamma\Big(\frac{1}{4}\Big) - \frac{3}{2} - \log_2 \pi + \frac{1}{\log 2}\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{L(\chi_k)}{\chi_k(\chi_k+1)} e^{2k\pi i x}.$$

4 Symmetric signed digit expansions

Recently, Heuberger and I [22] have considered a symmetric system with an *even* base q and digits $-\frac{q}{2},\ldots,\frac{q}{2}$. Such a system is a priori redundant because of the existence of both $\pm\frac{q}{2}$ but made unique by the condition that $n = \sum_{j\geq 0}\varepsilon_j q^j$, where $|\varepsilon_j| \leq q/2$ and $|\varepsilon_j| = q/2$ implies that $0 \leq \operatorname{sign}(\varepsilon_j)\varepsilon_{j+1} < q/2$. (Equivalent conditions where discussed in [22].) We call this expansion the *symmetric signed digit expansion* of n and denote it by $(\ldots\varepsilon_2(n)\varepsilon_1(n)\varepsilon_0(n))$. For $q=2$, this system was already considered by Reitwiesner in a computer science context [42]. Here is an algorithm to compute it from right to left (Algorithm 1):

Algorithm 1 Computation of the symmetric signed digit expansion of n

Input: $n > 0$, $q \geq 2$ integers.
Output: The representation ε.

$\varepsilon \leftarrow ()$; $m \leftarrow n$
while $m > 0$ **do**
  $a \leftarrow (m \bmod q)$
  if $a > q/2$ **or** $(a = q/2$ **and** $\{m/q^2\} \geq 1/2)$ **then**
    $a \leftarrow a - q$
  end if
  $m \leftarrow (m - a)/q$; $\varepsilon \leftarrow \varepsilon \,\&\, a$
end while

We were able to derive an *explicit formula* for the digit ε_r, which for $q = 2$ is

$$\varepsilon_r = \left\lfloor \frac{n}{2^{r+2}} + \frac{5}{6} \right\rfloor - \left\lfloor \frac{n}{2^{r+2}} + \frac{4}{6} \right\rfloor - \left\lfloor \frac{n}{2^{r+2}} + \frac{2}{6} \right\rfloor + \left\lfloor \frac{n}{2^{r+2}} + \frac{1}{6} \right\rfloor,$$

and was already obtained in [40]. For $q = 6$ it reads

$$\begin{aligned}
\varepsilon_r = &+ \lfloor y + \tfrac{248}{252} \rfloor + \lfloor y + \tfrac{241}{252} \rfloor + \lfloor y + \tfrac{234}{252} \rfloor - 5 \lfloor y + \tfrac{228}{252} \rfloor + \lfloor y + \tfrac{221}{252} \rfloor + \lfloor y + \tfrac{214}{252} \rfloor \\
&+ \lfloor y + \tfrac{206}{252} \rfloor + \lfloor y + \tfrac{199}{252} \rfloor + \lfloor y + \tfrac{192}{252} \rfloor - 5 \lfloor y + \tfrac{186}{252} \rfloor + \lfloor y + \tfrac{179}{252} \rfloor + \lfloor y + \tfrac{172}{252} \rfloor \\
&+ \lfloor y + \tfrac{164}{252} \rfloor + \lfloor y + \tfrac{157}{252} \rfloor + \lfloor y + \tfrac{150}{252} \rfloor - 5 \lfloor y + \tfrac{144}{252} \rfloor + \lfloor y + \tfrac{137}{252} \rfloor + \lfloor y + \tfrac{130}{252} \rfloor \\
&+ \lfloor y + \tfrac{122}{252} \rfloor + \lfloor y + \tfrac{115}{252} \rfloor - 5 \lfloor y + \tfrac{108}{252} \rfloor + \lfloor y + \tfrac{102}{252} \rfloor + \lfloor y + \tfrac{95}{252} \rfloor + \lfloor y + \tfrac{88}{252} \rfloor \\
&+ \lfloor y + \tfrac{80}{252} \rfloor + \lfloor y + \tfrac{73}{252} \rfloor - 5 \lfloor y + \tfrac{66}{252} \rfloor + \lfloor y + \tfrac{60}{252} \rfloor + \lfloor y + \tfrac{53}{252} \rfloor + \lfloor y + \tfrac{46}{252} \rfloor \\
&+ \lfloor y + \tfrac{38}{252} \rfloor + \lfloor y + \tfrac{31}{252} \rfloor - 5 \lfloor y + \tfrac{24}{252} \rfloor + \lfloor y + \tfrac{18}{252} \rfloor + \lfloor y + \tfrac{11}{252} \rfloor + \lfloor y + \tfrac{4}{252} \rfloor,
\end{aligned}$$

where $y = n/6^{r+2}$. In general, it has q^2 terms.

One sees the pattern $0, 1, 2, 3, -2, -1$ appearing 3 times followed by $0, 1, 2, -3, -2, -1$, also 3 times. The digits $0, 1, 2, 3, -2, -1$ are not symmetric around 0, and $0, 1, 2, -3, -2, -1$ are not, either. However, in combination, both coming with "probability" $1/2$, the distribution of the digits becomes symmetric.

In a recent paper with Grabner and Heuberger [16] we are addressing the subblock counting problem in symmetric signed digit expansions. To announce our principal findings, we need some notation. If a block $\mathbf{b} = (b_s, \ldots, b_0)$ is given, we denote its *value* by $\mathsf{value}(\mathbf{b}) = \sum_\ell b_\ell q^\ell$.

We also use Iverson's notation, popularized in [19]: $[P]$ is defined to be 1 if condition P is true, and 0 otherwise. With this notation we can count the number of subblock occurrences of $\mathbf{b}$ in (the symmetric signed digit expansions of) n via

$$\sum_{k \geq 0} [(\varepsilon_{k+r-1}(n), \ldots, \varepsilon_k(n)) = \mathbf{b}].$$

We only consider *admissible* blocks $\mathbf{b}$: these blocks represent the number $\mathsf{value}(\mathbf{b})$ in the symmetric signed digit expansion. For interest we note that there are $\frac{2+q}{1+q} q^r - \frac{1}{1+q}(-1)^r$ admissible blocks of length r; this was implicitly proved in [22]. We are studying the quantity

$$S_{\mathbf{b}}(N) = \sum_{n<N} \sum_{k \geq 0} [(\varepsilon_{k+r-1}(n), \ldots, \varepsilon_k(n)) = \mathbf{b}].$$

This is our main result:

