THE CANTOR—FIBONACCI DISTRIBUTION

HELMUT PRODINGER

Abstract. The Cantor–Fibonacci distribution is defined by a random series where the coefficients are 0 or 1 and no two adjacent 1's are allowed. A recurrence for the moments is obtained which is then evaluated asymptotically.

1. Introduction

The Cantor distribution with parameter ϑ, $0 < \vartheta \le \frac{1}{2}$, was introduced in [5] by the random series

$$\frac{\bar\vartheta}{\vartheta}\sum_{i\ge1} X_i\vartheta^i \ , \tag{1}$$

where the X_i are independent with the distribution

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

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.

We might of course alternatively think of an infinite (random) word $w_1w_2\dots$ over the alphabet $\{0,1\}$ and a mapping value, defined by

$$\mathsf{value}(w_1w_2\dots)=\frac{\bar\vartheta}{\vartheta}\sum_{i\ge1} w_i\vartheta^i \ . \tag{3}$$

Observe the recursion formula

$$\begin{aligned}\mathsf{value}(0w) &= \vartheta\cdot\mathsf{value}(w)\\ \mathsf{value}(1w) &= \bar\vartheta+\vartheta\cdot\mathsf{value}(w) \ .\end{aligned} \tag{4}$$

Whenever we are dealing with words over the alphabet $\{0,1\}$, we might also think about *restricted words*, according to the *Fibonacci restriction*, that two adjacent letters '1' are not allowed.

Compare some of the papers of the author for that [8, 9, 10].

The set of (finite) Fibonacci words $\mathcal{F}$ (no two adjacent letters '1') is given by

$$\mathcal{F}=\{0,10\}^*\{\varepsilon+1\}; \tag{5}$$

Date: May 9, 1996.

This research was partly done while the author was visiting the CRM (Centre de Recerca Matemàtica, Institut d'Estudis Catalans).

here, '$*$' means arbitrary repetitions and 'ε' the empty word. There is also the recurrence

$$\mathcal{F} = \varepsilon + 1 + \{0, 10\}\mathcal{F} \ . \tag{6}$$

In the original setting probabilities are simply introduced by saying that each letter can appear with probability $\frac{1}{2}$. Here, the situation is more complicated. We say that each Fibonacci word of length m is equally likely. There are F_{m+2} such words, with F_{m+2} denoting the $(m+2)$th Fibonacci number, whence the name. We will work with finite words, and then letting the length m tend to infinity. In the independent original case this can be avoided, but here it is necessary.

As an example, consider the classical Cantor case with $\vartheta = \frac{1}{3}$ and $m = 3$. Then the values

$$\begin{aligned}
2\Big(0\cdot\frac{1}{3}+0\cdot\frac{1}{9}+0\cdot\frac{1}{27}\Big) &= 0\\
2\Big(0\cdot\frac{1}{3}+0\cdot\frac{1}{9}+1\cdot\frac{1}{27}\Big) &= \frac{2}{27}\\
2\Big(0\cdot\frac{1}{3}+1\cdot\frac{1}{9}+0\cdot\frac{1}{27}\Big) &= \frac{2}{9}\\
2\Big(1\cdot\frac{1}{3}+0\cdot\frac{1}{9}+0\cdot\frac{1}{27}\Big) &= \frac{2}{3}\\
2\Big(1\cdot\frac{1}{3}+0\cdot\frac{1}{9}+1\cdot\frac{1}{27}\Big) &= \frac{20}{27}
\end{aligned} \tag{7}$$

appear, each with probability $\frac{1}{5}$.

It is also useful to have $F(z)$, the generating function of Fibonacci words, according to their lengths. From (5) and general principles we get it as

$$F(z) = \frac{1+z}{1-z-z^2} = \sum_{m\ge 0} F_{m+2} z^m \ ; \tag{8}$$

note that

$$F_n = \frac{1}{\sqrt{5}}\Big(\alpha^n - \beta^n\Big) \qquad \text{with} \quad \alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2} \ . \tag{9}$$

2. The Moments

Lad and Taylor in their original paper [5] have considered the nth moment of the Cantor distribution. They obtained a recursion for them, which was analyzed by Grabner and the author in [2]. Other papers dealing with the Cantor distribution are [3] and [4].

