RAIRO-Theor. Inf. Appl. 58 (2024) 11
https://doi.org/10.1051/ita/2024005

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

ON A PROBABILISTIC EXTENSION OF THE OLDENBURGER–KOLAKOSKI SEQUENCE

Chloé Boisson[1], Damien Jamet[2,*] and Irène Marcovici[3]

Abstract. The Oldenburger–Kolakoski sequence is the only infinite sequence over the alphabet $\{1,2\}$ that starts with 1 and is its own run-length encoding. In the present work, we take a step back from this largely known and studied sequence by introducing some randomness in the choice of the letters written. This enables us to provide some results on the convergence of the density of 1's in the resulting sequence. When the choice of the letters is given by an infinite sequence of i.i.d. random variables or by a Markov chain, the average densities of letters converge. Moreover, in the case of i.i.d. random variables, we are able to prove that the densities even almost surely converge.

Mathematics Subject Classification. 11K31, 11K36.

Received December 14, 2022. Accepted February 23, 2024.

1. Introduction

The Oldenburger–Kolakoski sequence 122112122122112 ... introduced by R. Oldenburger [1] and lately mentioned by W. Kolakoski [2] is the unique sequence $x_1x_2x_3\dots$ over the alphabet $\{1,2\}$ with $x_1 = 1$ and whose k-th block has length x_k for $k \in \mathbb{N}^\star$.

In [3], M.S. Keane asked whether the density of 1's in this sequence is 1/2. In [4], V. Chvátal showed that the upper density of 1's (resp. 2's) is less than 0.50084. This bound has been slightly improved by M. Rao but Keane's question still stands: "Is the density of 1's in Oldenburger–Kolakoski sequence defined and equal to 0.5?"

By definition, the Oldenburger–Kolakoski sequence $\mathcal{O} = (x_n)_{n\in\mathbb{N}^\star}$ is a fixed point of the run-length encoding operator denoted Δ:

$$\begin{aligned} \mathcal{O} \quad &= \underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\cdots = 1^1 2^2 1^2 2^1 1^1 2^2 1^1 2^2 1^2 \dots \\ \Delta(\mathcal{O}) &= \;1\;\;2\;\;2\;\;1\;\;1\;\;2\;\;1\;\;2\;\;2 \end{aligned} \tag{1.1}$$

$$\mathcal{O} = 1^{x_1}2^{x_2}1^{x_3}2^{x_4}1^{x_5}2^{x_6}1^{x_7}2^{x_8}1^{x_9} = \prod_{n\in\mathbb{N}} (1^{x_{2n+1}}2^{x_{2n+2}}) \tag{1.2}$$

Keywords and phrases: Kolakoski–Oldenburger – combinatorics on words – random variable – Markov chain.

[1] École Normale Supérieure de Lyon, 15 parvis René Descartes, F-69342 Lyon, France.
[2] Université de Lorraine, Loria, UMR 7503, Vandœuvre-lès-Nancy F-54506, France.
[3] Univ Rouen Normandie, CNRS, Normandie Univ, LMRS UMR 6085, F-76000 Rouen, France

[*] Correspondig author: damien.jamet@loria.fr

In [1], R. Oldenburger refers to sequences over an alphabet Σ as *trajectories* and refers to the sequence $\Delta(w)$ as the *exponent trajectory* of the trajectory w. He stated that "a periodic trajectory is distinct from its exponent trajectory" (Thm. 2, [1]) and, therefore, the Oldenburger–Kolakoski sequence is not periodic.

The Oldenburger–Kolakoski sequence is also connected to differentiable words, C^∞-words and smooth words [5–7]. A sequence w over the alphabet $\Sigma \subset \mathbb{N}^\star$ is *differentiable* if and only if $\Delta(w)$ is also defined over the same alphabet Σ. The sequence $\Delta(w)$ is called the *derivative sequence* of w[7]. A C^∞*-word*, or *smooth word*, is an infinitely differentiable sequence. Obviously, the Oldenburger–Kolakoski sequence is a C^∞-word since it is a fixed-point of the run-length encoding operator Δ.