Let $q \geq 2$ be an even integer and $r \geq 1$. For an admissible block $\mathbf{b} = (b_{r-1}, \dots, b_0)$ with $|b_{r-1}| < \frac{q}{2}$ and $\mathbf{b} \neq 0^r$ the number of occurrences of the block $\mathbf{b}$ in the symmetric signed digit expansions of the positive integers less than N satisfies

$$S_{\mathbf{b}}(N) = \frac{Q(b_0)}{q^r(q+1)} N \log_q N + h_0(\mathbf{b})N + NH_{\mathbf{b}}(\log_q N) + o(N),$$

where

$$\begin{aligned}
Q(\eta) =& q + \begin{cases} 2 & \text{for } \eta = 0, \\ 0 & \text{for } \eta = \pm\frac{q}{2}, \\ 1 & \text{else,} \end{cases} \\
H_{\mathbf{b}}(x) =& \sum_{k \in \mathbb{Z}\setminus\{0\}} h_k(\mathbf{b}) e^{2k\pi i x}, \\
h_k(\mathbf{b}) =& \frac{\log q}{2k\pi i(\log q + 2k\pi i)} \left(\zeta\left(\tfrac{2k\pi i}{\log q}, [\mathsf{value}(\mathbf{b}) < 0] + q^{-r}\,\mathsf{value}(\mathbf{b}) + \tfrac{R_{\min}(b_0)}{q^r(q+1)} \right) \right. \\
& \left. - \zeta\left(\tfrac{2k\pi i}{\log q}, [\mathsf{value}(\mathbf{b}) < 0] + q^{-r}\,\mathsf{value}(\mathbf{b}) + \tfrac{R_{\max}(b_0)}{q^r(q+1)} \right) \right) \text{ for } k \neq 0, \\
h_0(\mathbf{b}) =& \log_q \Gamma\left([\mathsf{value}(\mathbf{b}) < 0] + q^{-r}\,\mathsf{value}(\mathbf{b}) + \frac{R_{\min}(b_0)}{q^r(q+1)} \right) \\
& - \log_q \Gamma\left([\mathsf{value}(\mathbf{b}) < 0] + q^{-r}\,\mathsf{value}(\mathbf{b}) + \frac{R_{\max}(b_0)}{q^r(q+1)} \right) \\
& - \frac{Q(b_0)}{q^r(q+1)} \left(r + \frac{1}{2} + \frac{1}{\log q} - \frac{1}{q+1} \right) + \frac{1}{q^{r-1}(q+1)}, \\
R_{\min}(\eta) =& -\frac{q}{2} - \left[(\eta - 1) \bmod q \geq \frac{q}{2} \right], \qquad R_{\max}(\eta) = \frac{q}{2} + \left[\eta \bmod q < \frac{q}{2} \right].
\end{aligned}$$

The function $H_{\mathbf{b}}(x)$ is a periodic continuous function of period 1 and mean 0. As usual $\zeta(s, x)$ denotes the Hurwitz ζ–function.

The case of blocks $\mathbf{b}$ with most significant digit $b_{r-1} = \pm\frac{q}{2}$ can also be obtained from that, by a trivial argument taking differences.

The instance $r = 1$ (counting digits) was discussed in [22], although without mentioning the periodic fluctuations in explicit form. Thuswaldner [47] has used Dirichlet series and the Mellin–Perron summation formula to exhibit this fluctuating behaviour in the case $q = 2$ and $r = 1$.

The analysis is based on the Mellin–Perron formula, but with a slight twist:

$$\sum_{n<N} (N-n)a_n = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \sum_{n \geq 1} \frac{a_n}{(n-\alpha)^s} (N-\alpha)^{s+1} \frac{ds}{s(s+1)},$$

where $0 \leq \alpha < 1$ and c is in the half-plane of absolute convergence of the Dirichlet series. This slightly more general situation $0 \leq \alpha < 1$ was discussed in [17]. Without this version with the parameter α, one could still proceed successfully, as in [47], but that would be considerably more cumbersome and less elegant.

5 Representations of natural numbers as sums of 3 squares

This brief account is based on [46, 36, 43]. It is an old result that n is not representable as a sum of 3 squares iff $n = 4^i(8k+7)$, which means that in base-four representation, the last two nonzero digits are either 13 or 33. Let $Q(m)$ be the counting function of the numbers that are representable as a sum of 3 squares:

$$Q(m) = \sum_{0<n\le m} [n \text{ representable as a sum of 3 squares}] = m - \sum_{0<n\le m} \kappa(n),$$

with $\kappa(n) = 1$ if $n = 4^i(8k+7)$, $\kappa(n) = 0$ otherwise. It is a good idea to separate the main term and write $Q(m) = \frac{5}{6}m + \Delta(m)$. Osbaldestin and Shiu provide a Delange type analysis of the following formula:

$$\frac{1}{N} \sum_{0\le n<N} \Delta(n) = \frac{3}{8}\log_4 N + F(\log_4 N) + \frac{1}{8}\,[N \text{ is odd}]\,,$$

with $F(x) = \sum_{k\in\mathbb{Z}} f_k e^{2\pi i k x}$ and

$$\begin{aligned} f_0 &= -\frac{31}{48} - \frac{3}{8\log 4} - \frac{1}{\log 4}\Big(\zeta'(0,\tfrac{7}{16}) + \zeta'(0,\tfrac{15}{16})\Big), \\ f_k &= -\frac{1}{\log 4}\frac{1}{\chi_k(1+\chi_k)}\Big(\zeta(\chi_k,\tfrac{7}{16}) + \zeta(\chi_k,\tfrac{15}{16})\Big), \quad k \ne 0, \quad \chi_k = \frac{2\pi i k}{\log 4}. \end{aligned}$$

It was announced already in [10] that a proof of this result using the Mellin–Perron summation formula would be possible, and Marko Riedel has worked out the details in his Master's thesis [43].