In this paper, we want to analyze the nth moments of the Cantor–Fibonacci distribution. We do this by considering the generating functions

$$G_n(z) := \sum_{w\in\mathcal{F}} \Big(\mathsf{value}(w)\Big)^n z^{|w|} \ ; \tag{10}$$

where $|w|$ denotes the length of the Fibonacci word w. The quantity

$$\frac{[z^m]G_n(z)}{[z^m]F(z)} \tag{11}$$

is then the nth moment, when considering words of length m; finally we consider the limit of this for $m \to \infty$ and call it M_n.

We rewrite the recursion for value, when restricted to Fibonacci words;

$$\begin{aligned} \mathsf{value}(0w) &= \vartheta \cdot \mathsf{value}(w) \\ \mathsf{value}(10w) &= \bar{\vartheta} + \vartheta^2 \cdot \mathsf{value}(w) \; . \end{aligned} \tag{12}$$

Note also that $\mathsf{value}(\varepsilon) = 0$ and $\mathsf{value}(1) = \bar{\vartheta}$. Using the recursion $\mathcal{F} = \varepsilon + 1 + \{0, 10\}\mathcal{F}$, we find for $n \geq 1$

$$\begin{aligned} G_n(z) &= \bar{\vartheta}^n z + \vartheta^n z G_n(z) + \sum_{w \in \mathcal{F}} \Big(\bar{\vartheta} + \vartheta^2 \mathsf{value}(w) \Big)^n z^{2+|w|} \\ &= \bar{\vartheta}^n z + \vartheta^n z G_n(z) + z^2 \sum_{i=0}^{n} \binom{n}{i} \bar{\vartheta}^{n-i} \vartheta^{2i} \, G_i(z) \; . \end{aligned} \tag{13}$$

This recursion can be made explicit by rewriting it as

$$G_n(z) = \frac{1}{1 - \vartheta^n z - \vartheta^{2n} z^2} \left[\bar{\vartheta}^n z + z^2 \sum_{i=0}^{n-1} \binom{n}{i} \bar{\vartheta}^{n-i} \vartheta^{2i} \, G_i(z) \right] \tag{14}$$

Note the special instance $G_0(z) = F(z)$ and

$$G_1(z) = \frac{1}{1 - \vartheta z - \vartheta^2 z^2} \left[\bar{\vartheta} z + z^2 \bar{\vartheta} F(z) \right] = \frac{\bar{\vartheta} z}{(1 - \vartheta z - \vartheta^2 z^2)(1 - z - z^2)} \; . \tag{15}$$

From this we could get (by partial fraction decomposition) the coefficient of z^m explicitly.

However, since we only consider the limit for $m \to \infty$, life is easier. Both, $G_1(z)$ and $F(z)$ have the dominant singularity at $z = 1/\alpha$, and it is a simple pole. Consequently,

$$[z^m] G_1(z) \sim A \alpha^m, \qquad \text{with} \quad A = \lim_{z \to 1/\alpha} (1 - z\alpha) G_1(z) \tag{16}$$

and

$$[z^m] F(z) \sim B \alpha^m, \qquad \text{with} \quad B = \lim_{z \to 1/\alpha} (1 - z\alpha) F(z) \tag{17}$$

Therefore

$$\lim_{m \to \infty} \frac{[z^m] G_1(z)}{[z^m] F(z)} = \frac{A}{B} \; , \tag{18}$$

which we can compute as

$$\lim_{z \to 1/\alpha} \frac{\frac{\bar{\vartheta} z}{1 - \vartheta z - \vartheta^2 z^2}}{1 + z} = \frac{\bar{\vartheta}}{\alpha^2 - \alpha \vartheta - \vartheta^2} \; . \tag{19}$$

The fact that $z = 1/\alpha$ is a simple pole remains true for all the functions $G_n(z)$. We can thus divide $G_n(z)$ by $F(z)$, cancel the pole, and insert $z = 1/\alpha$ in the rest, to obtain the nth moment M_n as

$$M_n = \frac{1}{1 - \vartheta^n/\alpha - \vartheta^{2n}/\alpha^2} \left[\frac{1}{\alpha^2} \sum_{k=0}^{n-1} \binom{n}{k} \bar{\vartheta}^{n-k} \vartheta^{2k} M_k \right] . \tag{20}$$