Although not answering Keane's question fully, F.M. Dekking established connections between possible combinatorial properties of the Oldenburger–Kolakoski sequence [7]: if the Oldenburger–Kolakoski sequence is closed by complementation (that is, if w occurs in $\mathcal{O}$ then so does $\widetilde{w}$ with $\widetilde{1} = 2$ and $\widetilde{2} = 1$) then it is recurrent (any word that occurs in $\mathcal{O}$ does so infinitely often) (Prop. 1, [7]). Moreover, the Oldenburger–Kolakoski sequence is closed by complementation if and only if it contains every finite C^∞-word (Prop. 2, [7]).

A few years later, A. Carpi stated that the Oldenburger–Kolakoski sequence contains only a finite set of squares (words of the form xx where x is not empty) and does not contain any cube (word of the form xxx where x is not empty) [8]. Hence, since $\mathcal{O}$ contains only squares of bounded length then it cannot be the fixed point of a nondegenerated morphism: the image of a square $w = xx$ by such a morphism is still a square longer than w.

There are several ways to extend the definition of the Oldenburger–Kolakoski sequence, depending on whether one wants to preserve the fixed point property or to follow the construction scheme without requiring the resulting sequence to be a fixed point for the run-length encoding operator Δ. For instance, one can deal with other alphabets and thus construct *Generalized Oldenburger–Kolakoski* sequence (GOK-sequence for short) as follows: for any pair (a, b) of non-zero natural numbers, there exists a unique fixed point $\mathcal{O}_{a,b}$ of Δ over the alphabet $\{a, b\}$ starting with a. Also, according to this notation, the original Oldenburger–Kolakoski sequence is $\mathcal{O}_{1,2}$. For instance, if $a = 1$ and $b = 3$, the first terms of $\mathcal{O}_{1,3}$ are:

$$\mathcal{O}_{1,3} = \underbrace{1}_{1}\underbrace{333}_{3}\underbrace{111}_{3}\underbrace{333}_{3}\underbrace{1}_{1}\underbrace{3}_{1}\underbrace{1}_{1}\underbrace{333}_{3}\cdots = 1^1 3^3 1^3 3^3 1^1 3^1 1^1 3^3 \ldots \tag{1.3}$$

A significant result is, unlike the case of the original Oldenburger–Kolakoski sequence, that the densities of 1's in $\mathcal{O}_{1,3}$ and $\mathcal{O}_{3,1}$ are known and approximately 0.3972 [9].

Generalized Oldenburger–Kolakoski sequences are also connected with smooth words over arbitrary alphabets [10, 11]. As for the (Generalized) Oldenburger–Kolakoski sequences, the properties of smooth words are better known for alphabets with letters of the same parity: for instance, while the frequency of letters in an infinite smooth word over $\{1, 2\}$ is still unsolved, in [11] the authors showed that the frequency of letters for extremal smooth words (for the lexicographic order) over the alphabet $\{a, b\}$, where a and b are both even, is 0.5. They also computed the frequency for extremal smooth words over alphabets of type $\{1, b\}$, where b is odd. Moreover, if a and b have the same parity, then every infinite smooth word over the alphabet $\{a, b\}$ is recurrent [11]. Also, if a and b are both odd, then every infinite smooth word is closed under reversal but not under complementation [11]. On the other hand, if a and b are both even, then the extremal smooth words over the alphabet $\{a, b\}$ are neither closed under reversal nor closed under complementation [11].

For a more detailed survey on the Oldenburger–Kolakoski sequence and on generalizations over arbitrary two letter alphabets, see [12].

