ORDER STATISTICS FOR THE CANTOR-FIBONACCI DISTRIBUTION

LIGIA-LORETTA CRISTEA[*] AND HELMUT PRODINGER[†]

ABSTRACT. The Cantor distribution is obtained from strings consisting of words of 0 and 1 and appropriately attaching a value to them. The Cantor-Fibonacci distribution additionally rejects strings with two adjacent letters 1. Probabilities are brought in by assuming that each admissible string (word) of length m is equally likely; eventually the limit $m \to \infty$ is considered. In this way, one can still deal with discrete objects.

We assume that n random numbers (values of random strings) are drawn independently. The interest is in order statistics of these n values: the (average of) the smallest resp. largest of them. Recursions are obtained which are evaluated asymptotically.

Generalisations to the d-smallest resp. d-largest element are also considered.

1. INTRODUCTION

The Cantor-Fibonacci distribution has been introduced in [8] as a combination of the popular Cantor distribution (ternary expansions with digit 1 missing) and an idea that is immanent in research related to Fibonacci numbers (no two adjacent digits 1).

In this pioneering paper, the moments of the distributions were investigated, both, by giving a recursion formula for them, as well as an asymptotic expansion. The methods built on ones used in earlier papers on the (generalized) Cantor-distribution: The paper by Grabner and the second author [3] investigated the moments, solving open questions by [7]; the paper by Knopfmacher and the second author [6] dealt with the average of the minimum element of n elements, drawn from the Cantor distribution, again solving problems left open by [4].

Similar questions were also addressed in [1], where the interest was instead in just the Cantor-distribution in general q-ary expansions with missing digits.

Here, we want to go back to the Cantor-Fibonacci distribution, and consider the (average of the) minimal element, when we draw n (independent) random elements. Likewise, we are interested in the maximal element. (In previous research this followed from the minimum by symmetry, but not here). We go, however, one step further, and consider general *order statistics*: We think about the n elements being sorted as $y_1 \le y_2 \le \cdots \le y_n$. The minimum is then y_1, the maximum y_n, but we consider more generally the element y_d, the d-th order statistics. We get an exact recursion for the average of it. We can solve this one asymptotically for a) fixed d, and b) for fixed l, with $d = n - l$. Other regimes of d seem to be harder and are left for future research.

The *Cantor distribution with parameter* θ, $0 < \theta \leq \frac{1}{2}$ was introduced in [7] by the random series

$$\frac{\bar{\theta}}{\theta} \sum_{i \geq 1} X_i \theta^i, \tag{1.1}$$

where X_i are independent random variables with the distribution

$$\mathbb{P}\{X_i = 0\} = \mathbb{P}\{X_i = 1\} = \frac{1}{2}, \tag{1.2}$$

and $\bar{\theta} = 1 - \theta$. The name of this distribution stems from the fact that for the particular value $\theta = \frac{1}{3}$ one deals exactly with the numbers situated in the interval $[0,1]$ that admit a ternary expansion consisting only of the digits 0 and 2.

A slightly different point of view is to consider infinite (random) words $\omega_1 \omega_2 \ldots$ over the alphabet $\{0, 1\}$ (with the usual concatenation operation) and the mapping value, defined by

$$\mathsf{value}(\omega_1 \omega_2 \ldots) := \frac{\bar{\theta}}{\theta} \sum_{i \geq 1} \omega_i \theta^i. \tag{1.3}$$

Remark. One can easily see that the mapping value is monotonically increasing with respect to the lexicographic ordering of the set of infinite words $\omega_1 \omega_2 \ldots$ over $\{0, 1\}$.

It is straightforward to prove the recursions

$$\mathsf{value}(0\omega) = \theta \cdot \mathsf{value}(\omega), \qquad \mathsf{value}(1\omega) = \bar{\theta} + \theta \cdot \mathsf{value}(\omega). \tag{1.4}$$

In many situations, restrictions imposed on the mentioned words are of relevance. According to the *Fibonacci restriction* on words over the alphabet $\{0, 1\}$, two letters 1 are not allowed to occur in adjacent positions. In the sequel such words will be referred to as *Fibonacci words*. Fibonacci words of given length are enumerated by Fibonacci numbers, whence the name.

1.1. **The Cantor-Fibonacci distribution.** Here we recall how this distribution was introduced in [8]: Let us consider the set of (finite) Fibonacci words $\mathcal{F}$. This set (language) can be described as

$$\mathcal{F} = \{0, 10\}^* \{\varepsilon + 1\}, \tag{1.5}$$

where "$*$" denotes arbitrary repetitions and "ε" the empty word. It is easy to verify the recurrence

$$\mathcal{F} = \varepsilon + 1 + \{0, 10\}\mathcal{F}.$$

The next step is to define a distribution, i.e., to introduce probabilities on the space of Fibonacci words.

When defining the Cantor distribution, probabilities are simply introduced by assuming that each of the letters 0 and 1 can appear with probability $\frac{1}{2}$. See [7, 4] for more details.

