(q,δ)-NUMERATION SYSTEMS WITH MISSING DIGITS

FRÉDÉRIQUE BASSINO† AND HELMUT PRODINGER†

Abstract. We consider the (q,δ) numeration system, with basis $q \geq 2$ and the set of digits $\{\delta, \delta+1, \dots, q+\delta-1\}$ where $-(q-1) \leq \delta \leq 0$. We study properties of numbers where some digits *do not occur*. This is analogous to the Cantor set $\{0.a_1a_2\cdots \mid a_i \in \{0,2\}\}$.

We compute an asymptotic equivalent of the nth moment of the "Cantor (q,D)–distribution" which can be described as the numbers $0.w_1w_2\dots$ with $w_i \in D \subseteq \{\delta,\dots,q+\delta-1\}$, and each such letter can occur with the same probability $1/\mathit{Card} D$.

Furthermore, we consider n random strings according to this distribution and the expected minimum of them. We find a recursion which we solve asymptotically.

1. Introduction

We consider the (q,δ) numeration system, with basis $q \geq 2$ and the set of digits $\{\delta, \delta+1, \dots, q+\delta-1\}$ where $-(q-1) \leq \delta \leq 0$.[1] Every real number x has an essentially unique[2] representation

$$x = \sum_{k \leq n} a_k q^k$$

with $a_k \in \{\delta, \delta+1, \dots, q+\delta-1\}$. In particular, we are interested in properties of numbers where some digits *do not occur*. This is analogous to the Cantor set, which can be described as

$$\big\{0.a_1a_2\cdots \mid a_i \in \{0,2\}\big\}.$$

The Cantor distribution with parameter ϑ, $0 < \vartheta \leq \frac{1}{2}$, was introduced in [11] by the random series

$$\frac{\bar{\vartheta}}{\vartheta} \sum_{i \geq 1} X_i \vartheta^i \ ,$$

where the X_i are independent with the distribution

$$\mathbb{P}\{X_i = 0\} = \mathbb{P}\{X_i = 1\} = \frac{1}{2} \ ,$$

Date: November 18, 2002.

1991 *Mathematics Subject Classification.* Primary: 11A63, 11K16; Secondary: 60C05.

Key words and phrases. Digits, Depoissonization, Mellin Transform, Asymptotic Expansion, Moments.

†This author was supported by the CNRS/NRF-project n^o10959.

Part of this work was done during the first author's visit to the John Knopfmacher Centre for Applicable Analysis and Number Theory at the University of the Witwatersrand, Johannesburg, South Africa.

[1]Often, the letter d is used instead of δ. In this paper, however, we need the letter d for other purposes.

[2]Some numbers have two representations, which is the analogue of $1 = 0.999\dots$.

and $\bar{\vartheta} = 1 - \vartheta$. The name stems from the instance $\vartheta = \frac{1}{3}$, since then exactly those numbers from the interval $[0, 1]$ appear that have a ternary expansion solely consisting of the digits 0 and 2.

The moments of this distribution where considered in [11], and, more recently in [5], where an asymptotic formula for the nth moment was derived using a combination of analytic techniques, notably depoissonization ("de-Poissonization") and Mellin transforms.

In the first part of the present paper we derive analogous results: Let D be a (given) subset of the set of digits $\{\delta, \delta+1, \dots, q+\delta-1\}$; we set $\mathsf{d} = Card D$ and $D = \{d_1 < d_2 < \ldots < d_{\mathsf{d}}\}$.

We consider infinite (random) words $w_1 w_2 \ldots$ over the alphabet $D = \{d_1, d_2, \dots, d_{\mathsf{d}}\}$ and a mapping value, defined by

$$\mathsf{value}(w_1 w_2 \ldots) = \sum_{i \ge 1} w_i q^{-i} .$$

Each letter can appear with probability $\frac{1}{\mathsf{d}}$.

In this way we obtain a probability distribution on the interval $[\delta/(q-1), \delta/(q-1)+1]$, which we will call the Cantor–(q, D) distribution. In Section 2 we study its moments.

Another interesting topic related to the Cantor distribution was introduced in [6]; one assumes that n (independent) elements are drawn according to the Cantor distribution. One is interested in the expected value of the *minimum* of them. Hosking gave a recursion for these expectations, which was eventually solved in [9], both exactly and asymptotically. For the exact solution (involving Bernoulli numbers) a neat trick of Knuth's was essential; for the asymptotics one could then rely on *Rice's method* [3].

