Some Properties of the Cantor Distribution

Helmut Prodinger
Technische Universität Wien, Austria

December 9, 1996

[summary by Julien Clément]

Abstract

The Cantor distribution is defined as a random series

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

where ϑ is a parameter and the X_i are random variables that take the values 0 and 1 with probability 1/2. The moments and order statistics are discussed, as well as a "Fibonacci" variation. Connections to certain trees and splitting processes are also mentioned.

1. Cantor distribution

1.1. **Random series.** The Cantor distribution with parameter ϑ $(0<\vartheta\leq 1/2)$ was introduced in [5] by the random series

$$X=\frac{\overline{\vartheta}}{\vartheta}\sum_{i\geq 1} X_i\vartheta^i,$$

where the X_i are independent with the distribution $\Pr[X_i=0]=\Pr[X_i=1]=\frac{1}{2}$, and $\overline{\vartheta}=1-\vartheta$. The name stems from the special case $\vartheta=\frac{1}{3}$, since then this process gives exactly those numbers from the interval $[0,1]$ that have a ternary expansion solely consisting of the digits 0 and 2. We might alternatively consider an infinite (random) word $w_1w_2\cdots$ over the alphabet $\{0,1\}$ and a map value, defined by

$$\mathsf{value}(w_1w_2\cdots)=\frac{\overline{\vartheta}}{\vartheta}\sum_{i\geq 1} w_i\vartheta^i.$$

1.2. **Moments of the distribution.** We abbreviate $a_n=\mathrm{E}[X^n]$. The aim is to solve the recursion formula (from [5])

$$a_n=\frac{1}{2(1-\vartheta^n)}\sum_{k=0}^{n-1}\binom{n}{k}\overline{\vartheta}^{n-k}\vartheta^k a_k,\qquad a_0=1.$$

Let us introduce the exponential generating function $A(z)=\sum_{n\geq 0} a_n\frac{z^n}{n!}$. The functional equation involving $A(z)$, once solved by iteration, gives

$$A(z)=\prod_{k\geq 0}\frac{1+e^{\overline{\vartheta}\vartheta^k z}}{2}.$$

In order to derive an asymptotic equivalent of a_n, the Poisson generating function $B(z) = e^{-z}A(z)$ has to be considered. Using "Mellin" techniques to derive an asymptotic expansion of $\log B(z)$ when z tends to infinity and a "de-poissonization" argument which suggests the approximation $a_n \sim B(n)$, one gets

$$\mathrm{E}[X^n] = a_n = F(\log_{1/\vartheta} n) n^{-\log_{1/\vartheta} 2}\left(1 + O\left(\frac{1}{n}\right)\right).$$

The function $F(x)$ is periodic of period 1 and has known Fourier coefficients. The mean of $F(x)$ is for instance

$$-\frac{1}{2\log\vartheta}\int_0^\infty \prod_{k\geq 1}\frac{1+e^{-\overline{\vartheta}\vartheta^k z}}{2} e^{-\overline{\vartheta}x} x^{\log_{1/\vartheta} 2 - 1} dx.$$

1.3. **Order statistics.** Let us consider n random independent variables $Y_1, \ldots, Y_n$ from a Cantor distribution. The average value $\mathrm{E}[\min(Y_1, \ldots, Y_n)]$ of the smallest value among them is denoted by a_n. The coefficients a_n obey the following recursion

$$(2^n - 2\vartheta)a_n = \overline{\vartheta} + \vartheta \sum_{k=1}^{n-1}\binom{n}{k}\vartheta a_k.$$

Considering now not exactly the Poisson generating function $A(z) = \sum_{k\geq 0} a_n \frac{z^n}{n!}$ but rather

$$\widehat{A}(z) = \frac{1}{e^z - 1}A(z) = \sum_{n\geq 0}\widehat{a}_n \frac{z^n}{n!},$$

a simpler equation can be obtained. Indeed, one has

$$\widehat{A}(2z) = \vartheta\widehat{A}(z) + \frac{\overline{\vartheta}}{e^z + 1}.$$

The coefficients $\widehat{a}_n$ can be extracted directly from this equation (equating coefficients of $\frac{z^n}{n!}$ on both sides). Going back to the original coefficients a_n, we have the explicit solution

$$a_n = -\overline{\vartheta}\sum_{k=0}^{n-1}\binom{n}{k}\frac{B_{k+1}}{k+1}\frac{2^{k+1}-1}{2^k - \vartheta},$$

where B_n denotes a Bernoulli number. An approach based on Rice's method finally gives an asymptotic equivalent of a_n

$$a_n \sim n^{\log_2 \vartheta}\frac{2\vartheta - 1}{\vartheta \log 2}\left(\Gamma(-\log_2 \vartheta)\zeta(-\log_2\vartheta) + \delta(\log_2 n)\right),$$

where $\zeta(s)$, $\Gamma(s)$ and $\delta(s)$ denote respectively the Riemann's zeta function, the gamma function and a periodic function with period 1 and a very small amplitude (provided ϑ is not too close to 0).