In the present paper, we extend the Oldenburger–Kolakoski sequences by introducing the following notion of self-descriptive sequences. Given a sequence $T = t_1 t_2 t_3 \ldots$ over the alphabet $\{a, b\}$ (with $a, b \in \mathbb{N}^\star$ and $a \neq b$) called the *directive sequence*, we say that the *self-descriptive sequence* directed by T is the sequence $\mathcal{O}_T = x_1 x_2 x_3 \ldots$ over $\{a, b\}$ satisfying:

$$\mathcal{O}_T = x_1 x_2 x_3 \ldots = t_1^{x_1} t_2^{x_2} t_3^{x_3} t_4^{x_4} \ldots$$

Observe that the (original) Oldenburger–Kolakoski corresponds to the special case where the directive sequence is the periodic sequence $T = (12)^\infty$. We investigate the potential relationships between densities in T and (if any) in $\mathcal{O}_T$, for both deterministic and random directive sequences.

The present paper is organized as follows:

- In Section 2, we introduce the notion of self-descriptive sequence $\mathcal{O}_T$ directed by a sequence T over an alphabet $\{a, b\}$, and provide some related definitions.
- In Section 3, we deal with the case where the terms of the directive sequence $\mathbb{T} = (T_n)_{n\in\mathbb{N}^\star}$ are drawn randomly and independently. We show that, given $p \in]0, 1[$, if $P(T_n = 1) = p$ for all $n \geq 1$, then the density of 1's in $\mathcal{O}_\mathbb{T}$ tends to p, almost surely (Thm. 3.5).
- In Section 4, we consider the case where the directive sequence $\mathbb{T} = (T_n)_{n\in\mathbb{N}^\star}$ is given by a Markov chain with initial value 1 and with transition probability $p \in]0, 1[$ from 1 to 2 and from 2 to 1. We then show that, on average, the density of 1's in $\mathcal{O}_\mathbb{T}$ tends towards 1/2 (Thms. 4.1 and 4.4). The motivation behind this work is to get closer to the Oldenburger–Kolakoski sequence: indeed, when p tends towards 1, the directive sequence tends to the periodic sequence $(12)^\infty$ and thus $\mathcal{O}_\mathbb{T}$ converges to the Oldenburger–Kolakoski sequence.
- In both previous cases, the densities of the letters in the directive sequence $\mathbb{T}$ are, almost certainly or on average, identical to those in $\mathcal{O}_\mathbb{T}$. If the directive sequence T is periodic, simulations also suggest that T and the sequence directed by T have the same densities, and this seems to be true for even larger families of directive sequences. In Section 5, we provide, on the contrary, an example of deterministic sequences T and $\mathcal{O}_T$ that cannot share the same densities (Thm. 5.2).

2. Extending the construction scheme to any directive sequence

2.1. Notion of directive sequence

In the construction scheme of a Generalized Oldenburger–Kolakoski sequence, the blocks of $\mathcal{O}_{a,b}$ are composed, alternatively, of a's and b's as shown in (1.1) when $a = 1$ and $b = 2$ and in (1.3) when $a = 1$ and $b = 3$. In other words, the n^{th} block of $\mathcal{O}_{a,b} = (x_n)_{n\in\mathbb{N}^\star}$ is of length x_n and is filled with the letter τ_n, where $(\tau_n)_{n\in\mathbb{N}^\star} = (ab)^\omega$.

Beyond the specific case of the sequence $(ab)^\omega$, this construction scheme can be extended to any finite or infinite sequence $T = t_1t_2\ldots$ over $\{a, b\}$ as follows: "the n^{th} block of $\mathcal{O}_T = x_1x_2\ldots$ is of length x_n and is filled with the letter t_n" (see Program 1).

```
def O(T):
  X = []
  k = 0
  for x in T:
    X += [x] # concatenate 'x' at the end of X
    X += [x]*(X[k]-1) # concatenate 'X[k]-1' copie(s) of x
    k += 1
  return X

```

Program 1: `Python` program constructing $\mathcal{O}_T$ from T.

Whether T is finite or infinite, we say that the sequence $\mathcal{O}_T$ is *directed* by the sequence T, and that T is a *directive sequence* of $\mathcal{O}_T$. For instance, the sequence $\mathcal{O}_{a,b}$ is directed by $(ab)^\omega$ so that $\mathcal{O}_{a,b} = \mathcal{O}_{(ab)^\omega}$. In particular, the Oldenburger–Kolakoski sequence $\mathcal{O}$ is directed by $(12)^\omega$.