In the case of the *Cantor-Fibonacci distribution* we deal with a more complicated situation. Here the assumption is that all F_{m+2} Fibonacci words of length m appear with the same probability $1/F_{m+2}$, where F_{m+2} denotes the $(m+2)$-nd Fibonacci number, since they count Fibonacci words of length n. The idea in the sequel is to work with *finite words,* and the let the word length m tend to infinity.

We will often use the generating function

$$F(z) = \sum_{m \geq 0} F_{m+2} z^m = \frac{1+z}{1-z-z^2},$$ (1.6)

where

$$F_k = \frac{1}{\sqrt{5}}(\alpha^k - \beta^k), \quad \text{with} \quad \alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2}.$$ (1.7)

As it will be shown, the constant α plays an important role in the considerations to follow.

Remark. Since Fibonacci words form a subset of all infinite words over the alphabet $\{0, 1\}$, the mapping value is monotonically increasing with respect to the lexicographic ordering of (Fibonacci) words. Moreover, as one can easily check, it is strictly increasing.

1.2. **Order statistics for Fibonacci words.** As already mentioned, the aim of this paper is to study several problems of order statistics for Fibonacci words with respect to the Cantor-Fibonacci distribution.

The problem setting for the average *d-th minimum* can be described as follows:
Let $d \geq 1$ be a fixed integer. For an integer $n \geq d$ we choose at random (with respect to the Cantor-Fibonacci distribution) n Fibonacci words and evaluate the function value for each of them. We choose the string (or one of the strings, if there are more of them) for which the minimal value is attained. We repeat this procedure for the other $n-1$ words left and choose thus a second string. After repeating this d times we select the d-th string whose value we call the *d-th minimum*. For $d = 1$ one gets the (usual) minimum among the n strings. For given integers n and d as above we denote by $M_{n,d}$ the average value (with respect to the Cantor-Fibonacci distribution) of the d-th minimum among n random Fibonacci words. We study the asymptotic behaviour of $M_{n,d}$ for $n \to \infty$ and d fixed.

The problem of the *(l + 1)-st maximum* can be formulated in the following way:
Let $l \geq 0$ be a fixed integer. For $n \geq l + 1$ we consider at random n Fibonacci words. Among these we choose a string with the maximal value and call it the *maximum* (or *first maximum*). We repeat the procedure for the $n - 1$ strings left, inductively, and after $l + 1$ steps we obtain a $(l + 1)$-th string whose value we call the *(l + 1)-st maximum*. For $l = 0$ one gets the (usual) maximum value among the n Fibonacci words. We denote by $M_{n,n-l}$ the average value of the $(l+1)$-th maximum value among n random Fibonacci words. We study the asymptotic behaviour of $M_{n,n-l}$ for $n \to \infty$ and l fixed.

Remark. It is clear that $M_{n,d}$ is for $d = 1$ the average value of the (usual) minimum among n random Fibonacci words. Similarly is $M_{n,n-l}$ for $l = 0$ the average value of the maximum value among n random Fibonacci words.

In order to make the material more accessible to the reader we first approach, in Sections 2 and 3, in detail the most simple cases: the average minimum $a_n \ (= M_{n,1})$ and the average maximum $b_n \ (= M_{n,n})$. In the last section we pass to the more general setting in order to study the asymptotic behaviour of $M_{n,d}$ and $M_{n,n-l}$, for d resp. l fixed and $n \to \infty$. These messier computations will then only be sketched.

2. MINIMUM

2.1. **The problem setting.** Let $\mathcal{F}^m$ denote the set of all Fibonacci words of length m and by $\mathcal{F}^\infty$ the set of infinite Fibonacci words. Imagine that we pick at random (with respect to the Cantor-Fibonacci distribution), independently, n Fibonacci words from $\mathcal{F}^m$, for $n \geq 1$. We apply the function value defined in (1.3) to each of the chosen words and look for the minimum among these n values. The same can be done with all random choices of n words of $\mathcal{F}^\infty$. Let us denote by $a_n^{(m)}$ the average minimal value among all possible choices of n Fibonacci words of length m. By taking the limit $a_n := \lim_{m \to \infty} a_n^{(m)}$ we obtain the average minimal value among all choices of n words of $\mathcal{F}^\infty$. We are interested in the study of the asymptotic behaviour of a_n, for $n \to \infty$. The first step is to establish the recurrence relation

$$a_n^{(m)} = \theta \sum_{k=1}^{n} \binom{n}{k} \left(\frac{F_{m+1}}{F_{m+2}} \right)^k \left(\frac{F_m}{F_{m+2}} \right)^{n-k} a_k^{(m-1)} + \left(\frac{F_m}{F_{m+2}} \right)^n \left(\bar{\theta} + \theta^2 a_n^{(m-2)} \right).$$

This recursion is obtained based on the following idea: if among the n random Fibonacci strings there are any strings having the first digit 0 then the minimum value will be attained by (at least) one of these strings. In this case the first digit scales the recursively determined minimum $a_n^{(m-1)}$ by the factor θ. If all n random Fibonacci strings start with 10 then the value of the recursively determined minimum $a_n^{(m-2)}$ is scaled by the factor θ^2 and increased by the value $\bar{\theta}$. These observations easily follow from (1.4).