2. Cantor-Fibonacci distribution

2.1. **Fibonacci restriction.** The Cantor distribution might be viewed as a mapping value over a set of random words over a binary alphabet. We might also think about *restricted words*, according to the *Fibonacci restriction*, that two adjacent letters '1' are not allowed. The set of (finite) Fibonacci words $\mathcal{F}$ is given by

$$\mathcal{F} = \{0, 01\}^*\{\epsilon + 1\}.$$

In the original setting (*Cantor distribution*) probabilities are simply introduced by saying that each letter of $\{0, 1\}$ can appear with probability $\frac{1}{2}$. Here the situation is more complicated. We say that each word of Fibonacci of length m is equally likely. There are F_{m+2} such words, with F_{m+2} denoting the $(m+2)$th Fibonacci number. As an example, consider the classical Cantor case with $\vartheta = \frac{1}{3}$ and $m = 3$. Then the values

$$\mathsf{value}(000) = 0, \quad \mathsf{value}(001) = \frac{2}{27}, \quad \mathsf{value}(010) = \frac{2}{9}, \quad \mathsf{value}(100) = \frac{2}{3}, \quad \mathsf{value}(101) = \frac{20}{27}$$

appear, each with probability $\frac{1}{5}$. The generating function $F(z)$ of Fibonacci words, according to their lengths is easily derived from the definition of $\mathcal{F}$ above,

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

Note that

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

2.2. **Moments of the Cantor-Fibonacci distribution.** Let us consider the generating functions

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

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

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

is the nth moment, when considering words of length m. Then we let m tend to infinity to get a limit called M_n (note that taking limits wasn't necessary for the independent original case). The recursion for value, when restricted to Fibonacci words, is

$$\mathsf{value}(0w) = \vartheta \cdot \mathsf{value}(w)$$
$$\mathsf{value}(10w) = \overline{\vartheta} + \vartheta^2 \cdot \mathsf{value}(w).$$

These formulae translate almost directly to generating functions according to the recursive definition $\mathcal{F} = \epsilon + 1 + \{0, 10\}\mathcal{F}$. Thus it gives an explicit recursion formula for the functions $G_n(z)$

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

Since we only consider the limit for $m \to \infty$, we can get the asymptotic behaviour noting that both $F(z)$ and $G_n(z)$ have the same dominant singularity at $z = 1/\alpha$ and also that it is a simple pole. Consequently, we have (due to a "pole cancellation")

$$M_n = \lim_{m \to \infty} \frac{[z^m]G^n(z)}{[z^m]F(z)} = \lim_{z \to 1/\alpha} \frac{G_n(z)}{F(z)}.$$

Therefore we have the following theorem

Theorem 1. *The moments of the Cantor-Fibonacci distribution fulfill the following recursion:* $M_0 = 0$ *and for* $n \geq 1$

$$M_n = \frac{1}{\alpha^2 - \alpha \vartheta^n - \vartheta^{2n}} \sum_{i=1}^{n} \binom{n}{i} \overline{\vartheta}^{n-i} \vartheta^{2i} M_i.$$

2.3. **The asymptotic behaviour of the moments.** A rough estimate shows that $M_n \approx \lambda^n$. We might infer that $\lambda = \overline{\vartheta} + \lambda\vartheta^2$, so that $\lambda = \frac{1}{1+\vartheta}$. It is not rigourous but we can set

$$m_n := M_n \cdot (1+\vartheta)^n$$

anyway and show that this sequence has nicer properties. As before the recurrence on the coefficients m_n and then the exponential generating function $m(z) = \sum_n m_n \frac{z^n}{n!}$ need to be considered. Finally the Poisson transformed function $\widehat{m}(z) = e^{-z} m(z)$ obeys the functional equation

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

Because $m_n \sim \widehat{m}(n)$, the next step considers the behaviour of $\widehat{m}(z)$ for $z \to \infty$. Using the Mellin transform (and the Mellin inversion formula), we have the following theorem

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

$$M_n = \left(1+\overline{\vartheta}\right)^{-n} \Phi(-\log_\vartheta n) n^{\log_\vartheta \alpha}\left(1 + O\left(\frac{1}{n}\right)\right),$$

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

$$-\frac{1}{\log\vartheta}\int_0^\infty \frac{e^{-\overline{\vartheta} z}}{\alpha}\widehat{m}(\vartheta z) z^{-\log_\vartheta \alpha - 1} dz.$$

Note that here, $\frac{e^{-\overline{\vartheta} z}}{\alpha}\widehat{m}(\vartheta z)$ is merely considered as an auxiliary function. 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 through the recursion formula on the coefficient m_n. As an example, the classical case $\vartheta = \frac{1}{3}$ gives (apart from small fluctuations),

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

The fact that in an asymptotic formula the generating function itself, evaluated at a certain point, appears is not at all uncommon in combinatorial analysis.

References

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

[2] Flajolet (P.) and Sedgewick (R.). – Mellin transforms and asymptotics: finite differences and Rice's integral. *Theoretical Computer Science*, n° 144, 1994, pp. 101–124.

[3] Grabner (P.J.) and Prodinger (H.). – Asymptotic analysis of the moments of the Cantor distribution. *Statistics & Probability letters*, n° 26, 1996, pp. 243–248.

[4] Knopfmacher (A.) and Prodinger (H.). – Explicit and asymptotic formulae for the expected values of the order statistics of the Cantor distribution. *Statistics & Probability letters*, n° 27, 1996, pp. 189–194.

[5] Lad (F.R.) and Taylor (W.F.C.). – The moments of the Cantor distribution. *Statistics & Probability letters*, n° 13, 1992, pp. 307–310.

[6] Prodinger (H.). – The Cantor-Fibonacci distribution. *Applications of Fibonacci numbers*, 1998. – To appear.