6 The average number of registers to evaluate a binary tree of size n

The function "number of registers" *reg* is given recursively by $\mathrm{Reg}(\Box) = 0$ and

$$\mathrm{Reg}\left(\begin{smallmatrix} & \bigcirc & \\ t_1 & & t_2 \end{smallmatrix}\right) = \begin{cases} \max\{\mathrm{Reg}(t_1), \mathrm{Reg}(t_2)\} & \text{if } \mathrm{Reg}(t_1) \ne \mathrm{Reg}(t_2), \\ 1 + \mathrm{Reg}(t_1) & \text{otherwise.} \end{cases}$$

The problem is the evaluation of R_n, the average value of the function *reg*, based on the set of binary trees of size n. This was first studied in [14, 25]; for a tutorial see [38]. We have

$$R_n = \frac{1}{b_n}[z^n]E(z), \qquad \text{with} \qquad E(z) = \sum_{n\ge 0} z^n \sum_{t \text{ of size } n} \mathrm{Reg}(t)$$

and $b_n = \frac{1}{n+1}\binom{2n}{n}$ being the number of binary trees with n nodes. One can derive the following for the function $E(z)$, using the substitution $z = u/(1+u)^2$:

$$E(z) = \frac{1-u^2}{u} \sum_{p\ge 1} \frac{u^{2^p}}{1-u^{2^p}} = \frac{1-u^2}{u} \sum_{p\ge 1} \sum_{\lambda\ge 1} u^{\lambda 2^p} = \frac{1-u^2}{u} \sum_{n\ge 2} v_2(n) u^n,$$

where $v_2(n)$ is the exponent of 2 in the prime number decomposition of n. If we write in a unique way $n = 2^m(1+2i)$, ($m \ge 1$, $i \ge 0$), we have $v_2(n) = \sum_{p=1}^{m} 1 = m$. The quantity $v_2(n)$ (which we have seen already in Section 3) is often called "dyadic valuation," for obvious reasons. One can get the coefficients, e. g., using the Lagrange inversion formula:

$$[z^n]E(z) = \sum_{k\ge 1} v_2(n) \left[\binom{2n}{n-k} - 2\binom{2n}{n-k-1} + \binom{2n}{n-k-2} \right].$$

One strategy to proceed is the following (that was the method of Flajolet, Raoult, and Vuillemin): One application of partial summation brings, via $\nu(k) = k - \sum_{j\le k} v_2(j)$, the sum of digits function in, and a second application of partial summation makes the summatory function of the sum–of–digits function appear, which, according to Delange's analysis, is completely known! That means on the other side that the second difference of the binomial coefficients becomes eventually a fourth difference, but that causes no problems. The final result is this: The average number R_n of registers to evaluate a binary tree of size n satisfies

$$R_n = \log_4 n + D(\log_4 n) + o(1),$$

where $D(x)$ is a continuous periodic function of period 1; this function can be expanded in a Fourier series as $D(x) = \sum_{k\in\mathbb{Z}} d_k e^{2k\pi i x}$, with

$$d_0 = \frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2 \pi$$

and

$$d_k = \frac{1}{\log 2} \zeta(\chi_k) \Gamma\Big(\frac{\chi_k}{2}\Big)(\chi_k - 1), \quad k \ne 0, \quad \chi_k = \frac{2\pi i k}{\log 2}.$$

Another strategy, used by Kemp, is as follows: Approximation of the binomial coefficients leads eventually to the study of series of the type

$$f(x) := \sum_{k\ge 1} v_2(k) e^{-k^2 x^2};$$

the Mellin transform then brings in Dirichlet series $g(s) := \sum_{k\ge 1} v_2(k) k^{-s}$; but this one is easily evalutated, using $v_2(2k) = 1 + v_2(k)$ and $v_2(2k+1) = 0$:

$$g(s) = \sum_{k\ge 1} \big(1 + v_2(k)\big)(2k)^{-s} = 2^{-s}\big(\zeta(s) + g(s)\big),$$

from which one gets $g(s) = \zeta(s)/(2^s - 1)$.

A third strategy is more on the lines of [12] and based on singularity analysis of generating functions: One must study $E(z)$ for $z \to \frac{1}{4}$, i. e., for $u \to 1$, and sets

$u = e^{-t}$ and considers $E(z)$ for $t \to 0$. The second factor of $E(z)$, which is denoted by $V(t)$, is the challenging one. The Mellin transform of $V(t)$ is denoted by $V^*(s)$:

$$V^*(s) = \int_0^\infty x^{s-1} \sum_{k\geq 1} v_2(k) e^{-kx} dx = \sum_{k\geq 1} v_2(k) k^{-s} \int_0^\infty x^{s-1} e^{-x} dx.$$

This gives $V^*(s) = \frac{\Gamma(s)\zeta(s)}{2^s-1}$. The Mellin inversion formula then leads to

$$V(t) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} V^*(s) t^{-s} ds,$$

and we shift the line of integration to the left, taking the residues of the integrand $V^*(s)t^{-s}$ into account. Using some classical expansions we get

$$V(t) \sim \frac{1}{t} + \frac{1}{2}\log_2 t - \frac{1}{2}\log_2 2\pi + \frac{1}{4} + \frac{\gamma}{2\log 2} + \frac{1}{\log 2} \sum_{k\neq 0} \Gamma(\chi_k)\zeta(\chi_k) t^{-\chi_k} + \dots$$

which gives the local expansion for the function $E(z)$; the rest is then singularity analysis of generating functions [11].

7 Odd–even merge

In the analysis of a sorting algorithm (Batcher's odd–even merge [28]), one has to consider the quantity

$$B_n = \sum_{k\geq 1} \frac{\binom{2n}{n-k}}{\binom{2n}{n}} \Big(2F(k) + k \Big),$$

representing the number of exchanges in the odd–even merge of $2n$ elements. In this equation, $F(k) = \sum_{0\leq j<k} f(j)$ and $f(j)$ is the number of digits "1" in the Gray code representation of j. The numbers $\vartheta(j) := f(j) - f(j-1)$ equal ± 1: if k is of the form $k = 2^m(4i+1)$, $\vartheta(k) = 1$; if k is of the form $k = 2^m(4i+3)$, $\vartheta(k) = -1$. We might recognize a great similarity between $\vartheta(k)$ and $v_2(k)$ from Section 6.

Indeed, there are again basically 3 lines of attack: The Flajolet–Ramshaw approach [13] uses partial summation and uses an explicit result for the summatory function of the sum of digits function in the Gray code representation, as derived in Section 3; this is analogous to [14]. The next one, analogous to [25], was used by Sedgewick [44]: Approximating binomial coefficients, one has to study series à la

$$f(x) := \sum_{k\geq 1} \vartheta(k) e^{-k^2x^2},$$

or, after using Mellin transforms, Dirichlet series like $g(s) := \sum_{k\geq 1} \vartheta(k) k^{-s}$, but this one is again easily evaluated:

$$g(s) = \sum_{m,i\geq 0} \big(2^m(4i+1)\big)^{-s} - \sum_{m,i\geq 0} \big(2^m(4i+3)\big)^{-s} = \frac{1}{2^s(2^s-1)} \Big[\zeta(s,\tfrac14) - \zeta(s,\tfrac34)\Big],$$

where the Hurwitz $\zeta(s,a)$-function is defined by $\zeta(s,a) = \sum_{n\ge 0}(n+a)^{-s}$ for $\Re(s) > 1$. Recall that the function from Section 3 can be expressed as $L(s) = 4^{-s}\left[\zeta(s,\frac{1}{4}) - \zeta(s,\frac{3}{4})\right]$.