Now by taking the limit $m \to \infty$ we get:

$$a_n = \theta \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} \alpha^{-2(n-k)} a_k + \alpha^{-2n} \left(\bar{\theta} + \theta^2 a_n \right), \quad \text{for all integers } n \geq 1. \qquad (2.1)$$

Thus we have proven the following

Theorem 1. *The average minimum value among n Fibonacci words with respect to the Cantor-Fibonacci distribution satisfies the recursion*

$$a_n = \theta \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} \alpha^{-2(n-k)} a_k + \alpha^{-2n} \left(\bar{\theta} + \theta^2 a_n \right), \text{ for all integers } n \geq 1.$$

Remarks. For computational reasons it is convenient to set $a_0 = 0$.
In order to compute the elements a_n inductively, for $n = 1, 2, \ldots$, one can rewrite the recursion (2.1):

$$a_n = \frac{1}{1 - \alpha^{-n}\theta - \alpha^{-2n}\theta^2} \left(\theta \sum_{k=1}^{n-1} \binom{n}{k} \alpha^{-k} \alpha^{-2(n-k)} a_k + \alpha^{-2n} \bar{\theta} \right). \qquad (2.2)$$

2.2. **The asymptotics of the average minimum** a_n**.** The recurrence in (2.1) is useful for the study of the asymptotic behaviour of the average minimum a_n, for $n \to \infty$. We introduce the *exponential generating function*

$$A(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

From (2.1) we obtain by multiplication by α^{2n}

$$\alpha^{2n}a_n = \theta \sum_{k=1}^n \binom{n}{k} \alpha^k a_k + \bar{\theta} + \theta^2 a_n,$$

implying

$$A(\alpha^2 z) = \theta e^z A(\alpha z) + \bar{\theta}(e^z - 1) + \theta^2 A(z).$$

This yields, by multiplication with $e^{-\alpha^2 z}$,

$$e^{-\alpha^2 z}A(\alpha^2 z) = \theta e^{-\alpha^2 z}e^z A(\alpha z) + e^{-\alpha^2 z}\bar{\theta}(e^z - 1) + e^{-\alpha^2 z}\theta^2 A(z)$$
$$= \theta e^{-\alpha z}A(\alpha z) + \bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha^2 z}\theta^2 A(z).$$

Thus, for the *Poisson transformed function* $\widehat{A}(z) := e^{-z}A(z)$ we have

$$\widehat{A}(\alpha^2 z) = \theta \widehat{A}(\alpha z) + \bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha z}\theta^2 \widehat{A}(z). \tag{2.3}$$

As we are looking for the asymptotics of the average minimum a_n we are going to study the behaviour of $\widehat{A}(z)$ for $z \to \infty$. This is based on the fact that $a_n \sim \widehat{A}(n)$, which can be justified by using the technique of *depoissonisation* (for details about depoissonisation we refer to [5, 9]). The idea is to extract the coefficients a_n from $A(z)$ using Cauchy's integral formula and the saddle point method.

Let us rewrite (2.3) as

$$\widehat{A}(\alpha^2 z) = \theta \widehat{A}(\alpha z) + R(z), \tag{2.4}$$

where $R(z) = \bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha z}\theta^2 \widehat{A}(z)$ is considered to be an auxiliary function which we treat as a known function. These techniques have already been used in earlier papers, see the references given in the introduction.

We compute the Mellin transform $\widehat{A}^*(s)$ of the function $\widehat{A}(z)$:

$$\alpha^{-2s}\widehat{A}^*(s) = \theta \alpha^{-s}\widehat{A}^*(s) + R^*(s),$$

i.e.,

$$\widehat{A}^*(s) = \frac{R^*(s)}{\alpha^{-2s} - \theta \alpha^{-s}}.$$

From here the function $\widehat{A}(z)$ can be obtained by the Mellin inversion formula (for details regarding the Mellin transform we refer to [2]),

$$\widehat{A}(z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \widehat{A}^*(s) \cdot z^{-s} ds = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{R^*(s)}{\alpha^{-2s} - \theta \alpha^{-s}} \cdot z^{-s} ds,$$

where $0 < c < -\log_\alpha \theta$.