In Section 4 we are going to solve the analogous question in our model of the (q, δ)–system with allowed set of digits D. Unfortunately, the nice trick does no longer work in this more general case, and we thus have to use the technique of depoissonization; for more details about this technique, one can refer to [7] and [13]; the present approach is modelled after the analysis in [8], which is also covered in [7] and [13].

2. The Moments

Observe the recursion formula, valid for all $i \in \{1, \dots, \mathsf{d}\}$

$$\mathsf{value}(d_i w) = d_i \cdot q^{-1} + q^{-1} \cdot \mathsf{value}(w) . \tag{2.1}$$

Here, dw is the concatentation of the digit d and the infinite string w; denote the set of all infinite strings by $\mathcal{W}$. From the self–similarity of the measure, we can derive a recursion for the moments M_n as follows:

$$\begin{aligned} M_n &= \sum_{w \in \mathcal{W}} \big(\mathsf{value}(w)\big)^n \\ &= \frac{1}{\mathsf{d} q^n} \sum_{i=0}^{\mathsf{d}} \sum_{w \in \mathcal{W}} \big(d_i + \mathsf{value}(w)\big)^n \\ &= \frac{1}{\mathsf{d} q^n} \sum_{i=0}^{\mathsf{d}} \sum_{j=0}^{n} \binom{n}{j} d_i^{n-j} M_j. \end{aligned}$$

It can be made explicit by isolating the term M_n:

$$M_n = \frac{\mathsf{d}^{-1}q^{-n}}{1-q^{-n}} \sum_{i=1}^{\mathsf{d}} \sum_{j=0}^{n-1} d_i^{n-j} \binom{n}{j} M_j.$$

This recursion could be used to compute a list of the first few moments.

Theorem 1. *The moments of the Cantor–(q,D) distribution satisfy the following recursion: $M_0 = 1$ and for $n \geq 1$*

$$M_n = \frac{1}{\mathsf{d}(q^n-1)} \sum_{j=0}^{n-1} \sum_{i=1}^{\mathsf{d}} d_i^{n-j} \binom{n}{j} M_j.$$

For instance

$$\begin{aligned} M_1 &= \frac{1}{\mathsf{d}(q-1)} \sum_{i=1}^{\mathsf{d}} d_i, \\ M_2 &= \frac{1}{\mathsf{d}(q^2-1)} \sum_{i=1}^{\mathsf{d}} d_i^2 + \frac{1}{\mathsf{d}^2(q-1)^2(q+1)} \left(\sum_{i=1}^{\mathsf{d}} d_i \right)^2, \qquad (2.2) \\ \mathsf{Variance} = M_2 - M_1^2 &= \frac{1}{\mathsf{d}(q^2-1)} \sum_{i=1}^{\mathsf{d}} d_i^2 - \frac{q}{\mathsf{d}^2(q-1)^2(q+1)} \left(\sum_{i=1}^{\mathsf{d}} d_i \right)^2. \end{aligned}$$

3. The Asymptotic Behaviour of the Moments

The next problem is to investigate the asymptotic behaviour of the moments M_n, as $n \to \infty$. Remember that this investigation for the classical case was done in [5].

A rough estimation shows us that the moments decrease exponentially. Indeed, if we set $M_n \approx \lambda^{-n}$, we might infer that $\lambda = (q-1)/d_M$, where d_M is the digit of maximal modulus in D.

First, we assume that there is only one digit of maximal modulus; without loss of generality we may further assume that it is positive, since otherwise we would have simply to multiply the moments by $(-1)^n$ and work with the set of digits $-D$ instead.

We set

$$m_n := M_n \cdot \lambda^n$$

and show that this sequence has nicer properties. It satisfies the *modified recurrence*

$$m_n = \frac{1}{\mathsf{d}(q^n-1)} \sum_{i=1}^{\mathsf{d}} \sum_{j=0}^{n-1} \binom{n}{j} (\lambda d_i)^{n-j} m_j \ .$$

