Turkish Journal of Analysis and Number Theory, 2018, Vol. 6, No. 2, 57-60
Available online at http://pubs.sciepub.com/tjant/6/2/5

DOI:10.12691/tjant-6-2-5

A Necessary Condition for the Existence of the Asymptotic Density in Kolakoski Sequence

Abdallah Hammam*

Département de Mathématiques, Faculté des sciences, Université Moulay Ismaïl, 50020 Meknès, Morocco
*Corresponding author: a.hammam@fs.umi.ac.ma

Received January 08, 2018; Revised March 03, 2018; Accepted April 24, 2018

Abstract We investigate here the Oldenburger-Kolakoski sequence $(K_n)_{n\geq 1}$ with $K_1 = 1$. In the first part, we give some expressions of the discrepancy function $\delta(n)$ representing the difference between 2s and 1s in $K_1K_2...K_n$. The discrepancy could be interpreted as a perturbation of a certain equilibrium. Our main result is a necessary and sufficient condition for the existence of the asymptotic density. In the last section, we present an algorithm to generate the sequence terms and a formula for the term K_n.

Keywords: *Kolakoski sequence, recursive formula, asymptotic density*

Cite This Article: Abdallah Hammam, "A Necessary Condition for the Existence of the Asymptotic Density in Kolakoski Sequence." *Turkish Journal of Analysis and Number Theory*, vol. 6, no. 2 (2018): 57-60. doi: 10.12691/tjant-6-2-5.

1. Introduction

The Kolakoski sequence $(K_n)_{n\geq 1}$ with $K_1 = 1$ defines the unique infinite word $K_1K_2...$ which does not change by the Run Length Encoding operator Δ. It starts like this: 122112122122112... Many questions asked by Kimberling [2] about this sequence, are still open. For example, it is conjectured by Keane [1] that asymptotically, there is as many one's as two's, which means that the limiting density of 2*s* is $\frac{1}{2}$. Other information in concern with this sequence can be found in the Sloane's OEIS [3].

2. Notation

We now introduce some definitions and notation.

We let Δ and Δ_1^{-1} denote respectively the run-length encoding operator, defined by the following example, $\Delta(112212112) = 221121$ and its inverse. The Kolakoski infinite K word can be seen at the same time as a fixed point of Δ and of Δ_1^{-1} for $n > 0$. The density of 2*s* in $K_1K_2...K$ is given by

$$\rho_n = \frac{\left|\left\{ j \in \mathbb{N} : 1 \leq j \leq n \text{ and } K_j = 2 \right\}\right|}{n}.$$

We define the successive partial sums by

$$S_n = S_{1,n} = \sum_{j=1}^{n} K_j = n(1+\rho_n)$$

and

$$(\forall i \geq 1) S_{i+1,n} = S_{S_{i,n}}.$$

We will use the following cardinals $|1|_n$ and $|2|_n$ are respectively the numbers of 1*s* and 2*s* in $K_1K_2...K_n$ and $\left|K^{a_1a_2a_3}\right|_n = \left|\left\{ j \in \{1,2,...,n\} : K_j = K \right\}\right|$ with j, S_j, $S_{2,j}$ having respectively the same parity than a_1, a_2, a_3.

For instance,

$\left|2^2\right|_n$ = is the number of 2s in $K_1K_2...K_n$ with even indices.

$\left|1^{21}\right|_n$ is the number of 1s in $K_1K_2...K_n$ with even index and odd partial sum.

$\left|2^{212}\right|_n$ is the number of 2s in $K_1K_2...K_n$ with even index j, odd partial sum S_j and even $S_{2,j}$.

3. Some Expressions of the Discrepancy

The discrepancy function $\delta(n)$ represents the difference between the number of 2*s* and the number of 1*s* in the word $K_1K_2...K_n$. It can be expressed as follows

$$\delta(n) = |2|_n - |1|_n = \sum_{j=1}^{n} (-1)^{K_j} = (2\rho_n - 1)n.$$

Proposition 1.

$$\delta(S_n)=\sum_{j=1}^{n}(-1)^j K_j=(2\rho_{S_n}-1)(1+\rho_n)n$$

$$\delta(S_{2,n})=\sum_{j=1}^{n}\frac{1+(-1)^{K_j}}{2}\frac{3+(-1)^j}{2}(-1)^{S_j}$$
$$=(2\rho_{S_{2,n}}-1)(1+\rho_{S_n})(1+\rho_n)n$$

$$\delta(S_{3,n})=\sum_{j=1}^{n}\frac{1-(-1)^j}{2}\left(3-K_j\frac{3-(-1)^{S_j}}{2}\right)(-1)^{S_{2,j}}.$$