We shift the integral to the right and take the residues with negative sign into account in order to estimate $\widehat{A}(z)$. The function under the integral has simple poles at $s_k = -\log_\alpha \theta + \frac{2\pi i k}{\log \alpha}$, $k \in \mathbb{Z}$. Thus its residues with negative sign in s_k, $k \in \mathbb{Z}$, are

$$\frac{1}{\log \alpha}\, \theta^{-2} R^*\Big(-\log_\alpha \theta + \frac{2k\pi i}{\log \alpha}\Big) z^{\log_\alpha \theta - \frac{2k\pi i}{\log \alpha}},$$

with $R^*(s) = \int_0^\infty \big(\bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha z}\theta^2 \widehat{A}(z)\big) z^{s-1} dz.$

In particular, for $k = 0$ the residue with negative sign is

$$\frac{1}{\log \alpha} \, z^{\log_\alpha \theta} \, \theta^{-2} \int_0^\infty \Big(\bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha z} \theta^2 \widehat{A}(z) \Big) z^{-\log_\alpha \theta - 1} dz.$$

This term plays an essential role in the asymptotic behaviour of the average minimum a_n, as the contributions from the other poles only contribute small fluctuations. By collecting all these residues into a periodic function, one gets the series

$$\frac{\log \alpha}{\theta^2} \sum_{k \in \mathbb{Z}} z^{\log_\alpha \theta - \frac{2k\pi i}{\log \alpha}} \int_0^\infty \Big(\bar{\theta}(e^{-\alpha z} - e^{-\alpha z}) + e^{-\alpha^2 z} \theta^2 \widehat{A}(z) \Big) z^{-\log_\alpha \theta + \frac{2k\pi i}{\log \alpha} - 1} dz.$$

Putting everything together, we have obtained the following result.

Theorem 2. *The average a_n of the minimum value among n random sequences with respect to the Cantor-Fibonacci distribution admits the asymptotic estimate*

$$a_n = \Phi(-\log_\alpha n) \cdot n^{\log_\alpha \theta} \Big(1 + \mathcal{O}\Big(\frac{1}{n}\Big)\Big), \tag{2.5}$$

for $n \to \infty$, where $\Phi(x)$ is a periodic function having period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) of Φ is given by the expression

$$\frac{1}{\theta^2 \log \alpha} \int_0^\infty \Big(\bar{\theta}(e^{-\alpha z} - e^{-\alpha^2 z}) + e^{-\alpha z} \theta^2 \widehat{A}(z) \Big) z^{-\log_\alpha \theta - 1} dz. \tag{2.6}$$

Remark. One can compute this integral numerically by taking for $\widehat{A}(z)$ the first few terms of its Taylor expansion, which can be found from the recurrence (2.1) for the numbers a_n. In order to do this we rewrite (2.6) as

$$\frac{1}{\log \alpha} \Big(\frac{\bar{\theta}^2}{\theta} \, \Gamma(-\log_\alpha \theta) + \theta^2 \sum_{n \geq 0} \frac{a_n}{n! \, \alpha^{2n}} \, \Gamma(n - \log_\alpha \theta) \Big).$$

For instance, for $\theta = \frac{1}{3}$, this sum evaluates to $3.31661\ldots$, and the ratio $a_{300}/300^{\log_\alpha \theta}$ to $3.27556\ldots$.

3. MAXIMUM

3.1. **The problem setting.** We use the notations of the previous section. Again we pick at random (with respect to the Cantor-Fibonacci distribution), independently, n Fibonacci words from $\mathcal{F}^m$, for $n \geq 1$. We apply the function value defined in (1.3) to each of the chosen words and look for the maximum among these n values. The same can be done with all random choices of n words of $\mathcal{F}^\infty$. Let us denote by $b_n^{(m)}$ the average maximal value among all possible choices of n Fibonacci words of length m. By taking the limit $b_n := \lim_{m \to \infty} b_n^{(m)}$ we obtain the average maximal value among all choices of n words of $\mathcal{F}^\infty$. We are interested in the study of the asymptotic behaviour of b_n, for $n \to \infty$.

First we find a recurrence relation between b_n and b_k with $1 \leq k \leq n$:

$$b_n^{(m)} = \sum_{k=1}^n \binom{n}{k} \Big(\frac{F_m}{F_{m+2}}\Big)^k \Big(\frac{F_{m+1}}{F_{m+2}}\Big)^{n-k} \big(\bar{\theta} + \theta^2 b_k^{(m-2)} \big) + \Big(\frac{F_{m+1}}{F_{m+2}}\Big)^n \theta b_n^{(m-1)}.$$

One can deduce the above relation proceeding analogously as in the study of the minimum: If among the n random Fibonacci strings there are any strings starting with

1, i.e., with 10, then the maximum value will be attained for (at least) one of these strings and the value of the recursively determined maximum $b_n^{(m-2)}$ is scaled by the factor θ^2 and increased by $\bar{\theta}$. In the complementary case the value of the recursively determined maximum $b_n^{(m-1)}$ is scaled by the factor θ. By taking the limit for $m \to \infty$ we obtain the relation

$$b_n = \sum_{k=1}^{n} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} (\bar{\theta} + \theta^2 b_k) + \alpha^{-n} \theta b_n, \tag{3.1}$$

and by rearrangement of terms we get the following

Theorem 3. *The average maximum value among n Fibonacci words with respect to the Cantor-Fibonacci distribution satisfies the recursion*

$$b_n = \bar{\theta}\Big(1 - \frac{1}{\alpha^n}\Big) + \theta^2 \sum_{k=1}^{n} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} b_k + \alpha^{-n} \theta b_n, \tag{3.2}$$

for all integers $n \geq 1$.