The third approach, using singularity analysis of generating functions, is presented in the tutorial [38]: One has to study the function

$$\frac{u(1+u)}{(1-u)^3}\sum_{i\ge 1}\vartheta(i)u^i$$

near $u=1$. Eventually one finds:

The average number of exchanges in the odd–even merge of $2n$ elements satisfies

$$B_n \sim \frac{1}{4}n\log_2 n + nB\big(\log_4 n\big),$$

where $B(x)$ is a continuous periodic function of period 1; this function can be expanded as a Fourier series $B(x) = \sum_{k\in\mathbb{Z}} b_k e^{2k\pi i x}$, with

$$b_0 = -\frac{1}{2\log 2} - \frac{\gamma}{4\log 2} - \frac{3}{4} + 2\log_2\Gamma\Big(\frac{1}{4}\Big) - \log_2\pi$$

and for $k \ne 0$

$$b_k = \frac{1}{\log 2}\zeta\big(\chi_k,\tfrac{1}{4}\big)\frac{\Gamma(\chi_k/2)}{1+\chi_k}.$$

8 A problem of Yekutieli and Mandelbrot

The following problem was left open (and attacked empirically) in [49] and later solved in [39].

If we have an extended binary tree, we label the leaves with 0, and, recursively, if the left subtree of a node is labeled with a and the right subtree with b, we label the node with $\max\{a,b\}$ if $a\ne b$ and with $a+1$ otherwise. The value attached to the root is called the *register function* of the tree t. The value attached to a particular node is the register function of the subtree having this node as its root, as already discussed earlier in this paper.

The authors in [49] consider the *bifurcation ratio (at the root)*. It is meant to be the *number of maximal subtrees (which is not the same as the number of internal nodes (!)) having register function exactly 1 less than the register function of the entire tree.* It was observed empirically that the expected value of this parameter is asymptotically a periodic function of $\log_4 n$ if all trees of size n (n internal nodes) are considered to be equally likely. Here, we want to settle this problem by explicitly describing the periodic function in terms of the Fourier coefficients. In principle, a full asymptotic expansion could be given, but the computation of the lower order term becomes more and more complicated.

Now, let $w_{p,k,n}$ be the number of binary trees with n nodes, register function p, and Yekutieli–Mandelbrot–parameter k, and let

$$W_p(z,y) = \sum_{n,k\ge 0} w_{p,k,n}\, y^k z^n$$

be its bivariate generating function.

To find the expected values, we have to work with $T_p(z) = \frac{\partial}{\partial y} W_p(z,y)|_{y=1}$ and $T(z) = \sum_{p\ge 1} T_p(z)$. The coefficient of z^n in $T(z)$, divided by $\frac{1}{n+1}\binom{2n}{n}$, is the expected value sought by Yekutieli and Mandelbrot. One finds for $p \ge 1$

$$W_p(z,y) = zy^2 R_{p-1}^2(z) + 2zyW_p(z,y)R_{p-1}(z) + 2zW_p(z,y)\big(B(z) - S_{p-1}(z)\big),$$

with $R_p(z)$ being the generating function of binary trees and register function $= p$, and $S_p(z) = \sum_{j\ge p} R_j(z)$. Therefore

$$T_p(z) = 2zR_{p-1}^2(z) + 2zR_p(z)R_{p-1}(z) + 2zT_p(z)\big(B(z) - S_p(z)\big),$$

and eventually (again with $z = u/(1+u)^2$)

$$T(z) = 2u + 2\frac{1-u^2}{u}\sum_{p\ge 1}\frac{u^{3\cdot 2^{p-1}}}{(1+u^{2^p})^2(1-u^{2^p})}.$$

One has to study the series

$$\begin{aligned}\sigma &= \sum_{p\ge 1}\frac{u^{3\cdot 2^{p-1}}}{(1+u^{2^p})^2(1-u^{2^p})} = \sum_{p\ge 0, k\ge 1} u^{2^p(2k+1)}(-1)^{k-1}\left\lfloor \tfrac{k+1}{2}\right\rfloor \\ &= -\sum_{p,k\ge 0} k u^{2^p(4k+1)} + \sum_{p,k\ge 0}(k+1)u^{2^p(4k+3)}.\end{aligned}$$

Now, one performs a Mellin transform analysis in order to find the local behaviour

$$T(z) \sim 3 - \left(\frac{4\mathcal{C}}{(\log 2)\pi} + 5\right)\sqrt{1-4z} - \frac{8}{\log 2}\sum_{k\ne 0}\Gamma(\chi_k)\beta(\chi_k - 1)(1-4z)^{(1-\chi_k)/2}.$$

Here, $\beta(s) = \zeta(s,\frac14) - \zeta(s,\frac34)$, which is the function that appeared already in the odd–even merge problem, $\chi_k = 2\pi i k/\log 2$, and $\mathcal{C} = \sum_{k\ge 0}(-1)^k/(2k+1)^2 = 0.9159655942\ldots$ is Catalan's constant. Consequently we get this result: The average value of the Yekutieli–Mandelbrot parameter, if all binary trees of size n are considered to be equally likely, is given by

$$\frac{2\mathcal{C}}{(\log 2)\pi} + \frac{5}{2} + \delta(\log_4 n) + \mathcal{O}\Big(\frac{1}{n}\Big).$$

The periodic function $\delta(x)$ has mean value 0 and admits the following representation as a Fourier series,

$$\delta(x) = -\frac{2}{\log 2}\sum_{k\ne 0}(\chi_k - 1)\Gamma\Big(\frac{\chi_k}{2}\Big)\beta(\chi_k - 1)e^{2k\pi i x}.$$

One could also perform an analysis along the lines of Kemp resp. Sedgewick [25, 44]; it would be based on the exact formula

$$[z^n]T(z) = \frac{2}{n+1}\binom{2n}{n} + 2\sum_{m\ge 1}\psi(m)\left[\binom{2n}{n+1-m} - 2\binom{2n}{n-m} + \binom{2n}{n-1-m}\right]$$

with $\psi(m) = \begin{cases} -k & \text{if } m = 2^i(4k+1) \quad \text{for some } i \text{ and } k, \\ k+1 & \text{if } m = 2^i(4k+3) \quad \text{for some } i \text{ and } k.\end{cases}$