To study this sequence further, we rewrite it as

$$m_n \cdot \mathsf{d}(q^n-1) = \sum_{i=1}^{\mathsf{d}} \left(\sum_{j=0}^{n} \binom{n}{j} (\lambda d_i)^{n-j} m_j - m_n \right)$$

or

$$m_n = \frac{1}{\mathsf{d}q^n} \sum_{i=1}^{\mathsf{d}} \sum_{j=0}^{n} \binom{n}{j} (\lambda d_i)^{n-j} m_j$$

and note that this holds for all $n \geq 0$. Then we introduce the *exponential generating function*

$$m(z) = \sum_{n\geq 0} m_n \frac{z^n}{n!}$$

and get

$$m(z) = \frac{1}{\mathsf{d}} \sum_{i=1}^{\mathsf{d}} e^{d_i \lambda \frac{z}{q}}\, m\Big(\frac{z}{q}\Big) = \prod_{k\geq 1} \frac{\sum_{i=1}^{\mathsf{d}} e^{z d_i \lambda / q^k}}{\mathsf{d}}.$$

As in [5], we have to consider the *Poisson transformed function* $\widehat{m}(z) = e^{-z} m(z)$, which satisfies the functional equation

$$\widehat{m}(z) = \frac{1}{\mathsf{d}} \sum_{i=1}^{\mathsf{d}} e^{\frac{z}{q}(d_i\lambda + 1 - q)}\, \widehat{m}\Big(\frac{z}{q}\Big)\ . \tag{3.1}$$

This functional equation (3.1) can also be solved by iteration:

$$\widehat{m}(z) = \prod_{k\geq 1} \frac{\sum_{i=1}^{\mathsf{d}} e^{z(d_i\lambda+1-q)/q^k}}{\mathsf{d}}.$$

However, this product is not too useful, and we have to go back to the functional equation.

The next step is to consider the behaviour of $\widehat{m}(z)$ for $z \to \infty$. The reason is that $m_n \sim \widehat{m}(n)$. The justification for this is a technique called *depoissonization.*

The general references for that are [7] and [13]. We follow [5], where the technique is explained in more detail. The idea is roughly as follows: One extracts the coefficients m_n of $m(z)$ via Cauchy's integral formula taking a circle of radius n and uses Stirling's formula for the approximation of the quantity $n!/n^n$ which occurs.

It is suggestive to use a new name $R(z)$ for $\frac{1}{\mathsf{d}} \sum_{i\neq M} e^{\frac{z}{q}(d_i\lambda+1-q)} \widehat{m}\big(\frac{z}{q}\big)$ and consider it to be an auxiliary (and *known*) function;

$$\widehat{m}(z) = \frac{1}{\mathsf{d}} \widehat{m}\Big(\frac{z}{q}\Big) + R(z)\ . \tag{3.2}$$

We compute the *Mellin transform* of (3.2) (see [2] for definitions and properties);

$$\widehat{m}^*(s) = \frac{q^s}{\mathsf{d}} \widehat{m}^*(s) + R^*(s) = \frac{R^*(s)}{1 - \frac{q^s}{\mathsf{d}}}\ .$$

The function $m(z)$ can be recovered from this by *Mellin's inversion formula*,

$$\widehat{m}(z) = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \frac{R^*(s)}{1 - \frac{q^s}{\mathsf{d}}}\, z^{-s}\, ds\ ,$$

where $0 < c < \log_q \mathsf{d}$. By shifting the integral to the right and taking the *negative residues* into account, we get the desired asymptotic behaviour of $\widehat{m}(z)$. There are simple poles at $s = \log_q \mathsf{d} + \frac{2k\pi \mathrm{i}}{\log q}$, $k \in \mathbb{Z}$. The negative residue there is

$$\frac{1}{\log q} R^*\Big(\log_q \mathsf{d} + \frac{2k\pi\mathrm{i}}{\log q}\Big)\, z^{-\log_q \mathsf{d} - \frac{2k\pi\mathrm{i}}{\log q}}\ .$$

The value for $k=0$ is of special interest; it is, to make it more explicit,

$$\frac{1}{\log q} z^{-\log_q \mathsf{d}} \int_0^\infty \frac{1}{\mathsf{d}} \sum_{i\neq M} e^{\frac{z}{q}(d_i\lambda+1-q)}\ \widehat{m}\Big(\frac{z}{q}\Big)\ z^{\log_q \mathsf{d}-1}\, dz\ .$$

Traditionally, one collects all the terms into a periodic function.