Notice that, in general, the directed sequence $\mathcal{O}_T$ may no longer be a fixed point of the operator Δ.

Let us illustrate the construction of a directed sequence by considering the example of the directive sequence $\mathcal{T} = (\tau_n)_{n\in\mathbb{N}^\star} = 21122\ldots$ We read the digits of $\mathcal{T}$, one by one, from left to right:

Step 1: Since $\tau_1 = 2$, then $\mathcal{O}_\mathcal{T}$ starts with 2. Hence, the first block of $\mathcal{O}_\mathcal{T}$ is of length 2. It follows that $\mathcal{O}_\mathcal{T} = 22\ldots$. Furthermore, since the second term of $\mathcal{O}_\mathcal{T}$ is 2, we also know that the second block of

$\mathcal{O}_{\mathcal{T}}$ is of length 2 without knowing, for the moment, how it will be filled. We will denote:

$$\mathcal{O}_{\tau_1} = \underbracket{22}\underbracket{??}.$$

Step 2: Since $\tau_2 = 1$, we now know that we can fill the second block of $\mathcal{O}_{\mathcal{T}}$ by 1's. Furthermore, knowing the third and the fourth letters of $\mathcal{O}_{\mathcal{T}}$ provides information on the length of its third and its fourth blocks:

$$\mathcal{O}_{\tau_1\tau_2} = \underbracket{22}\underbracket{11}\underbracket{?}\underbracket{?}.$$

Step 3: Since $\tau_3 = 1$, we fill the third block of $\mathcal{O}_{\mathcal{T}}$ with a 1 and it follows that:

$$\mathcal{O}_{\tau_1\tau_2\tau_3} = \underbracket{22}\underbracket{11}\underbracket{1}\underbracket{?}\underbracket{?}.$$

And so on...

Let us go back to the general case of a directive sequence $T = (t_n)_{n\in\mathbb{N}^\star}$. For $n \in \mathbb{N}^\star$, let $w_n = \mathcal{O}_{t_1\dots t_n}$. We denote by $|w_n|$ the length of w_n, including question marks. Observe that it is equal to the sum of the digits 1 and 2 occurring in w_n. For instance, in the example above, $\mathcal{O}_{\tau_1\tau_2\tau_3}$ is of length 7, that is $2+2+1+1+1$.

2.2. Partitions of the set of directive sequences

We now introduce some subsets of directive sequences that will be crucial in the following sections. For this purpose, let us classify the sequences T according to the information they provide on the length of the blocks of $\mathcal{O}_T$: let k and n be two integers such that $1 \le k \le n$ and let $\mathcal{S}_{n,k}$ be the set of sequences $T = (t_n)_{n\in\mathbb{N}^\star}$ such that the length of the block of $\mathcal{O}_T$ containing its n^{th} letter is known when reading t_k but not before. Formally, recalling that $w_n = \mathcal{O}_{t_1\dots t_n}$ for $n \ge 1$ and setting $w_0 = \varepsilon$, namely the empty word, we have:

$$\mathcal{S}_{n,k} = \Big\{T = (t_n)_{n\in\mathbb{N}^\star} \in \{1,2\}^{\mathbb{N}^\star} : |w_{k-1}| < n \text{ and } |w_k| \ge n\Big\}.$$

Since $|w_n| \ge n$, the set $\{\mathcal{S}_{n,k} : k \in [\![1;n]\!]\}$ is a partition of $\{1,2\}^{\mathbb{N}^\star}$. Moreover,

$$\mathcal{S}_{n,n} = \Big\{T = (t_n)_{n\in\mathbb{N}^\star} \in \{1,2\}^{\mathbb{N}^\star} : |w_{n-1}| = n-1\Big\},$$

and $|w_{n-1}| = n-1$ if and only if $t_1 = t_2 = \cdots = t_{n-1} = 1$. Consequently,