9 Missing digits

Interesting phenomena occur if several digits are forbidden. We describe here the instance of $q = 3$ and digits $\{0, 2\}$, i. e., the digit 1 is forbidden. The set of strings $0.a_1a_2\ldots$, with $a_i \in \{0, 2\}$ is the classical *Cantor set*. Thus considering $2\sum_{i\ge 1} X_i 3^{-i}$, where the X_i are independent and identically distributed with probability distribution $\mathbb{P}\{X = 0\} = \mathbb{P}\{X = 1\} = 1/2$, we get a probability distribution on the interval $[0, 1]$ which is conveniently called *Cantor distribution*. In [18], the moments of a slightly more general distribution were investigated, motivated by an earlier paper of Lad and Taylor [31]; we sketch the procedure. Denote a_n the nth moment of the Cantor distribution; it is not hard to see that

$$a_n = \frac{1}{2(3^n-1)}\sum_{i=0}^{n-1}\binom{n}{i}2^{n-i}a_i, \quad n \ge 1, \quad a_0 = 1.$$

If one rewrites it as $(2-3^{-n})a_n = \sum_{i=0}^{n}\binom{n}{i}2^{n-i}a_i$ and introduces the exponential generating function $A(z) = \sum_{n\ge 0} a_n z^n/n!$, it translates into

$$2A(z) - 2 - A(\tfrac{z}{3}) + 1 = e^{2z/3}A(\tfrac{z}{3}) - 1,$$

or

$$A(z) = \frac{1+e^{2z/3}}{2}A(\tfrac{z}{3}) = \prod_{k\ge 1}\frac{1+e^{2z/3^k}}{2};$$

the last step was by iteration of the functional equation. Slightly more useful than this infinite product is the *Poisson transformed* generating function $B(z) = e^{-z}A(z)$, since by a process called *Depoissonization* one finds that $a_n \sim B(n)$. There is a well written survey paper on the subject by Jacquet and Szpankowski [24]. We get

$$B(z) = \frac{1+e^{-2z/3}}{2}B(\tfrac{z}{3}) = \prod_{k\ge 1}\frac{1+e^{-2z/3^k}}{2}.$$

To find the asymptotic behaviour of $B(z)$ for large z (and thus $B(n)$ and thus a_n) we compute the *Mellin transform* $B^*(s)$ of $B(z)$, viz.

$$B^*(s) = \frac{1}{2}3^s B^*(s) + \frac{1}{2}\int_0^\infty B(\tfrac{z}{3})e^{-2z/3}z^{s-1}dz,$$

or

$$B^*(s) = \frac{1}{2-3^s}\int_0^\infty \prod_{k\ge 2}\frac{1+e^{-2z/3^k}}{2}\; e^{-2z/3}z^{s-1}dz.$$

The standard reference for the Mellin transform in the context of asymptotic enumeration is [9]. The Mellin inversion formula gives

$$B(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} B^*(s)z^{-s}ds,$$

where the constant c might be chosen to be $-1/2$. Now one shifts the line of integration to the right and collects negative residues. They come from the solutions of $2-3^s=0$, i. e., $s=\log_3 2+2\pi ik/\log 3$, for $k\in\mathbb{Z}$. If one calls them f_k, then $B(n)\sim\sum_{k\in\mathbb{Z}}f_k n^{-\log_3 2-2\pi ik/\log 3}$, which is of the form $n^{-\log_3 2}\delta(\log_3 n)$, with a periodic function $\delta(x)$ of period 1. Typically, the amplitudes of such periodic functions are quite small, so the most interesting term is obtained for $k=0$. The negative residue at $s=\log_3 2$ is

$$\frac{1}{2\log 3}\int_0^\infty\prod_{k\geq 2}\frac{1+e^{-2z/3^k}}{2}\,e^{-2z/3}z^{\log_3 2-1}dz=0.734\ldots\ ,$$

and so $a_n\sim B(n)\simeq f_0 n^{-\log_3 2}\approx 0.734\,n^{-0.631}$.

Another interesting question was triggered by Hosking [23] and solved in [27]: Assume that we draw n random numbers (independently) according to the Cantor distribution. What is the expected value of the minimum of them? If one would draw from the interval $[0,1]$ according to the uniform distribution, then it is fairly easy to see that it is $1/(n+1)$. If we call that minimum a_n, we get

$$a_n=\frac{1}{3}2^{-n}\sum_{k=1}^{n}\binom{n}{k}a_k+\frac{1}{3}2^{-n}\big(2+a_n\big),\quad a_0:=0.$$

This recursion is obtained by observing that the minimum is obtained from one of the random strings starting with a zero, provided there is one. In the rare event that all strings start with 2, this first digit contributes $2/3$ (additively), and the minimum is sought from all the n participants. Rearranging the recursion and introducing the exponential generating function $A(z):=\sum_{k\geq 0}a_k z^k/k!$, one finds

$$A(2z)=\frac{1}{3}(1+e^z)A(z)+\frac{2}{3}(e^z-1).$$

Now we introduce, following Knuth [28], the function $\hat{A}(z):=A(z)/(e^z-1)=\sum_{k\geq 0}\hat{a}_k z^k/k!$ and get

$$\hat{A}(2z)=\frac{1}{3}\hat{A}(z)+\frac{2}{3}\frac{1}{e^z+1},$$

from which one finds

$$\hat{a}_n=-\frac{2}{3}\frac{B_{n+1}}{n+1}\frac{2^{n+1}-1}{2^n-\frac{1}{3}},$$

with Bernoulli numbers B_n. Hence one gets the explicit solution

$$a_n=-\frac{2}{3}\sum_{k=0}^{n-1}\binom{n}{k}\frac{B_{k+1}}{k+1}\frac{2^{k+1}-1}{2^k-\frac{1}{3}}.$$