Proof. If we replace n by S_n in the expression of $\delta(n)$, use the fact that

$$\Delta_1^{-1}(K_1K_2...K_n)=K_1K_2...K_{S_n},$$

we get

$$\delta(S_n)=|2|_{S_n}-|1|_{S_n}$$
$$=2\left(\left|2^2\right|_n-\left|2^1\right|_n\right)+\left(\left|1^2\right|_n-\left|1^1\right|_n\right)$$
$$=\sum_{j=1}^{n}2(K_j-1)(-1)^j+(2-K_j)(-1)^j=\sum_{j=1}^{n}(-1)^j K_j$$

$$\delta(S_{2,n})=2\left(\left|2^2\right|_{S_n}-\left|2^1\right|_{S_n}\right)+\left(\left|1^2\right|_{S_n}-\left|1^1\right|_{S_n}\right)$$
$$=2\left(\left|1^{22}\right|_n-\left|1^{21}\right|_n\right)+\left(\left|1^{12}\right|_n-\left|1^{11}\right|_n\right)$$
$$=\sum_{j=1}^{n}\frac{1+(-1)^{K_j}}{2}(-1)^{S_j}\left(2\frac{1+(-1)^j}{2}+\frac{1-(-1)^j}{2}\right)$$
$$=\sum_{j=1}^{n}\frac{1+(-1)^{K_j}}{2}\frac{3+(-1)^j}{2}(-1)^{S_j}$$

$$\delta(S_{3,n})=2\left(\left|1^{22}\right|_{S_n}-\left|1^{21}\right|_{S_n}\right)+\left(\left|1^{12}\right|_{S_n}-\left|1^{11}\right|_{S_n}\right)$$
$$=2\left(\left|1^{122}\right|_n+\left|2^{122}\right|_n+\left|2^{111}\right|_n-\left|1^{121}\right|_n-\left|2^{121}\right|_n-\left|2^{112}\right|_n\right)$$
$$+\left(\left|1^{112}\right|_n+\left|2^{112}\right|_n+\left|2^{121}\right|_n-\left|1^{111}\right|_n-\left|2^{111}\right|_n-\left|2^{122}\right|_n\right)$$
$$=2\left(\left|1^{122}\right|_n-\left|1^{121}\right|_n\right)+\left(\left|1^{112}\right|_n-\left|1^{111}\right|_n\right)$$
$$+\left(\left|2^{122}\right|_n-\left|2^{121}\right|_n\right)+\left(\left|2^{111}\right|_n-\left|2^{112}\right|_n\right)$$
$$=\sum_{j=1}^{n}\frac{1-(-1)^j}{2}(-1)^{S_{2,j}}\left(\begin{array}{l}\frac{1+(-1)^{S_j}}{2}(3-K_j)\\+\frac{1-(-1)^{S_j}}{2}(3-2K_j)\end{array}\right)$$
$$=\sum_{j=1}^{n}\frac{1-(-1)^j}{2}\left(3-K_j\frac{3-(-1)^{S_j}}{2}\right)(-1)^{S_{2,j}}.$$

From here, we conclude that if the limiting density $(\lim_{n\to\infty}\rho_n)$ exists and is not $\frac{1}{2}$, then the discrepancies $\delta(S_n)$ and $\delta(S_{2,n})$ should have the same sign for large enough n.

4. The Discrepancy is a Perturbation of an Equilibrium

In this section, we will consider the integers n such that $n, S_n, S_{2,n}$ are even

Proposition 2.

$$\sum_{j=1}^{n}(-1)^j=0$$

$$\delta(n)=\sum_{j=1}^{n}(-1)^j(-1)^{K_j+j}$$

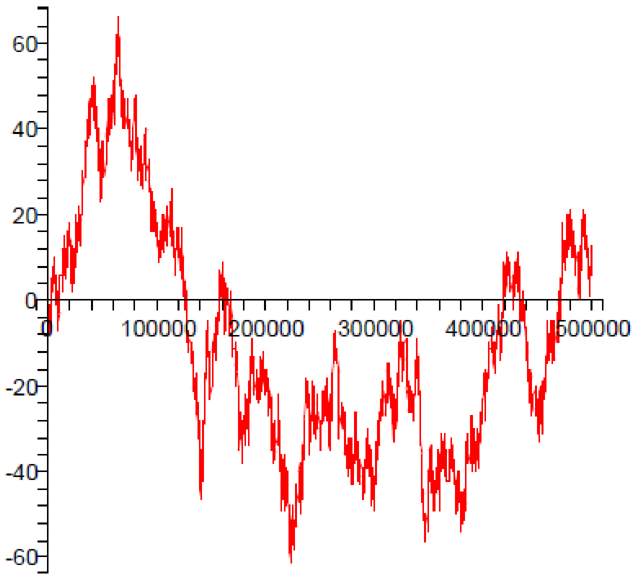

Figure 1. The sign of the discrepancy changes

Proposition 3.