$$\mathcal{S}_{n,n} = 1^{n-1} \cdot \{1,2\}^{\mathbb{N}}. \tag{2.1}$$

Furthermore,

$$\mathcal{S}_{n,n-1} = \Big\{T = (t_n)_{n\in\mathbb{N}^\star} \in \{1,2\}^{\mathbb{N}^\star} : |w_{n-2}| < n \text{ and } |w_{n-1}| \ge n\Big\},$$

and $|w_{n-2}| < n$ implies $t_1 = \cdots = t_{n-2} = 1$. Indeed, let k be the smallest integer such that $t_k = 2$. If $k \le n-2$, then $w_k = \underbracket{1}^{k-1}\underbracket{22}\underbracket{??}$ and $|w_k| = k+3$. Hence, $w_{n-2} \ge n+1$. It follows that

$$\mathcal{S}_{n,n-1} = 1^{n-2} \cdot 2 \cdot \{1,2\}^{\mathbb{N}}. \tag{2.2}$$

As an example, let us illustrate the definition of $\mathcal{S}_{8,k}$ with the same sequence $\mathcal{T} = (\tau_n)_{n\in\mathbb{N}^\star} = 21122\dots$ as in the previous paragraph. After having read τ_1, τ_2 and τ_3, we obtain $w_3 = \underbracket{22}\underbracket{11}\underbracket{1}\underbracket{?}\underbracket{?}$ and we still do not know the length of the block of $\mathcal{O}_{\mathcal{T}}$ containing its 8^{th} term. But, whatever the value of τ_4, $|w_4| \ge 8$, so that $\mathcal{T} \in \mathcal{S}_{8,4}$.

More generally, one shows that the sets $\mathcal{S}_{8,k}$, with $k \in [\![1;8]\!]$, partition the set $\{1,2\}^{\mathbb{N}^\star}$ as follows:

$$\begin{aligned} &S_{8,1} = \emptyset, && S_{8,5} = \{1^3\cdot 2, 1^4 \cdot 2\}\cdot\{1,2\}^{\mathbb{N}},\\ &S_{8,2} = 22\cdot\{1,2\}^{\mathbb{N}}, && S_{8,6} = 1^5\cdot 2\cdot\{1,2\}^{\mathbb{N}},\\ &S_{8,3} = \{122, 212\}\cdot\{1,2\}^{\mathbb{N}}, && S_{8,7} = 1^6\cdot 2\cdot\{1,2\}^{\mathbb{N}},\\ &S_{8,4} = \{112, 121, 211\}\cdot\{1,2\}^{\mathbb{N}}, && S_{8,8} = 1^7\cdot\{1,2\}^{\mathbb{N}}. \end{aligned}$$

3. Sequence directed by independent random variables

In the present section, $\mathbb{T} = (T_n)_{n\in\mathbb{N}^\star}$ is a sequence of independent and identically distributed random variables (i.i.d. for short) over the two-letter alphabet $\mathcal{A} = \{1,2\}$, with $\mathbb{P}(T_n = 1) = p$ and $\mathbb{P}(T_n = 2) = 1-p$ for each $n \in \mathbb{N}^\star$, where p is a fixed parameter in $]0,1[$. The sequence $\mathbb{T}$ is thus distributed according to the *product distribution* $(p\delta_1 + (1-p)\delta_2)^{\otimes\mathbb{N}^\star}$. We denote $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes\mathbb{N}^\star}$.

Let $\mathbb{X} = (X_n)_{n\in\mathbb{N}^\star}$ be the sequence directed by $\mathbb{T}$. The sequence $\mathbb{X}$ is a random sequence with *a priori* unknown distribution. Assume that one wants to compute the n^{th} letter X_n of $\mathbb{X}$, for some large integer n. Unless the sequence $\mathbb{T}$ begins with a long succession of 1's (an event which has a low probability to occur), one just has to read the first terms $T_1, \ldots, T_k$ of $\mathbb{T}$, until knowing the length and position of the block containing X_n. Then, one fills that block with 1's with probability p, or with 2's with probability $1-p$, independently of previous draws. The resulting value of X_n obtained that way will have the desired distribution. This leads to the fact that $\lim_{n\to+\infty}\mathbb{P}(X_n = 1) = p$, and we can even use this observation to compute more precisely the value of $\mathbb{P}(X_n = 1)$, as shown in the following proposition.