Remarks. It is convenient to set $b_0 := \frac{1}{1+\theta}$.
In order to compute the elements b_n, for $n = 1, 2, \ldots$, one can rewrite the above recursion as

$$b_n = \frac{1}{1 - \alpha^{-2n}\theta^2 - \alpha^{-n}\theta} \Big(\bar{\theta}\Big(1 - \frac{1}{\alpha^n}\Big) + \theta^2 \sum_{k=1}^{n-1} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} b_k \Big), \quad n \geq 1. \tag{3.3}$$

3.2. **The asymptotics of the average maximum b_n.** Our goal is now to study the asymptotic behaviour of b_n, for $n \to \infty$. We proceed analogously to the previous section, but here R and Φ denote new functions, related to the behaviour of the average maximum.

Since we expect b_n to approach $\mathsf{value}(0.10101010\ldots) = \frac{1}{1+\theta}$, we set $b_n := \frac{1}{1+\theta} - c_n$ for all integers $n \geq 0$ and study the behaviour of c_n in order to get the desired information about b_n.

We look for a recurrence relation for c_n. The relation (3.3) becomes

$$\frac{1}{1+\theta} - c_n = \theta\Big(1 - \frac{1}{\alpha^n}\Big) + \alpha^{-n}\theta^2 \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} \Big(\frac{1}{1+\theta} - c_k\Big) + \alpha^{-n}\theta\Big(\frac{1}{1+\theta} - c_n\Big),$$

which leads us to

$$c_n = -\theta\left(1 - \frac{1}{\alpha^n}\right) - \alpha^{-n}\frac{\theta^2}{1+\theta} \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k}$$

$$+ \alpha^{-n}\,\theta^2 \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} c_k - \frac{\theta}{1+\theta}\alpha^{-n} + \frac{1}{1+\theta} + \alpha^{-n}\theta c_n,$$

and herefrom

$$c_n = \alpha^{-n}\Big(\bar{\theta} - \frac{\theta}{1+\theta}\Big) - \theta + \frac{1}{1+\theta} - \alpha^{-n}\frac{\theta^2}{1+\theta}\left(\Big(1 + \frac{1}{\alpha}\Big)^n - 1\right)$$

$$+ \alpha^{-n}\theta^2 \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} c_k + \alpha^{-n}\theta c_n,$$

which finally leads us to

$$\alpha^n c_n = \frac{\bar{\theta}}{1+\theta} + \theta^2 \sum_{k=1}^{n} \binom{n}{k} \alpha^{-k} c_k + \theta c_n. \qquad (3.4)$$

Let us now define the *exponential generating function*

$$C(z) = \sum_{n \geq 0} c_n \frac{z^n}{n!}.$$

From (3.4) one can immediately deduce

$$C(\alpha z) = \frac{\bar{\theta}}{1+\theta}(e^z - 1) + \theta^2 \, e^z C\left(\frac{z}{\alpha}\right) + \theta C(z).$$

This implies that for the *Poisson transformed function* $\widehat{C}(z) = C(z) \cdot e^{-z}$ we have

$$\widehat{C}(\alpha z) = \frac{\bar{\theta}}{1+\theta}(e^{\frac{-z}{\alpha}} - e^{-\alpha z}) + \theta^2 \widehat{C}\left(\frac{z}{\alpha}\right) + \theta e^{-\frac{z}{\alpha}}\widehat{C}(z).$$

By the substitution $z := z/\alpha$ we can rewrite this as

$$\widehat{C}(z) = \frac{\bar{\theta}}{1+\theta}\left(e^{-\frac{z}{\alpha^2}} - e^{-z}\right) + \theta^2 \widehat{C}\left(\frac{z}{\alpha^2}\right) + \theta e^{-\frac{z}{\alpha^2}}\widehat{C}\left(\frac{z}{\alpha}\right),$$

i.e.,

$$\widehat{C}(z) = \theta^2 \widehat{C}\left(\frac{z}{\alpha^2}\right) + R(z), \qquad (3.5)$$

where $R(z) = \frac{\bar{\theta}}{1+\theta}(e^{-\frac{z}{\alpha^2}} - e^{-z}) + \theta \, e^{-\frac{z}{\alpha^2}}\widehat{C}\left(\frac{z}{\alpha}\right)$ is considered to be an auxiliary function which we treat as a known function. Now we compute the Mellin transform $\widehat{C}^*(s)$ of the function $\widehat{C}(z)$,

$$\widehat{C}^*(s) = \theta^2 \alpha^{2s} \widehat{C}^*(s) + R^*(s) = \frac{R^*(s)}{1 - \theta^2 \cdot \alpha^{2s}}. \qquad (3.6)$$