Theorem 2. *The nth moment M_n of the Cantor–(q,D) distribution has for $n\to\infty$ the following asymptotic behaviour*

$$M_n = \Big(\frac{d_M}{q-1}\Big)^n \Phi(-\log_q n)\, n^{-\log_q \mathsf{d}} \Big(1+O\Big(\frac{1}{n}\Big)\Big)\ ,$$

where $\Phi(x)$ is a periodic function with period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) is given by

$$\frac{1}{\log q} \int_0^\infty \frac{1}{\mathsf{d}} \sum_{i\neq M} e^{\frac{z}{q}(d_i\lambda+1-q)}\ \widehat{m}\Big(\frac{z}{q}\Big)\ z^{\log_q \mathsf{d}-1}\, dz\ .$$

This integral can be computed numerically by replacing $\widehat{m}\big(\frac{z}{q}\big)$ by the first few values of its Taylor expansion, which can be obtained by iterating the recurrence for the numbers m_n.

Example. We consider $q=5$ and $D=\{-1,1,3\}$, so $\mathsf{d}=3$, $d_M=3$, and $\lambda=\frac{4}{3}$. Then

$$m_n = \frac{1}{3(5^n-1)} \sum_{j=0}^{n-1} \binom{n}{j} \Big((-\tfrac{4}{3})^{n-j} + (\tfrac{4}{3})^{n-j} + 4^{n-j}\Big) m_j\ .$$

So we replace $R(z)$ by

$$R^{\text{approx}} = \frac{1}{3}\Big(e^{-16z/15}+e^{-8z/15}\Big)\ e^{-z/5}\ \Big(1+\tfrac{1}{20}z+\tfrac{1}{288}z^2+\tfrac{19}{144000}z^3+\tfrac{5887}{1347840000}z^4+\dots\Big)$$

and compute

$$\frac{1}{\log 5}\int_0^\infty R^{\text{approx}}\ z^{\log 3/\log 5-1}\ dz = 0.59896\dots\ .$$

We find $M_{100}\ (\frac{4}{3})^{100}\ 100^{\log 3/\log 5} \approx 0.60351$. The reason that this works is as follows: The radius of convergence of $m(z)$ is infinity, so we replace it by its Taylor series. It is multiplied by terms of the form e^{-z} and integrated from 0 to ∞; to have any degree of accuracy, we can integrate from 0 to K, say, and throw in enough terms of the Taylor series. Maple evaluates the integral as a finite sum of Gamma functions.

The case when $-d_1 = d_{\mathsf{d}}$ requires special care. If one has e. g. a symmetric set of digits D, then all odd moments vanish. A similar phenomenon occurs if the largest positive and negative digits coincide. Depoissonization, as it is described in [7] and in [13], does not cover this instance. But one could just extract coefficients $n![z^n]m(z)$ via Cauchy's integral formula using as the path of integration a circle of radius n. We have

$$n![z^n]m(z) = \frac{1}{2\pi\mathrm{i}}\int_{|z|=n} \frac{m(z)}{z^{n+1}}dz.$$

After a separation of the integral between positive and negative half plane and a change of variable in the second term of the sum, we get

$$n![z^n]m(z) = \frac{n!}{2\pi\mathrm{i}}\int_{|z|=n,\Re z\geq 0}\frac{m(z)}{z^{n+1}}dz + (-1)^n\frac{1}{2\pi\mathrm{i}}\int_{|z|=n,\Re z\geq 0}\frac{m(-z)}{z^{n+1}}dz.$$

We now use the fact that the saddle point in the integral

$$\frac{n!}{2\pi i}\int_{\mathcal{C}}\frac{e^z}{z^{n+1}}dz = 1$$

lies at $z = n + O(1)$. For more details about the saddle point method, one can refer to [1] or to [10]. As $m(z) = e^z(e^{-z}m(z))$ where

$$e^{-z}m(z) = \prod_{k\geq 1}\frac{\sum_{i=1}^{\mathsf{d}} e^{z(d_i\lambda+1-q)/q^k}}{\mathsf{d}}$$

is bounded, the previous saddle point is asymptotically not affected by multiplying e^z by this infinite product. Thus we get

$$\frac{n!}{2\pi\mathrm{i}}\int_{|z|=n,\Re z\geq 0}\frac{m(z)}{z^{n+1}}dz \sim e^{-n}m(n)\Big(1+O\Big(\frac{1}{n}\Big)\Big),$$

and similarly

$$\frac{n!}{2\pi\mathrm{i}}\int_{|z|=n,\Re z\geq 0}\frac{m(-z)}{z^{n+1}}dz \sim e^{-n}m(-n)\Big(1+O\Big(\frac{1}{n}\Big)\Big).$$

Compare [4] for such an approach.