Proposition 3.1. *If $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes\mathbb{N}^\star}$ with $p \in]0,1[$, then for any $n \geq 2$,*

$$\mathbb{P}(X_n = 1) = p(1 - p^{n-2} + p^{n-1}).$$

Proof. Let us decompose the event $\{X_n = 1\}$ according to the partition $\{\mathcal{S}_{n,k} : k \in [\![1;n]\!]\}$ of $\{1,2\}^{\mathbb{N}^\star}$ introduced in the previous section, and use the equalities (2.1) and (2.2). The following two particular cases are obvious:

$$\mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,n}) = p, \quad \text{and} \quad \mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,n-1}) = 0.$$

Let us now suppose that $\mathbb{T} \in \mathcal{S}_{n,k}$, with $k \leq n-2$ and let us show that:

$$\mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) = p.$$

In fact, it is sufficient to prove that, if $\mathbb{T} \in \mathcal{S}_{n,k}$, then the n^{th} letter of w_k is a question mark and will be determined, conditionally on the event $\mathbb{T} \in \mathcal{S}_{n,k}$, by an independent draw.

By definition, $|w_{k-1}| < n$ and $|w_k| \geq n$. It follows that $T_1 \ldots T_k$ contains at least one 2. Otherwise, we would have $|w_k| = k \leq n-2$ which is in contradiction with $|w_k| \geq n$.

Let i be the smallest integer such that $T_i = 2$. Then, $w_{i-1} = \underbracket{1}^{i-1}$ and $w_i = \underbracket{1}^{i-1}\underbracket{22}\underbracket{??}$. Moreover, for all $j \geq i$, w_j ends with a block of question marks. Indeed when one reads a letter in T, one fills exactly one block and creates at least one block of question marks.

Two cases occur:

1. If $i = k$, then $w_k = \underbracket{1}^{k-1}\underbracket{22}\underbracket{??}$, $|w_k| = k+3$ and $k+2 \leq n \leq k+3$ since $k \leq n-2$ by hypothesis and $|w_k| \geq n$ by definition of $\mathcal{S}_{n,k}$.
2. If $i \leq k-1$, then the last block of w_{k-1} is a block of question marks which ends at a position strictly smaller than n. When reading T_k, we fill exactly one block which already exists in w_{k-1}, so it can't be the block containing the n^{th} term of w_k. Since $\mathbb{T} \in \mathcal{S}_{n,k}$, $|w_k| \geq n$ and the n^{th} letter of w_k is a question mark.

In both cases, the n^{th} letter of w_k is a question mark so that $\mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) = p$.

Finally, we use the formula of total probability:

$$\begin{aligned}\mathbb{P}(X_n = 1) &= \sum_{k=1}^{n} \mathbb{P}(\mathbb{T} \in \mathcal{S}_{n,k}) \times \mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) \\ &= \sum_{k=1}^{n-2} \mathbb{P}(\mathbb{T} \in \mathcal{S}_{n,k}) \times \mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) \\ &\qquad + \mathbb{P}(T_1 \dots T_{n-1} = 1^{n-2} \cdot 2) \times \mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,n-1}) \\ &\qquad + \mathbb{P}(T_1 \dots T_{n-1} = 1^{n-1}) \times \mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,n}) \\ &= \left(1 - p^{n-2}(1-p) - p^{n-1}\right) \times p + p^{n-2}(1-p) \times 0 + p^{n-1} \times p \\ &= p(1 - p^{n-2} + p^{n-1}).\end{aligned}$$

□

As a corollary, we obtain the following convergence of the proportion of 1's in $\mathbb{X}$.