In order to study the asympotic behaviour of that quantity one can use Rice's method [15], which is related to Mellin transforms, and eventually write

$$a_n=-\frac{2}{3(n+1)}\cdot\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{(-1)^n n!}{z(z-1)\ldots(z-n)}\big(-z\zeta(1-z)\big)\frac{2^z-1}{2^{z-1}-\frac{1}{3}}dz,$$

where the contour $\mathcal{C}$ encloses the poles $1,2,\ldots,n-1$ and no others. Deforming the contour of integration reduces the problem to the computation of residues; the dominant ones are the solutions of $2^{z-1}-\frac{1}{3}=0$, i. e., $1-\log_2 3+2k\pi i/\log 2$, for all $k\in\mathbb{Z}$. Eventually one finds, apart from a periodic function as above, that

$$a_n \simeq n^{\log_2 3}\frac{2}{3\log 2}\Gamma(\log_2 3)\zeta(\log_2 3).$$

Another type of moments are *Cantor's singular moments*: Consider the unique non–decreasing function on $[0,1]$ such that, if $x=2\sum_{j\ge1}\varepsilon_j 3^{-j}$ with $\varepsilon_j\in\{0,1\}$, then $F(x)=\sum_{j\ge1}\varepsilon_j 2^{-j}$. It was a question in the problem section of the American Mathematical Monthly [6] to compute the moments; the following solution came out:

$$J_n=\int_0^1\big(F(x)\big)^n dx=\frac{2}{3(n+1)}\sum_{j=0}^{n}\binom{n+1}{j}\frac{B_j}{3\cdot 2^{j-1}-1}\qquad\text{for } n\ge1 \text{ and } J_0=1;$$

the case $n=-1$ was left open and treated by me in [41]: Since one can show that $J_{-1}=\sum_{n\ge0}J_n$, and, again by Rice's method,

$$J_n=\frac{2}{3}\cdot\frac{1}{2\pi i}\int_{-\frac12-i\infty}^{-\frac12+i\infty}\frac{\Gamma(n+1)\Gamma(1-s)}{\Gamma(n+2-s)}\frac{\zeta(1-s)}{3\cdot 2^{s-1}-1}\,ds,$$

one finds

$$\begin{aligned}\sum_{n=0}^{N}J_n &= J_0+\frac{2}{3}\cdot\frac{1}{2\pi i}\int_{-\frac12-i\infty}^{-\frac12+i\infty}\frac{\Gamma(N+2)\Gamma(1-s)}{\Gamma(N+2-s)s}\frac{\zeta(1-s)}{3\cdot 2^{s-1}-1}\,ds\\ &\quad-\frac{2}{3}\cdot\frac{1}{2\pi i}\int_{-\frac12-i\infty}^{-\frac12+i\infty}\frac{\Gamma(2)\Gamma(1-s)}{\Gamma(2-s)s}\frac{\zeta(1-s)}{3\cdot 2^{s-1}-1}\,ds.\end{aligned}$$

Now we perform the limit $N\to\infty$ and get eventually (the computations are not displayed here):

$$\begin{aligned}\sum_{n\ge0}J_n &= \frac{4}{3}+\frac{2}{3}\sum_{k,m\ge1}3^{-k}\cdot\frac{1}{2\pi i}\int_{\frac32-i\infty}^{\frac32+i\infty}\frac{1}{s(s-1)}\Big(\frac{2^k}{m}\Big)^s ds\\ &= -\frac{1}{3}+\frac{2}{3}\sum_{k\ge1}\Big(\frac{2}{3}\Big)^k H_{2^k}=3.36465\,07281\,00925\,16083\,89349\,6289\ldots,\end{aligned}$$

with harmonic numbers $H_n=\sum_{1\le k\le n}\frac{1}{k}$.

More general results of this type are currently worked out in collaboration with F. Bassino [1].

10 Von Neumann's addition algorithm

This brief description is based on a recent paper with Heuberger [21].

Knuth [29] has analyzed von Neumann's addition algorithm: Assume that two integers are given in q–ary notation, say $(\ldots y_2y_1y_0)_q$ and $(\ldots y_2y_1y_0)_q$; then the integer $(\ldots s_2s_1s_0)_q$ with $s_i=(x_i+y_i)\mod q$ is formed, as well as $(\ldots c_2c_1c_0)_q$ (the carries), where $c_{i+1} = [x_i + y_i \geq q]$. The process is iterated by adding $(\dots s_2 s_1 s_0)_q$ with $s_i = (x_i + y_i) \mod q$ and $(\dots c_2 c_1 c_0)_q$ until the string of carries contains only zeros. Knuth studied the average number of iterations, assuming two random integers with n digits. The result is $\sim \log_q n$; it turns out that the longest subsequence of the form $\dots i(q-1)(q-1)\dots(q-1)j\dots$ with $i \neq q-1$ and $j \geq q$ in $\big(\dots(x_2+y_2)(x_1+y_1)(x_0+y_0)\big)_q$ is responsible for the number of iterations. We extended Knuth's results to other positional number systems, namely for the basis q and the set of q digits $\{d, d+1, \dots, d+q-1\}$. Note carefully that carries might now be ± 1 and that the sequence of sums might be oscillating, being smaller or larger than the true value of the sum of the two integers. This is in sharp contrast to the traditional q–ary system, where the sums are monotonically increasing until the algorithm stops. Thus it is perhaps natural that the description of subsequences being responsible for the number of iterations is significantly more complicated. Here is an example for $q = 5$, $d = -1$, and $\bar{1} = -1$:

$$
\begin{array}{rcrcr}
(12\bar{1}\bar{1}\bar{1}13)_{(5,-1)} &=& (\dots x_2 x_1 x_0)_{(5,-1)} &=& 21108 \\
(22\bar{1}0\bar{1}23)_{(5,-1)} &=& (\dots y_2 y_1 y_0)_{(5,-1)} &=& 36863 \\
\hline
(3\bar{1}3\bar{1}331)_{(5,-1)} &=& (\dots z_2 z_1 z_0)_{(5,-1)} &=& 45591 \\
(1\bar{1}0\bar{1}010)_{(5,-1)} &=& (\dots c_2 c_1 c_0)_{(5,-1)} &=& 12380 \\
\hline
(\bar{1}3333\bar{1}1)_{(5,-1)} &=& (\dots z_2 z_1 z_0)_{(5,-1)} &=& -3929 \\
(1\bar{1}0\bar{1}0100)_{(5,-1)} &=& (\dots c_2 c_1 c_0)_{(5,-1)} &=& 61900 \\
\hline
(13323\bar{1}\bar{1}1)_{(5,-1)} &=& (\dots z_2 z_1 z_0)_{(5,-1)} &=& 135971 \\
(\bar{1}0001000)_{(5,-1)} &=& (\dots c_2 c_1 c_0)_{(5,-1)} &=& -78000 \\
\hline
(0332\bar{1}\bar{1}\bar{1}1)_{(5,-1)} &=& (\dots z_2 z_1 z_0)_{(5,-1)} &=& 57346 \\
(00010000)_{(5,-1)} &=& (\dots c_2 c_1 c_0)_{(5,-1)} &=& 625 \\
\hline
(0333\bar{1}\bar{1}\bar{1}1)_{(5,-1)} &=& (\dots z_2 z_1 z_0)_{(5,-1)} &=& 57971 \\
(00000000)_{(5,-1)} &=& (\dots c_2 c_1 c_0)_{(5,-1)} &=& 0
\end{array}
$$