We next study both terms of the sum by making use of the Mellin transform as in the previous instance.

Theorem 3. *In the instance of two dominant digits d_1, d_{d}, with $-d_1 = d_{\mathsf{d}} = d_M$, the nth moment M_n of the Cantor–(q,D) distribution has for $n\to\infty$ the following asymptotic behaviour*

$$M_n = \Big(\frac{d_M}{q-1}\Big)^n\Big[\Phi_1(-\log_q n) + (-1)^n\Phi_2(-\log_q n)\Big]\, n^{-\log_q \mathsf{d}}\Big(1+O\Big(\frac{1}{n}\Big)\Big)\ ,$$

where $\Phi_1(x)$ and $\Phi_2(x)$ are periodic functions with period 1 and known Fourier coefficients. The means (zeroth Fourier coefficients) are given by

$$\frac{1}{\log q}\int_0^\infty \frac{1}{\mathsf{d}}\sum_{i\neq d} e^{\frac{z}{q}(d_i\lambda+1-q)}\ e^{-z/q}\ m\Big(\frac{z}{q}\Big)\ z^{\log_q \mathsf{d}-1}\,dz,$$
$$\frac{1}{\log q}\int_0^\infty \frac{1}{\mathsf{d}}\sum_{i\neq 1} e^{\frac{z}{q}(-d_i\lambda+1-q)}\ e^{-z/q}\ m\Big(\frac{-z}{q}\Big)\ z^{\log_q \mathsf{d}-1}\,dz,$$

respectively.

While the even moments are always positive, the sign of the odd ones depends on the largest (in modulus) digit $d_i\in D$ such that not both d_i and $-d_i$ are in D.

Example. We consider $q = 7$ and $D = \{-3, 2, 3\}$, so $\mathsf{d} = 3$, $d_M = 3$, and $\lambda = 2$. Then

$$m_n = \frac{1}{3(7^n - 1)} \sum_{j=0}^{n-1} \binom{n}{j} \Big((-6)^{n-j} + 4^{n-j} + 6^{n-j} \Big) m_j \ .$$

So we replace $R(z)$ by

$$R^{\text{approx}} = \frac{1}{3}\Big(e^{-12z/7} + e^{-2z/7}\Big)\ e^{-z/7}\ \Big(1 + \tfrac{1}{63}z + \tfrac{101}{63504}z^2 + \tfrac{251}{32577552}z^3 + \dots\Big)$$

and compute

$$\frac{1}{\log 7}\int_0^\infty R^{\text{approx}}\ z^{\log 3/\log 7 - 1}\ dz = 0.63967\dots\ .$$

This is the first contribution; call it C_1. Now we do the same for the set of digits $-D = \{-3, -2, 3\}$. Then we use

$$R^{\text{approx}} = \frac{1}{3}\Big(e^{-12z/7} + e^{-10z/7}\Big)\ e^{-z/7}\ \Big(1 - \tfrac{1}{63}z + \tfrac{101}{63504}z^2 - \tfrac{251}{32577552}z^3 + \dots\Big)$$

and find

$$\frac{1}{\log 7}\int_0^\infty R^{\text{approx}}\ z^{\log 3/\log 7 - 1}\ dz = 0.39769\dots\ .$$

This is the second contribution; call it C_2.

We find $M_{100}\ 2^{100}\ 100^{\log 3/\log 7} \approx 1.04057$; this is close to $C_1 + C_2 = 1.03737$. On the other hand, $M_{99}\ 2^{99}\ 99^{\log 3/\log 7} \approx 0.26770$; this is to be compared with $C_1 - C_2 = 0.24197$.