Corollary 3.2. *If* $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$ *with* $p \in]0,1[$, *then*

$$\lim_{n\to+\infty} \frac{\mathbb{E}(|X_1 \dots X_n|_1)}{n} = p$$

Proof. Since $\mathbb{X}$ has values in $\{1,2\}$, we have $|X_1 \dots X_n|_1 = 2n - (X_1 + ... + X_n)$. It follows that

$$\frac{\mathbb{E}(|X_1 \dots X_n|_1)}{n} = 2 - \frac{\sum_{k=1}^{n} \mathbb{E}(X_k)}{n}.$$

By Proposition 3.1, we have $\lim_{n\to+\infty} \mathbb{P}(X_n = 1) = p$ and $\lim_{n\to+\infty} \mathbb{P}(X_n = 2) = 1 - p$. We deduce that $\lim_{n\to+\infty} \mathbb{E}(X_n) = 2 - p$. By Cesàro lemma, we obtain:

$$\lim_{n\to+\infty} \frac{\mathbb{E}(|X_1 \dots X_n|_1)}{n} = 2 - (2-p) = p.$$

□

Each time we run a simulation with $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$, the frequency of 1's in $\mathbb{X}$ seems to converge to p. We thus expect the sequence $|X_1 \dots X_n|_1/n$ to converge almost surely to p, and not only in expectation. Since the random variables $(X_n)_{n\in\mathbb{N}^\star}$ are correlated, we can not directly apply the strong law of large numbers (SLLN) to prove the almost sure convergence of $(X_1 + ... + X_n)/n$. However, the correlations being sufficiently weak, we can apply the following stronger version of the SLLN.

Theorem 3.3 (Lyons [13])**.** *Let* $(Y_n)_{n\in\mathbb{N}^\star}$ *be a sequence of real-valued random variables such that for all* $n \in \mathbb{N}^\star$, $|Y_n| \leq 1$ *and*

$$\forall n, m \in \mathbb{N}^\star,\ \mathbb{E}(Y_m Y_n) \leq \Phi(|n-m|), \qquad \textit{with } \Phi \geq 0 \textit{ and } \sum_{n\geq 1} \frac{\Phi(n)}{n} < +\infty.$$

Then $\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 0$ *almost surely.*

In order to apply Theorem 3.3, let us first prove the following lemma.

Lemma 3.4. *If $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$ with $p \in]0,1[$, then for any $m \geq 1$ and any $n \geq m+2$,*

$$\mathbb{P}(X_m = 2 \text{ and } X_n = 1) = p \times \mathbb{P}(X_m = 2).$$

Proof. Since $\mathcal{S}_{n,n} \cap (X_m = 2) = \emptyset$ and $\mathcal{S}_{n,n-1} \cap (X_m = 2) = \emptyset$, we have

$$\mathbb{P}(X_m = 2 \cap X_n = 1) = \sum_{k=1}^{n-2} \mathbb{P}(X_m = 2 \cap X_n = 1 \cap \mathbb{T} \in \mathcal{S}_{n,k}).$$

It follows that:

$$\begin{aligned}\mathbb{P}(X_m = 2 \cap X_n = 1) = \sum_{k=1}^{n-2} \mathbb{P}(\mathbb{T} \in \mathcal{S}_{n,k}) &\times \mathbb{P}(X_n = 1 \mid \mathbb{T} \in \mathcal{S}_{n,k}) \\ &\times \mathbb{P}(X_m = 2 \mid X_n = 1 \cap \mathbb{T} \in \mathcal{S}_{n,k}).\end{aligned}$$

First, recall that for $k \leq n-2$, $\mathbb{P}(X_n = 1 \mid \mathbb{T} \in \mathcal{S}_{n,k}) = p$. Let us now prove that:

$$\mathbb{P}(X_m = 2 \mid X_n = 1 \cap \mathbb{T} \in \mathcal{S}_{n,k}) = \mathbb{P}(X_m = 2 \mid \mathbb{T} \in \mathcal{S}_{n,k}).$$

It is equivalent to proving that when the events $X_m = 2$ and $\mathbb{T} \in \mathcal{S}_{n,k}$ are not incompatible,

$$\mathbb{P}(X_n = 1 \mid X_m = 2 \cap \mathbb{T} \in \mathcal{S}_{n,k}) = \mathbb{P}(X_n = 1 \mid \mathbb{T} \in \mathcal{S}_{n,k}).$$

Let i be the integer such that the letter X_m is given by T_i. We can decompose the event $\mathbb{T} \in \mathcal{S}_{n,k}$ into the two following cases.