At this stage the function $\widehat{C}(z)$ can be obtained by applying the Mellin inversion formula

$$\widehat{C}(z) = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \widehat{C}^*(s) \cdot z^{-s} ds = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \frac{R^*(s)}{1 - \theta^2 \cdot \alpha^{2s}} \cdot z^{-s} ds, \qquad (3.7)$$

where $0 < c < -\log_\alpha \theta$.
Now we shift the above integral to the right and take the residues with negative sign in order to estimate $\widehat{C}(z)$ in (3.7). The function under the integral has simple poles at $s_k = -\log_\alpha \theta + \frac{2k\pi \mathrm{i}}{\log \alpha^2}$, $k \in \mathbb{Z}$. For these the residues with negative sign are

$$\frac{1}{\log \alpha^2} R^*\left(-\log_\alpha \theta + \frac{2k\pi \mathrm{i}}{\log \alpha^2}\right) z^{\log_\alpha \theta - \frac{2k\pi \mathrm{i}}{\log \alpha^2}},$$

where $R^*(s) = \int_0^\infty \left(\frac{\bar{\theta}}{1+\theta}(e^{\frac{-z}{\alpha^2}} - e^{-z}) + \theta e^{-\frac{z}{\alpha^2}}\widehat{C}\left(\frac{z}{\alpha}\right)\right) z^{s-1} dz$.
For $k = 0$ the residue with negative sign is

$$\frac{1}{\log \alpha^2} \, z^{\log_\alpha \theta} \int_0^\infty \left(\frac{\bar{\theta}}{1+\theta}(e^{-\frac{z}{\alpha^2}} - e^{-z}) + \theta e^{-\frac{z}{\alpha^2}}\widehat{C}\left(\frac{z}{\alpha}\right)\right) z^{-\log_\alpha \theta - 1} dz.$$

This term constitutes the main contribution to the asymptotic behaviour of c_n. We collect all mentioned residues into a periodic function and obtain the series

$$\frac{1}{\log \alpha^2} \sum_{k \in \mathbb{Z}} z^{\log_\alpha \theta - \frac{2k\pi i}{\log \alpha^2}} \int_0^\infty \Big(\frac{\bar{\theta}}{1+\theta} (e^{-\frac{z}{\alpha^2}} - e^{-z}) + \theta e^{-\frac{z}{\alpha^2}} \widehat{C}\Big(\frac{z}{\alpha}\Big) \Big) z^{-\log_\alpha \theta + \frac{2k\pi i}{\log \alpha^2} - 1} dz.$$

Herefrom we get for $b_n = \frac{1}{1+\theta} - c_n$ the following

Theorem 4. *The average b_n of the maximum value among n random sequences with respect to the Cantor-Fibonacci distribution admits the asymptotic estimate*

$$b_n = \frac{1}{1+\theta} - \Phi(-\log_{\alpha^2} n) \cdot n^{\log_\alpha \theta}\Big(1 + \mathcal{O}\Big(\frac{1}{n}\Big)\Big), \tag{3.8}$$

for $n \to \infty$, where $\Phi(x)$ is a periodic function having period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) of Φ is given by the expression

$$\frac{1}{\log \alpha^2} \int_0^\infty \Big(\frac{\bar{\theta}}{1+\theta} (e^{-\frac{z}{\alpha^2}} - e^{-z}) + \theta e^{-\frac{z}{\alpha^2}} \widehat{C}\Big(\frac{z}{\alpha}\Big) \Big) z^{-\log_\alpha \theta - 1} dz. \tag{3.9}$$

Remark. One can compute this integral numerically by taking for $\widehat{C}(z)$ the first few terms of its Taylor expansion, which can be found by taking into account the relation $c_n = \frac{1}{1+\theta} - b_n$, for $n \geq 0$ and the recurrence (3.2) for b_n, for $n \geq 1$. For this purpose we can rewrite the expression in (3.9) as

$$\frac{\bar{\theta}^2}{\theta^2 \log \alpha^2} \, \Gamma(-\log_\alpha \theta) + \frac{\theta}{\log \alpha^2} \sum_{n \geq 0} \frac{c_n}{n! \, \alpha^n} \, \Gamma(n - \log_\alpha \theta).$$

Again, for $\theta = \frac{1}{3}$, this series evaluates to $5.35114\ldots$, and $c_{300}/300^{\log_\alpha \theta}$ to $5.27105\ldots$.

4. Generalisations

4.1. **The d-th minimum, $d \geq 1$.** Let $d \geq 1$ be a *fixed* integer. For $n \geq d$ let $M_{n,d}$ be the average value of the d-smallest element. In the notation of the previous sections, $a_n = M_{n,1}$ and $b_n = M_{n,n}$.

Here is the general recursion:

Theorem 5. *Let $d \geq 1$ be a* fixed *integer. The average value of the d-th minimum among n random Fibonacci words with respect to the Cantor-Fibonacci distribution satisfies the recursion*

$$M_{n,d} = \theta \sum_{k=d}^n \binom{n}{k} \alpha^{-k} \alpha^{-2(n-k)} M_{k,d} + \sum_{k=0}^{d-1} \binom{n}{k} \alpha^{-k} \alpha^{-2(n-k)} \Big(\bar{\theta} + \theta^2 M_{n-k,d-k} \Big), \tag{4.1}$$

for all integers $n \geq d$.