4. Expected value of the minimum order statistic of the (q, D)–distribution

We consider *random strings* $c_1\, c_2\, c_3 \dots$ where the c_i's $\in D = \{d_1 < d_2 < \dots < d_\mathsf{d}\}$ are equally likely. We then consider the random variable value that maps $c_1\, c_2\, c_3 \dots$ to the real number

$$\mathsf{value}(c_1\, c_2\, c_3 \dots) = \sum_{i\ge 1} c_i q^{-i}\ \in \Big[\frac{d_1}{q-1}, \frac{d_\mathsf{d}}{q-1}\Big]\ .$$

The strings now have a natural order from the usual ordering of the real numbers. This is easily seen to be equivalent to the lexicographic ordering of strings, i.e. $c_1\, c_2\, c_3 \dots < c_1'\, c_2'\, c_3' \dots$ iff there is a k such that $c_i = c_i'$ for $i = 1, \dots, k-1$ and $c_k < c_k'$. It thus makes sense to speak of order statistics for strings. Suppose that n independent random strings $w_1, \dots, w_n$ are produced. We denote by a_n the average value of the minimum of the n real numbers $\mathsf{value}(w_1), \dots, \mathsf{value}(w_n)$. We derive the following recursion for the expected minimum

$$a_n = \frac{1}{q\mathsf{d}^n}\Bigg[\sum_{i=1}^{\mathsf{d}-1}\Bigg(\sum_{k=1}^{n}\binom{n}{k}(\mathsf{d}-i)^{n-k}(d_i + a_k)\Bigg) + d_\mathsf{d} + a_n\Bigg]\ , \quad n \ge 1\ ,$$

or

$$(\mathsf{d}^n q - \mathsf{d})a_n = \sum_{i=1}^{\mathsf{d}}\Big[(\mathsf{d}-i+1)^n - (\mathsf{d}-i)^n\Big]d_i + \sum_{k=1}^{n-1}\sum_{i=1}^{\mathsf{d}-1}\binom{n}{k}(\mathsf{d}-i)^{n-k}a_k\ , \quad n \ge 1\ .$$

This recursion is obtained by considering the smallest digit d_i that at least one of the n random strings has in its first position. The minimum value will be one of these, and be determined recursively; the first position adds the quantity $\frac{d_i}{q}$ to the recursively determined minimum.

Now, if n is large, it is almost certain that there is a string starting with $d_1d_1d_1\dots$ among the n random strings, producing the minimal value (in the limit) $\frac{d_1}{q-1}$. Remember that in the classical Cantor case $d_1 = 0$, and the question was to analyze how fast a_n approaches zero. In order to obtain meaningful results, we define $\alpha_n := a_n - \frac{d_1}{q-1}$ and rewrite the recursion:

$$(\mathsf{d}^n q - \mathsf{d})\Big(\alpha_n + \frac{d_1}{q-1}\Big) = \sum_{i=1}^{\mathsf{d}} \Big[(\mathsf{d}-i+1)^n - (\mathsf{d}-i)^n\Big] d_i + \sum_{k=1}^{n-1}\sum_{i=1}^{\mathsf{d}-1}\binom{n}{k}(\mathsf{d}-i)^{n-k}\Big(\alpha_k + \frac{d_1}{q-1}\Big)$$

or

$$(\mathsf{d}^n q - \mathsf{d})\alpha_n = -(\mathsf{d}-1)^n d_1 + \sum_{i=2}^{\mathsf{d}} \Big[(\mathsf{d}-i+1)^n - (\mathsf{d}-i)^n\Big] d_i + \sum_{k=1}^{n-1}\sum_{i=1}^{\mathsf{d}-1}\binom{n}{k}(\mathsf{d}-i)^{n-k}\alpha_k.$$

Then we introduce the *exponential generating function* (with $\alpha_0 := 0$)

$$a(z) = \sum_{n\ge 0} \alpha_n \frac{z^n}{n!}$$

and get

$$q a(\mathsf{d}z) - \mathsf{d}a(z) = (1 - e^{(\mathsf{d}-1)z})d_1 + \sum_{i=2}^{\mathsf{d}} \Big(e^{(\mathsf{d}-i+1)z} - e^{(\mathsf{d}-i)z}\Big) d_i + \sum_{i=1}^{\mathsf{d}-1}\Big(e^{(\mathsf{d}-i)z} - 1\Big)a(z),$$

or

$$a(\mathsf{d}z) = \frac{1}{q}\Big(\frac{1-e^{\mathsf{d}z}}{1-e^z}\Big)a(z) + \frac{d_1}{q}(1 - e^{(\mathsf{d}-1)z}) + \frac{1}{q}\sum_{i=1}^{\mathsf{d}}\Big(e^{(\mathsf{d}-i+1)z} - e^{(\mathsf{d}-i)z}\Big)d_i.$$