Theorem 1. *The moments of the Cantor–Fibonacci distribution fulfill the following recursion: $M_0 = 1$ and for $n \ge 1$*

$$M_n = \frac{1}{\alpha^2 - \alpha\vartheta^n - \vartheta^{2n}} \sum_{k=0}^{n-1} \binom{n}{k} \bar\vartheta^{n-k}\vartheta^{2k} M_k \ . \tag{21}$$

For instance

$$\begin{aligned} M_1 &= \frac{\bar\vartheta}{\alpha^2 - \alpha\vartheta - \vartheta^2} \ , \\ M_2 &= \frac{\bar\vartheta^2(\alpha^2 - \alpha\vartheta + \vartheta^2)}{(\alpha^2 - \alpha\vartheta^2 - \vartheta^4)(\alpha^2 - \alpha\vartheta - \vartheta^2)} \ , \\ \text{Variance} = M_2 - M_1^2 &= \frac{\bar\vartheta^4\alpha^3}{(\alpha^2 - \alpha\vartheta^2 - \vartheta^4)(\alpha^2 - \alpha\vartheta - \vartheta^2)^2} \ . \end{aligned} \tag{22}$$

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 (= non–Fibonacci) case was done in [2].

A rough estimation shows us that the moments decrease exponentially. Indeed, if we set $M_n \approx \lambda^n$, we might infer that $\lambda = \bar\vartheta + \lambda\vartheta^2$, or

$$\lambda = \frac{1}{1+\vartheta} \ . \tag{23}$$

This is not rigorous, but we can set

$$m_n := M_n \cdot (1+\vartheta)^n \tag{24}$$

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

$$m_n = \frac{1}{\alpha^2 - \alpha\vartheta^n - \vartheta^{2n}} \sum_{k=0}^{n-1} \binom{n}{k} (1-\vartheta^2)^{n-k}\vartheta^{2k} m_k \ . \tag{25}$$

To study this sequence further, we rewrite it as

$$m_n(\alpha^2 - \alpha\vartheta^n - \vartheta^{2n}) = \sum_{k=0}^{n} \binom{n}{k} (1-\vartheta^2)^{n-k}\vartheta^{2k} m_k - \vartheta^{2n} m_n \tag{26}$$

or

$$m_n(\alpha^2 - \alpha\vartheta^n) = \sum_{k=0}^{n} \binom{n}{k} (1-\vartheta^2)^{n-k}\vartheta^{2k} m_k \tag{27}$$

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

$$m(z) = \sum_{n\ge 0} m_n \frac{z^n}{n!} \tag{28}$$

and get

$$\alpha^2 m(z) - \alpha m(\vartheta z) = e^{(1-\vartheta^2)z} m(\vartheta^2 z). \tag{29}$$

As in [2], we have to consider the *Poisson transformed function* $\widehat{m}(z) = e^{-z}m(z)$, which fulfills the functional equation

$$\widehat{m}(z) = \frac{e^{-\bar{\vartheta} z}}{\alpha}\widehat{m}(\vartheta z) + \frac{1}{\alpha^2}\widehat{m}(\vartheta^2 z) \ . \tag{30}$$

This functional equation can – at least in principle – be solved by *iteration*. Even though it is hard to write down the explicit solution, we can consider it to be a *known function*.

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 *de–Poissonization*. A preliminary reference for that is [11]. However, Jacquet and Szpankowski are currently working on a monography about de–Poissonization, which will be the ultimate reference for at least the next ten years. We follow [2], where the technique is explained in more detail.