Our aim is now to pass, analogously to the previous sections, from the above recurrence to an equation in terms of generating functions.

We define the *exponential generating function*

$$A_d(z) := \sum_{n \geq d} M_{n,d} \frac{z^n}{n!},$$

for any fixed integer $d \geq 1$.

By multiplying (4.1) by $\alpha^{2n} \frac{z^n}{n!}$ and then summing up for $n \geq d$ we obtain

$$A_d(\alpha^2 z) = \theta e^z A_d(\alpha z) + K_d(z),$$

where

$$K_d(z) := \sum_{n \geq d} \frac{z^n}{n!} \sum_{k=0}^{d-1} \binom{n}{k} \alpha^k \left(\bar{\theta} + \theta^2 M_{n-k,d-k} \right).$$

This implies

$$e^{-\alpha^2 z} A_d(\alpha^2 z) = \theta e^{(1-\alpha^2)z} A_d(\alpha z) + e^{-\alpha^2 z} K_d(z),$$

i.e., for the *Poisson transformed function* $\widehat{A}_d(z) := e^{-z} A_d(z)$,

$$\widehat{A}_d(\alpha^2 z) = \theta \widehat{A}_d(\alpha z) + R_d(z), \tag{4.2}$$

with $R_d(z) = e^{-\alpha^2 z} K_d(z)$.

Remark. One can easily check that, as expected, for $d = 1$ the function $R_d(z)$ coincides with $R(z)$ occurring in Section 2.

It is easy to see that from here on the computations leading to the asymptotic behaviour of $M_{n,d}$ follow exactly those shown in Section 2, whereas in all formulæ containing $R(z)$ we will replace this by the function $R_d(z)$ defined above. (We leave these details to the reader as a straightforward exercise.) Thus we obtain the following result.

Theorem 6. *The average $M_{n,d}$ of the d-th minimum value among n random sequences with respect to the Cantor-Fibonacci distribution admits the asymptotic estimate*

$$M_{n,d} = \Phi_d(-\log_\alpha n) \cdot n^{\log_\alpha \theta} \left(1 + \mathcal{O}\left(\frac{1}{n} \right) \right), \tag{4.3}$$

for $n \to \infty$, where $\Phi_d(x)$ is a periodic function having period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) of Φ_d is given by the expression

$$\frac{1}{\theta^2 \log \alpha} \int_0^\infty \left(e^{-\alpha^2 z} \sum_{n \geq d} \frac{z^n}{n!} \sum_{k=0}^{d-1} \binom{n}{k} \alpha^k \left(\bar{\theta} + \theta^2 M_{n-k,d-k} \right) \right) z^{-\log_\alpha \theta - 1} dz. \tag{4.4}$$

4.2. The $(l+1)$-st maximum $M_{n,n-l}$, $l \geq 0$. Rewriting (4.1) in terms of $M_{n,n-l}$, we get

Theorem 7. *Let $l \geq 0$ be a* fixed *integer. The average value $M_{n,n-l}$ of the $(l+1)$-st maximum among n random Fibonacci words with respect to the Cantor-Fibonacci distribution satisfies the recursion*

$$M_{n,n-l} = \sum_{k=l+1}^{n} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} (\bar{\theta} + \theta^2 M_{k,k-l}) + \theta \sum_{k=0}^{l} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} M_{n-k,n-l}, \tag{4.5}$$

for all $n \geq l + 1$.

As in the previous sections, we use the recursion (4.5) in order to deduce an equation for the generating functions.

Since we expect $M_{n,n-l}$ to approach $\mathsf{value}(0.10101010\dots) = \frac{1}{1+\theta}$ for $n \to \infty$, we set, analogously to Section 3, $M_{n,n-l} := \frac{1}{1+\theta} - P_{n,n-l}$ for all integers $n \geq l+1$ and study the behaviour of $P_{n,n-l}$ for $n \to \infty$.

With the above substitution (4.5) becomes

$$\frac{1}{1+\theta} - P_{n,n-l} = \sum_{k=l+1}^{n} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} \left(\bar{\theta} + \theta^2 \left(\frac{1}{1+\theta} - P_{k,k-l} \right) \right)$$
$$+ \theta \sum_{k=0}^{l} \binom{n}{k} \alpha^{-2k} \alpha^{-(n-k)} \left(\frac{1}{1+\theta} - P_{n-k,n-l} \right),$$

for all integers $n \geq l+1$.