$$\sum_{j=1}^{n}\frac{1-(-1)^{K_j}}{2}(-1)^{S_j}=0$$

$$\delta(S_n)=\sum_{j=1}^{n}\frac{1-(-1)^{K_j}}{2}(-1)^{S_j}(-1)^{S_j+j+1}.$$

Proposition 4.

$$\sum_{j=1}^{n}\frac{1-(-1)^{K_j}}{2}\frac{1-(-1)^j}{2}(-1)^{S_{2,j}}=0$$

$$\delta(S_{2,n})=\sum_{j=1}^{n}\frac{1-(-1)^{K_j}}{2}\frac{1-(-1)^j}{2}(-1)^{S_{2,j}}(-1)^{S_{2,j}+S_j+1}$$

The proof of these propositions is based on the identity $\left|1^2\right|_n + \left|2^2\right|_n = \left|1^1\right|_n + \left|2^1\right|_n$ where we replace n by S_n and $S_{2,n}$.

The discrepancy function can be seen as perturbation of some equilibrium. Hence, $\delta(n)$, $\delta(S_n)$ and $\delta(S_{2,n})$ are three perturbations of three equilibriums and there is no reason they will have same sign. The change of the sign of the discrepancy means that if the density ρ_n has a limit, it will be $\frac{1}{2}$.

5. A Necessary Condition for the Existence of the Asymptotic Density

Let $\beta_n = \frac{\left|2^2\right|_n}{n}$ and $\gamma_n = \frac{\left|1^{22}\right|_n}{n}$ be the respective densities of 2^2 and 1^{22} in $K_1K_2...K_n$.

Proposition 5.

$$\lim_{n\to+\infty} \rho_n = \rho \Leftrightarrow \lim_{n\to+\infty} \beta_n = \rho(\rho+1) - \frac{1}{2}$$

$$\lim_{n\to+\infty} \rho_n = \rho \Leftrightarrow \lim_{n\to+\infty} \gamma_n = \rho\left(\rho(\rho+1) - \frac{1}{2}\right).$$

In particular,

$$\lim_{n\to+\infty} \rho_n = \frac{1}{2} \Leftrightarrow \lim_{n\to+\infty} \beta_n = \frac{1}{4} \Leftrightarrow \lim_{n\to+\infty} \gamma_n = \frac{1}{8}.$$

Proof. Using the definition of the 2*s* density, we have

$$\rho_{S_n} = \frac{\left|2\right|_{S_n}}{S_n} = \frac{2\left|2^2\right|_n + \left|1^2\right|_n}{n(1+\rho_n)}$$

but $\left|2^2\right|_n + \left|1^2\right|_n = \frac{n}{2} + \frac{(-1)^n - 1}{4}$

thus

$$\rho_{S_n} = \frac{\left|2^2\right|_n + \frac{n}{2} + \frac{(-1)^n - 1}{4}}{n(1+\rho_n)}.$$

If we assume that the asymptotic density exist and $\lim_{n\to+\infty} \rho_n = \rho$, then, as a subsequence, we should have $\lim_{n\to+\infty} \rho_{S_n} = \rho$ and consequently

$$\lim_{n\to+\infty} \frac{\left|2^2\right|_n}{n} = \rho(\rho+1) - \frac{1}{2}.$$

On the other hand

$$\frac{\left|2^2\right|_{S_n}}{S_n} = \frac{\left|2^2\right|_n + \left|1^{22}\right|_n}{n(1+\rho_n)}.$$

By the same, we will have

$$\lim_{n\to+\infty} \frac{\left|2^2\right|_{S_n}}{S_n} = \rho(\rho+1) - \frac{1}{2},$$

which gives the condition

$$\lim_{n\to+\infty} \frac{\left|1^{22}\right|_n}{n} = \rho\left(\rho(\rho+1) - \frac{1}{2}\right).$$

6. An Interesting Algorithm to Generate the Sequence Terms

A natural way to generate the sequence (K_n) is the one based on a first integration using the inverse run-length encoding operator Δ_1^{-1}. The blue ones create ones while the red ones create twos. The blue twos will generate blocs of two ones and the red twos will give blocs of two twos.

1 2 2 1 1 2 ....
1 22 11 2 1 22 ....