We have to consider the *Poisson transformed function* $\widehat{a}(z) = e^{-z}a(z)$, which satisfies the functional equation

$$\widehat{a}(\mathsf{d}z) = \frac{1}{q}\Big(\frac{1-e^{-\mathsf{d}z}}{1-e^{-z}}\Big)\widehat{a}(z) + \frac{d_1}{q}(e^{-\mathsf{d}z} - e^{-z}) + \frac{1}{q}(e^z - 1)\sum_{i=2}^{\mathsf{d}} e^{-iz}d_i\ .$$

The next step is to consider the behaviour of $\widehat{a}(z)$ for $z \to \infty$. The reason is that $\alpha_n \sim \widehat{a}(n)$. The justification for this is again the technique of *depoissonization.* We set

$$b(z) = \frac{1-e^{-\mathsf{d}z}}{1-e^{-z}} \qquad \text{and} \qquad \phi(z) = \prod_{j=0}^{\infty} b(z\mathsf{d}^j) = \frac{1}{1-e^{-z}}$$

and

$$c(z) = \frac{d_1}{q}(e^{-z} - e^{-z/\mathsf{d}}) + \frac{1}{q}(e^{z/\mathsf{d}} - 1)\sum_{i=2}^{\mathsf{d}} e^{-iz/\mathsf{d}}d_i.$$

We then get

$$\widehat{a}(z) = \sum_{n=0}^{\infty} q^{-n}\, c(q^{-n}z)\prod_{k=1}^{n} b(q^{-k}z).$$

As $\widehat{a}(0) = 0$, we finally obtain

$$\widehat{a}(z)\phi(z) = \sum_{n=0}^{\infty} q^{-n}\, c(\mathsf{d}^{-n}z)\, \phi(\mathsf{d}^{-n}z). \tag{4.1}$$

We compute the *Mellin transform* of (4.1); since it is a harmonic sum (see [2] for more background), we obtain

$$\big(\widehat{a}(z)\phi(z)\big)^*(s) = \sum_{n\ge 0} q^{-n}\, \mathsf{d}^{ns}\, \big(c(z)\phi(z)\big)^*(s) = \frac{1}{1-\frac{\mathsf{d}^s}{q}} \big(c(z)\phi(z)\big)^*(s).$$

(The Mellin transform

$$\big(c(z)\phi(z)\big)^*(s) = \int_0^\infty \left[\frac{d_1}{q}(e^{-z} - e^{-z/\mathsf{d}}) + \frac{1}{q}(e^{z/\mathsf{d}} - 1)\sum_{i=2}^{\mathsf{d}} e^{-iz/\mathsf{d}} d_i\right] \frac{e^z}{e^z - 1} z^{s-1} dz$$

can be expressed by Hurwitz' zeta functions: Recall that the Hurwitz zeta function is defined as

$$\zeta(s,a) = \sum_{k=0}^{\infty} \frac{1}{(k+a)^s}$$

for $\Re(s) > 1$ and $a \ge 0$. The classical formula

$$\Gamma(s)\zeta(s,a) = \int_0^\infty z^{s-1} \frac{e^{-az}}{1-e^{-z}} dz$$

eventually gives us

$$\big(c(z)\phi(z)\big)^* = \Gamma(s)\bigg(\frac{d_1}{q}\Big(\zeta\big(s,1\big) - \zeta\big(s,\frac{1}{\mathsf{d}}\big)\Big) + \frac{1}{q}\sum_{i=2}^{\mathsf{d}} \Big(\zeta\big(s,\frac{i-1}{\mathsf{d}}\big) - \zeta\big(s,\frac{i}{\mathsf{d}}\big)\Big) d_i\bigg).$$

This can be used for numerical calculations.)