It might be suggestive to use a new name $R(z)$ for $\frac{e^{-\bar{\vartheta} z}}{\alpha}\widehat{m}(\vartheta z)$ and consider it to be an auxiliary (and known) function;

$$\widehat{m}(z) = \frac{1}{\alpha^2}\widehat{m}(\vartheta^2 z) + R(z) \ . \tag{31}$$

We compute the *Mellin transform* of (31) (see [1] for definitions and properties);

$$\widehat{m}^*(s) = \frac{1}{\alpha^2}\vartheta^{-2s}\widehat{m}^*(s) + R^*(s) = \frac{R^*(s)}{1-\alpha^{-2}\vartheta^{-2s}} \ . \tag{32}$$

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

$$\widehat{m}(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{R^*(s)}{1-\alpha^{-2}\vartheta^{-2s}}\, z^{-s}\, ds \ , \tag{33}$$

where $0 < c < -\log_\vartheta \alpha$. 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 $s = -\log_\vartheta \alpha + \frac{2k\pi i}{\log \vartheta}$, $k \in \mathbb{Z}$. The negative residue there is

$$-\frac{1}{2\log\vartheta}R^*\Big(-\log_\vartheta \alpha + \frac{2k\pi i}{\log\vartheta}\Big)\, z^{\log_\vartheta \alpha - \frac{2k\pi i}{\log\vartheta}} \ . \tag{34}$$

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

$$-\frac{1}{2\log\vartheta}z^{\log_\vartheta \alpha}\int_0^\infty \frac{e^{-\bar{\vartheta} z}}{\alpha}\,\widehat{m}(\vartheta z)\, z^{-\log_\vartheta \alpha - 1}\, dz \ . \tag{35}$$

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

Theorem 2. *The nth moment M_n of the Cantor–Fibonacci distribution has for $n \to \infty$ the following asymptotic behaviour*

$$M_n = (1+\vartheta)^{-n}\Phi(-\log_\vartheta n)\, n^{\log_\vartheta \alpha}\Big(1 + O\Big(\frac{1}{n}\Big)\Big) \ , \tag{36}$$

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

$$-\frac{1}{2\log\vartheta}\int_0^\infty \frac{e^{-\bar{\vartheta} z}}{\alpha}\,\widehat{m}(\vartheta z)\, z^{-\log_\vartheta \alpha - 1}\, dz \ . \tag{37}$$

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

As an example, we treat the classical case $\vartheta = \frac{1}{3}$. Then (apart from small fluctuations),

$$M_n \sim .6160498 \cdot n^{-.4380178} \cdot 0.75^n. \tag{38}$$

This is consistent with the actual values; the reader is encouraged to check it. It is fun!!

The fact that in an asymptotic formula the generating function of the sequence in question, evaluated at a certain point, appears, is not at all uncommon in combinatorial analysis. We cite for instance the classical papers [7], [6].

Acknowledgement. This paper was worked out mainly while I was on holiday in Thailand. I completed it and wrote it up while I was an invited researcher in the *Centre de Recerca Matemâtica, Institut d'Estudis Catalans, Barcelona*. I am thankful for the hospitality encountered there.

References

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

[2] P. Grabner and H. Prodinger. Asymptotic analysis of the moments of the Cantor distribution. *Statistics and Probability Letters (to appear)*, 1994.

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

[4] A. Knopfmacher and H. Prodinger. Exact and asymptotic formulæ for average values of order statistics of the Cantor distribution. *Statistics and Probability Letters (to appear)*, 1994.

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

[6] R. Otter. The number of trees. *Ann. Math.*, 49:583–599, 19348.

[7] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica*, 68:145–254, 1937.

[8] H. Prodinger. On the number of Fibonacci partitions of a set. *The Fibonacci Quarterly*, 19:463–466, 1981.

[9] H. Prodinger. Ordered Fibonacci partitions. *Canadian Mathematical Bulletin*, 26:312–316, 1983.

[10] H. Prodinger. Über längste 1-Teilfolgen in 0-1–Folgen. In E. Hlawka, editor, *Zahlentheoretische Analysis II*, volume 1262 of *Lecture Notes in Mathematics*, pages 124–133, 1987.

[11] B. Rais, P. Jaquet, and W. Szpankowski. Limiting distribution for the depth in Patricia tries. *SIAM Journal on Discrete Mathematics*, 6:197–213, 1993.

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria.

E-mail address: `Helmut.Prodinger@tuwien.ac.at`