The generating function where the coefficient of z^n counts the number of pairs of integers of length $\leq n$, such that $\leq k+2$ iterations are necessary, is given by

$$G^{\leq k}(z) = \frac{s_0(z) + (z/q)^k r_1(z) + (z/q^2)^k r_2(z) + (z^2/q^3)^k r_3(z)}{(1-z)s_0(z) + (z/q)^k s_1(z) + (z/q^2)^k s_2(z) + (z^2/q^3)^k s_3(z)},$$

where $s_0(z) = -2q^4(q-1)(q^2 - z(1-d))(q^2 - z(q+d))$. The terms $r_1(z)$, $r_2(z)$, $r_3(z)$, $s_1(z)$, $s_2(z)$, $s_3(z)$ are polynomials in z, q, d which are independent of k.

From this one can derive, essentially by bootstrapping, approximations, and Mellin transforms, that the expected number t_n of carry propagations satisfies

$$t_n = \log_q n + \log_q \delta + \frac{\gamma}{\log q} + \frac{1}{2} + \psi(\log_q n + \log_q \delta) + O\left(\frac{\log^4 n}{n}\right),$$

where $\delta = \frac{(q^3 + (2d-2)q^2 + (2d-1)(d-1)q - (d-1)d)(q-1)(q+1)}{2(q^2-q-d)(q^2+d-1)}$ and $\psi(x)$ is again a periodic function.

For the instance of the symmetric signed digit expansion, one can also perform such an analysis, but this is more involved, and we refer to the original paper.

11 Mergesort

Because of space restrictions, we cannot describe the algorithm and/or any details. Basically, there is a top–down version where one needs to study functions like

$$(k+1)n-2(2^{k+1}-1)+2\sum_{0\le j\le k}\frac{2^j}{\lfloor\frac{n}{2^{j+1}}\rfloor+2}+2\sum_{0\le j\le k}\frac{(1-b_j)\big(2^j-(b_{j-1}\dots b_0)_2\big)}{\big(\lfloor\frac{n}{2^{j+1}}\rfloor+1\big)\big(\lfloor\frac{n}{2^{j+1}}\rfloor+2\big)},$$

where $(b_k\dots b_0)_2$ is the binary representation of n. This can be efficiently done by a method introduced by Flajolet and Golin [8]. This worked well because such quantities were obtained by so–called divide–and–conquer recursions.

The bottom–up version that I studied with Panny [37] led to even more involved expressions, and no divide–and–conquer recursions were available, so we had to resort to some Delange type analysis. More recently, Hwang and his coworkers [2] developed techniques to deal with quite general versions of Mergesort, including the top–down and bottom–up variants; further mergesort papers of Hwang's can be found on his homepage http://algo.stat.sinica.edu.tw/.

12 Jump interpolation search trees

Güntzer and Paul have used symmetric signed digit representations for the classical case $q=2$ to construct a data structure that they called *jump interpolation search trees* [20]. All numbers with representations of length $\le n$ are in the tree, and the number 0, its length being zero; it serves as the root. Node y is a child of node x, if the least significant nonzero digit in y is replaced by 0, resulting in x. Thus, the depth of a node is the number of nonzero digits. This construction can be *verbatim* translated to the case of general even q. The computations are analogous to the ones in [40]. The admissible words allow the representation

$$\varepsilon+(\tfrac{q}{2}+L+\tfrac{-q}{2}+\bar L)\Big(L\tfrac{q}{2}+0\tfrac{q}{2}+\bar L\tfrac{-q}{2}+0\tfrac{-q}{2}+L+\bar L+0\Big)^*$$

with $L=\{1,\dots,\frac{q}{2}-1\}$ and $\bar L=\{-\frac{q}{2}+1,\dots,-1\}$,

To mark the length of the representations by z and the depth by u, we replace L and $\bar L$ by $(\frac{q}{2}-1)zu$, $\pm\frac{q}{2}$ by zu, and 0 by z, and of course f^* by $1/(1-f)$. The coefficient of z^n then refers to words of length n; since we are interested in words of length $\le n$, we divide the result by $1-z$ and obtain

$$\frac{1+(-1+2u)z-u(uq-2u+2)z^2}{(1-z)(1+(2u-1-uq)z-u(uq-2u+2)z^2)}. \tag{6}$$

From this, we get first, by setting $u=1$, that the number of words of length $\le n$ is given by

$$\frac{q^2}{(q+1)(q-1)}q^n+\frac{q-2}{2(q-1)}-\frac{q}{2(q+1)}(-1)^n. \tag{7}$$

Differentiating (6) with respect to u, and then setting $u = 1$ leads to $\frac{zq(1-z+(q-2)z^2)}{(1-z)(1+z)^2(1-zq)^2}$, and the coefficient of z^n in it is given by

$$\Big[\tfrac{q(q^2-2)}{(q-1)(q+1)^2}n + \tfrac{q(q^2-q+2)}{(q-1)^2(q+1)^3}\Big]q^n + \frac{q(q-2)}{4(1-q)^2} - \Big[\tfrac{q^2}{2(q+1)^2}n + \tfrac{(q^2+3q+6)q}{4(q+1)^3}\Big](-1)^n. \quad (8)$$

Taking the quotient of (8) and (7) gives us then that the average depth of a random node in the tree is

$$\frac{q^2-2}{q(q+1)}n - \frac{q^2-q+2}{q(q+1)^2(q-1)} + o(1);$$

for $q = 2$, we get again the old result $n/3$ from [20].

References

[1] F. Bassino and H. Prodinger. (q, δ)-numeration systems with missing digits. *In preparation*, 2002.

[2] W. M. Chen, H. K. Hwang, and G. H. Chen. The cost distribution of queue-mergesort, optimal mergesorts, and power-of-2 rules. *J. Algorithms*, 30:423–448, 1999.

[3] H. Delange. Sur la fonction sommatoire de la fonction somme des chiffres. *Enseignement Mathématique*, 21:31–47, 1975.