The function $\widehat{a}(z)$ can be recovered from this by *Mellin's inversion formula*,

$$\widehat{a}(z)\phi(z) = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \frac{\big(c(z)\phi(z)\big)^*(s)}{1-\frac{\mathsf{d}^s}{q}}\, z^{-s}\, ds\ ,$$

where $0 < c < \log_{\mathsf{d}} q$. By shifting the integral to the right and taking the *negative residues* into account, we get the desired asymptotic behaviour of $\widehat{a}(z)$. There are simple poles at $s = \log_{\mathsf{d}} q + \frac{2k\pi \mathrm{i}}{\log \mathsf{d}}$, $k \in \mathbb{Z}$. The negative residue there is

$$\frac{1}{\log \mathsf{d}} \big(c(z)\phi(z)\big)^* \Big(\log_{\mathsf{d}} q + \frac{2k\pi\mathrm{i}}{\log \mathsf{d}}\Big)\, z^{-\log_{\mathsf{d}} q - \frac{2k\pi \mathrm{i}}{\log \mathsf{d}}}.$$

The value for $k = 0$ is of special interest; it is, to make it more explicit,

$$\frac{1}{\log \mathsf{d}} z^{-\log_{\mathsf{d}} q} \int_0^\infty c(z)\phi(z)\, z^{\log_{\mathsf{d}} q - 1}\, dz.$$

Moreover $\phi(z) \sim 1$ as $z \to \infty$. One collects all the terms into a periodic function.

Theorem 4. *The expected value of the minimum order statistics of the Cantor–(q, D) distribution has for $n \to \infty$ the following asymptotic behaviour*

$$a_n = \frac{d_1}{q-1} + \Phi(-\log_{\mathsf{d}} q)\, n^{-\log_{\mathsf{d}} q}\left(1 + O\Big(\frac{1}{n}\Big)\right),$$

where $\Phi(x)$ is a periodic function with period 1 and known Fourier coefficients. The mean *(zeroth Fourier coefficient) is given by*

$$\frac{1}{\log \mathsf{d}} \int_0^\infty c(z)\phi(z)\, z^{\log_{\mathsf{d}} q - 1}\, dz.$$

As before, the integral can be computed numerically.

Example. We consider again the example $q = 5$ with $D = \{-1, 1, 3\}$ and $\mathsf{d} = 3$. Then $a_{100} + \frac{1}{4} \approx 0.00205441$. Further,

$$c(z) = \frac{1}{5}\big(-4e^{-z} + 2e^{-z/3} + 2e^{-2z/3} \big),$$

and

$$\frac{1}{\log 3} \int_0^\infty c(z)\phi(z)\, z^{\log 5/\log 3 - 1}\, dz \approx 1.77099.$$

Eventually, $1.77099 \cdot 100^{\log 5/\log 3} \approx 0.00208$.

References

[1] N. G. de Bruijn. *Asymptotic Method in Analysis.* North Holland, Amsterdam, The Netherlands, 1961.
[2] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoret. Comput. Sci.*, 144:3–58, 1995.
[3] P. Flajolet and R. Sedgewick. Mellin Transforms and Asymptotics : Finite Differences and Rice's Integrals. *Theoret. Comput. Sci.*, 144:101–124, 1995.
[4] P. Grabner. Searching for losers. *Random Structures and Algorithms*, 4:99–110, 1993.
[5] P. Grabner and H. Prodinger. Asymptotic analysis of the moments of the Cantor distribution. *Statistics and Probability Letters*, 26:243–248, 1996.
[6] J. R. M. Hosking. Moments of order statistics of the Cantor distribution. *Statistics and Probability Letters*, 19:161–165, 1994.
[7] P. Jacquet and W. Szpankowski. Analytical depoissonization and its applications. *Theoret. Comput. Sci.*, 201:1–62, 1998.
[8] P. Kirschenhofer, H. Prodinger and W. Szpankowski. Analysis of a splitting process arising in probabilistic counting and other related algorithms. *Random Structures and Algorithms*, 9:379–401, 1996.
[9] 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.
[10] A. M. Odlyzko. Asymptotic enumeration methods. In *Handbook of Combinatorics*, vol. 2, 1063–1229, R. L. Graham, M. Groetschel, and L. Lovasz, eds., Elsevier, 1995.
[11] F. R. Lad and W. F. C. Taylor. The moments of the Cantor distribution. *Statistics and Probability Letters*, 13:307–310, 1992.
[12] H. Prodinger. The Cantor-Fibonacci distribution. Proceedings *"Applications of Fibonacci Numbers"*, vol. 7, 311–318, Kluwer, 1998.
[13] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* John Wiley & Sons, New York, 2001.

Frédérique Bassino, Institut Gaspard Monge, Unversité de Marne-la-Vallée, 77454 Marne-la-Vallée Cedex 2, France

E-mail address: `bassino@univ-mlv.fr`

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

E-mail address: `helmut@staff.ms.wits.ac.za`