We rearrange some terms, multiply by $\alpha^n \frac{z^n}{n!}$ and sum up for $n \geq l+1$ and finally introduce the *exponential generating function*

$$C_l(z) = \sum_{n \geq l+1} P_{n,n-l} \frac{z^n}{n!}$$

to obtain

$$C_l(\alpha z) = \theta^2 e^z C_l(\frac{z}{l}) + H_l(z),$$

where

$$H_l(z) = \frac{1}{1+\theta} \sum_{n \geq l+1} \frac{\alpha^n z^n}{n!}$$
$$+ \sum_{n \geq l+1} \left(\theta \sum_{k=0}^{l} \binom{n}{k} \alpha^{-k} P_{n-k,n-l} - \frac{\theta}{1+\theta} \sum_{k=0}^{l} \binom{n}{k} \alpha^{-k} - \frac{1}{1+\theta} \sum_{k=l+1}^{n} \binom{n}{k} \alpha^{-k} \right) \frac{z^n}{n!}$$
$$= \sum_{n \geq l+1} \frac{z^n}{n!} \sum_{k=0}^{l} \binom{n}{k} \alpha^{-k} \left(\theta P_{n-k,n-l} + \frac{\bar{\theta}}{1+\theta} \right).$$

Thus for the *Poisson transformed function* $\widehat{C}_l(z) = C(z) \cdot e^{-z}$ we deduce

$$\widehat{C}_l(\alpha z) = \theta^2 \widehat{C}_l(\frac{z}{\alpha^2}) + e^{-\alpha z} H_l(z), \tag{4.6}$$

Herefrom we obtain, by substituting $z = z/\alpha$,

$$\widehat{C}_l(z) = \theta^2 \widehat{C}_l(\frac{z}{\alpha^2}) + Q_l(z), \tag{4.7}$$

where $Q_l(z) = e^{-z} H_l(\frac{z}{\alpha})$.

Remark. As one would expect, for $l = 0$, $Q_l(z)$ coincides with the auxiliary function $R(z)$ that occurs in Section 3.

As in the generalisation of the average minimum, from here on the computations leading to the asymptotic behaviour of $M_{n,n-l}$ follow exactly those from the study of the average maximum b_n in Section 3, whereas in all formulæ in Section 3 containing $R(z)$ we will replace this by the function $Q_l(z)$ defined above. Thus one gets the proof of the following theorem:

Theorem 8. *The average $M_{n,n-l}$ of the $(l+1)$-st maximum value among n random sequences with respect to the Cantor-Fibonacci distribution admits the asymptotic estimate*

$$M_{n,n-l} = \frac{1}{1+\theta} - \Phi_l(-\log_{\alpha^2} n) \cdot n^{\log_\alpha \theta}\Big(1 + \mathcal{O}\Big(\frac{1}{n}\Big)\Big), \qquad (4.8)$$

for $n \to \infty$, where $\Phi_l(x)$ is a periodic function having period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) of Φ_l is given by the expression

$$\frac{1}{\log \alpha^2} \int_0^\infty \left(e^{-z} \sum_{n \ge l+1} \frac{z^n}{n! \alpha^n} \sum_{k=0}^l \binom{n}{k} \alpha^{-k} \Big(\theta P_{n-k,n-l} + \frac{\bar\theta}{1+\theta} \Big) \right) z^{-\log_\alpha \theta - 1} dz. \qquad (4.9)$$

Of course, for numerical purposes, the constants in (4.4) and in (4.9) could be expressed as a series involving Gamma functions.

References

[1] F. Bassino and H. Prodinger. (q,δ)-numeration systems with missing digits. *Monatshefte für Mathematik*, 141:89–99, 2004.

[2] Ph. Flajolet, X. Gourdon, and Ph. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoret. Comput. Sci.*, 144:3–58, 1995.

[3] P. J. Grabner and H. Prodinger. Asymptotic analysis of the moments of the Cantor distribution. *Statist. Probab. Lett.*, 26(3):243–248, 1996.

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

[5] Ph. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. *Theoret. Comput. Sci.*, 201(1-2):1–62, 1998.

[6] A. Knopfmacher and H. Prodinger. Explicit and asymptotic formulae for the expected values of the order statistics of the Cantor distribution. *Statist. Probab. Lett.*, 27(2):189–194, 1996.

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

[8] H. Prodinger. The Cantor-Fibonacci distribution. In *Applications of Fibonacci numbers, Vol. 7 (Graz, 1996)*, pages 311–318. Kluwer Acad. Publ., Dordrecht, 1998.

[9] W. Szpankowski. *Average case analysis of algorithms on sequences.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2001. With a foreword by Philippe Flajolet.

LIGIA-L. CRISTEA, JOHANNES KEPLER UNIVERSITY, INSTITUTE FOR MATHEMATICAL ANALYSIS, ALTENBERGERSTR. 69, 4040-LINZ, AUSTRIA, AND DEPARTMENT OF MATHEMATICS, POLITEHNICA UNIVERSITY OF TIMIŞOARA, ROMANIA

E-mail address: `ligia-loretta.cristea@jku.at`

HELMUT PRODINGER, MATHEMATICS DEPARTMENT, STELLENBOSCH UNIVERSITY, 7602 STELLENBOSCH, SOUTH AFRICA

E-mail address: `hproding@sun.ac.za`