[4] L. Devroye and P. Kruszewski. A note on the Horton–Strahler number for random trees. *Inform. Process. Lett.*, 56:95–99, 1995.

[5] L. Devroye and P. Kruszewski. On the Horton–Strahler number for random tries. *RAIRO Inform. Théor. Appl.*, 30:443–456, 1996.

[6] H. G. Diamond and B. Reznick. Problems and solutions. *American Mathematical Monthly*, 106:175–176, 1999.

[7] G. Doetsch. *Handbuch der Laplace-Transformation, 3 volumes.* Birkhäuser Verlag, Basel, 1972.

[8] P. Flajolet and M. Golin. Mellin transforms and asymptotics. The mergesort recurrence. *Acta Inform.*, 31:673–696, 1994.

[9] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[10] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. Tichy. Mellin transforms and asymptotics: Digital sums. *Theoretical Computer Science*, 123:291–314, 1994.

[11] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[12] P. Flajolet and H. Prodinger. Register allocation for unary–binary trees. *SIAM J. Comput.*, 15:629–640, 1986.

[13] P. Flajolet and L. Ramshaw. A note on Gray code and odd–even merge. *SIAM J. Comput.*, 9(1):142–158, 1980.

[14] P. Flajolet, J.-C. Raoult, and J. Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoretical Computer Science*, 9:99–125, 1979.

[15] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, 144:101–124, 1995.

[16] P. Grabner, C. Heuberger, and H. Prodinger. Subblock occurrences in signed digit representations. *Submitted*.

[17] P. Grabner, P. Kirschenhofer, and H. Prodinger. The sum–of–digits function for complex bases. *J. London Math. Soc.*, 57:20–40, 1998.

[18] P. Grabner and H. Prodinger. Asymptotic analysis of the moments of the Cantor distribution. *Statistics and Probability Letters*, 26:243–248, 1996.

[19] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics. A foundation for computer science*. Addison-Wesley, second edition, 1994.

[20] U. Güntzer and M. Paul. Jump interpolation search trees and symmetric binary numbers. *Information Processing Letters*, 26:193–204, 1987/88.

[21] C. Heuberger and H. Prodinger. Carry propagation in signed digit representations. *submitted*, 2001.

[22] C. Heuberger and H. Prodinger. On minimal expansions in redundant number systems: Algorithms and quantitative analysis. *Computing*, 66:377–393, 2001.

[23] J. R. M. Hosking. Moments of order statistics of the Cantor distribution. *Statistics and Probability Letters*, 19:161–165, 1994.

[24] P. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. *Theoretical Computer Science*, 201:1–62, 1998.

[25] R. Kemp. The average number of registers needed to evaluate a binary tree optimally. *Acta Inform.*, 11:363–372, 1978/79.

[26] P. Kirschenhofer and H. Prodinger. Subblock occurrences in positional number systems and Gray code representation. *Journal of Information and Optimization Sciences*, 5:29–42, 1984.

[27] A. Knopfmacher and H. Prodinger. Exact and asymptotic formulæ for average values of order statistics of the Cantor distribution. *Statistics and Probability Letters*, 27:189–194, 1996.

[28] D. E. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973. Second edition, 1998.

[29] D. E. Knuth. The average time for carry propagation. *Indagationes Mathematicæ*, 40:238–242, 1978.

[30] P. Kruszewski. A note on the Horton–Strahler number for random binary search trees. *Inform. Process. Lett.*, 69:47–51, 1999.

[31] F. R. Lad and W. F. C. Taylor. The moments of the Cantor distribution. *Statistics and Probability Letters*, 13:307–310, 1992.

[32] A. Meir, J. W. Moon, and J. R. Pounder. On the order of random channel networks. *SIAM J. Algebraic Discrete Methods*, 1:25–33, 1980.

[33] J. W. Moon. An extension of Horton's law of stream numbers. *Math. Colloq. Univ. Cape Town*, 12:1–15 (1980), 1978/79.

[34] J. W. Moon. On Horton's law for random channel networks. *Ann. Discrete Math.*, 8:117–121, 1980. Combinatorics 79 (Proc. Colloq., Univ. Montréal, Montreal, Que., 1979), Part I.

[35] M. Nebel. On the Horton–Strahler number for combinatorial tries. *Theor. Inform. Appl.*, 34:279–296, 2000.

[36] A. H. Osbaldestin and P. Shiu. A correlated digital sum problem associated with sums of three squares. *Bull. London Math. Soc.*, 21:369–374, 1989.

[37] W. Panny and H. Prodinger. Bottom-up mergesort—a detailed analysis. *Algorithmica*, 14:340–354, 1995.

[38] H. Prodinger. Some analytic techniques for the investigation of the asymptotic behaviour of tree parameters. *EATCS Bulletin*, 47:180–199, 1992.

[39] H. Prodinger. On a problem of Yekutieli and Mandelbrot about the bifurcation ratio of binary trees. *Theoretical Computer Science*, 181:181–194, 1997.

[40] H. Prodinger. On binary representations of integers with digits -1, 0, 1. *Integers*, pages A8, 14 pp. (electronic), 2000.

[41] H. Prodinger. On Cantor's singular moments. *Southwest J. Pure Appl. Math.*, 2000(1):27–29 (electronic), 2000.

[42] G. Reitwiesner. Binary arithmetic. *Vol. 1 of Advances in Computers, Academic Press*, pages 231–308, 1960.

[43] M. Riedel. Applications of the Mellin–Perron formula in number theory. Technical report, University of Toronto, 1996.

[44] R. Sedgewick. Data movement in odd–even merging. *SIAM J. Comput.*, 7:239–272, 1978.

[45] L. A. Shepp and S. P. Lloyd. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.*, 121:340–357, 1966.

[46] P. Shiu. Counting sums of three squares. *Bull. London Math. Soc.*, 20:203–208, 1988.

[47] J. Thuswaldner. Summatory functions of digital sums occurring in Cryptography. *Periodica Math. Hungarica*, 38:111–130, 1999.

[48] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis.* Cambridge University Press, fourth edition, 1927. Reprinted 1973.

[49] I. Yekutieli and B. Mandelbrot. Horton–Strahler ordering of random binary trees. *J. Phys. A*, 27:285–293, 1994.

Helmut Prodinger
The John Knopfmacher Centre for Applicable Analysis and Number Theory
School of Mathematics, University of the Witwatersrand
P. O. Wits, 2050 Johannesburg, South Africa
`helmut@staff.ms.wits.ac.za`
`http://www.wits.ac.za/helmut/index.htm`