To make the algorithm faster, we will use the third integration defined by Δ_1^{-3}. This is illustrated as follows

1 2 2 1 1 2 ....
1 221121 221 2211 2 112212 ....

For $n = 1,2,3,...$, the multi-run generated by K_n after three integrations, will depend on the parity of n, S_n and $S_{2,n}$. Its length will be $S_{3,n} - S_{3,n-1}$.

We can improve the speed of the algorithm with the Δ_1^{-4} integration but the computations become complex.

1 2 2 1 1 2 ....
1 221121221 22112 112212 11 212211211 ....

So, with Δ_1^{-3}, if we know $K_1, K_2..., K_m$, we can generate $K_{m+1}, K_{m+2}, ..., K_q$ where $q \approx \frac{27m}{8}$.

7. A New Formula for the n^{th} Term K_n

Lemma 6. *For every* $n > 1$:

$$S_{2,n} = S_{2,n-1} + K_n \frac{3+(-1)^n}{2}$$

$$S_{3,n} = S_{3,n-1} + K_n \frac{3+(-1)^n}{2} \frac{3+(-1)^{S_n}(2-K_n)}{2}$$

Proof. By definition, we have

$$2S_n = 3n + \delta(n)$$

replacing n by S_n, we get $2S_{2,n} = 3S_n + \delta(S_n)$ and $2S_{2,n-1} = 3S_{n-1} - \delta(S_{n-1})$. Using the expression of $\delta(S_n)$ given above, the difference yields to

$$S_{2,n} = S_{2,n-1} + K_n \frac{3+(-1)^n}{2}.$$

If we replace again n by $S_{2,n}$, the first identity becomes

$$2S_{3,n} = 3S_{2,n} + \delta(S_{2,n})$$

which yields to

$$2S_{3,n} = 2S_{2,n-1} + 3(S_{2,n} - S_{2,n-1}) + \delta(S_{2,n}) - \delta(S_n)$$

and finally gives

$$S_{3,n} = S_{3,n-1} + K_n \frac{3+(-1)^n}{2} \frac{3+(-1)^{S_n}(2-K_n)}{2}.$$

Proposition 7. *Let $n \in \mathbb{N}$ and p the unique integer such that*

$$S_{3,p-1} < n \le S_{3,p}.$$

If we assume known the terms $K_1, K_2, ..., K_p$, then, using the recursive formulas above, we can compute $S_1, S_{2,p}$ and $S_{3,p}$ and predict the word generated by K_p, represented by the number N_p, after three integrations. The difference $S_{3,p} - n$ gives the position and the following value of K_n:

$$K_n = \left\lfloor \frac{N_p - 10^{S_{3,p}-n+1} * \left\lfloor \frac{N_p}{10^{S_{3,p}-n+1}} \right\rfloor}{10^{S_{3,p}-n}} \right\rfloor$$

Where N_p is a number which depends on the parity of K_p, p, S_p and $S_{2,p}$ as illustrated by the table below.

Table 1. The number N_p generated by K_p, p, S_p and Δ_1^{-3}

$K_p^{p,S_p,S_{2,p}}$	1^{111}	1^{112}	1^{122}	1^{121}	1^{211}	1^{212}	1^{222}	1^{221}
N_p	1	2	22	11	21	12	1122	2211
$K_p^{p,S_p,S_{2,p}}$	2^{111}	2^{112}	2^{122}	2^{121}	2^{211}	2^{212}	2^{222}	2^{221}
N_p	221	112	122	211	221121	112212	121122	212211

8. Concluding Remarks

The question about the existence of the limiting density is still unanswered. We just replaced it by an equivalent one concerning the density of even twos. The expressions of the discrepancy presented here show the complexity of Kolakoski sequence and justify why the answer to Keane conjecture is not easy. We believe that is not possible to have an explicit expression of the n^{th} term K_n. All one can do, is to find K_n from some previous terms $K_1, K_2, K_3, ...K_p$ with p as small as possible. We proved that $p \approx \frac{8}{27}n$, improving the former result $p \approx \frac{4}{9}n$. In the end, the advantage of the algorithm used to generate $K_1, K_2, K_3, ...$ is that, it could be easily generalized and made faster and faster.

Acknowledgments

The author would like to express his gratitude to the referees for their careful reading of the manuscript and for giving me some of their precious time. I also thank some of my friends for fruitful discussions.

References

[1] Keane and C. Series (eds.), Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, Oxford University Press, 1991, pp. 35-70.

[2] C.Kimberling,http://faculty.evansville.edu/ck6/integer/index.html.

[3] N. J. A. Sloane, The On-line Encyclopedia of Integer Sequences, published electronically at http://oeis.org.