1. If $i > k$, then after reading $T_1 \dots T_k$, we know the size of the blocks containing X_m and X_n but not their content. In other words, there are question marks at positions m and n in w_k. Furthermore, since $n - m \geq 2$, then X_m and X_n do not belong to the same block and their values will be determined by independent draws. In particular, the addtionnal information that $X_m = 2$ does not affect the probability of having $X_n = 1$.
2. If $i \leq k$, then after reading $T_1 \dots T_k$, one already knows whether $X_m = 2$, but still does not know the value of X_n, which will be drawn independently.

In all cases, we have:

$$\mathbb{P}(X_n = 1 \mid X_m = 2 \cap \mathbb{T} \in \mathcal{S}_{n,k}) = \mathbb{P}(X_n = 1 \mid \mathbb{T} \in \mathcal{S}_{n,k}) = p.$$

We deduce that

$$\begin{aligned}\mathbb{P}(X_m = 2 \cap X_n = 1) &= \sum_{k=1}^{n-2} \mathbb{P}(\mathbb{T} \in \mathcal{S}_{n,k}) \times p \times \mathbb{P}(X_m = 2 \mid \mathbb{T} \in \mathcal{S}_{n,k}) \\ &= p \times \mathbb{P}(X_m = 2).\end{aligned}$$

□

We can now state the following theorem.

Theorem 3.5. *If $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$ with $p \in]0,1[$, then*

$$\lim_{n\to+\infty} \frac{|X_1 \ldots X_n|_1}{n} = p \textit{ almost surely.}$$

Proof. In order to apply Theorem 3.3, we need to center the random variables $(X_n)_{n\in\mathbb{N}^\star}$. For $n \in \mathbb{N}^\star$, we thus introduce the random variables $\tilde{X}_n = X_n - (2-p)$, in order to have $|\tilde{X}_n| \le 1$ and $\lim_{n\to+\infty} \mathbb{E}(\tilde{X}_n) = 0$. Now, let us exploit Lemma 3.4 to compute $\mathbb{E}(\tilde{X}_m \tilde{X}_n)$, for $n \ge m+2$. We have

$$\begin{aligned}
&\mathbb{P}(\tilde{X}_m = p \cap \tilde{X}_n = p-1) = p \times \mathbb{P}(X_m = 2), \\
&\mathbb{P}(\tilde{X}_m = p \cap \tilde{X}_n = p) = (1-p) \times \mathbb{P}(X_m = 2), \\
&\mathbb{P}(\tilde{X}_m = p-1 \cap \tilde{X}_n = p-1) = 1 - \mathbb{P}(X_n = 2) - p\,\mathbb{P}(X_m = 2), \\
&\mathbb{P}(\tilde{X}_m = p-1 \cap \tilde{X}_n = p) = \mathbb{P}(X_n = 2) - (1-p)\,\mathbb{P}(X_m = 2).
\end{aligned}$$

Gathering these values and using Proposition 3.1, we obtain

$$\mathbb{E}(\tilde{X}_n \tilde{X}_m) = -(1-p)^2 p^{n-1} \le 0.$$

Let us define a function $\Phi : \mathbb{N}^\star \to \mathbb{R}$ by $\Phi(0) = \Phi(1) = 1$ and for all $k \ge 2, \Phi(k) = 0$. Then $\mathbb{E}(\tilde{X}_m \tilde{X}_n) \le \Phi(|n-m|)$ for all $m, n \in \mathbb{N}^\star$, and Φ satisfies obviously $\Phi \ge 0$ and $\sum_{n\ge 1} \frac{\Phi(n)}{n} < +\infty$. By Theorem 3.3, we deduce that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \tilde{X}_k = 0 \text{ almost surely.}$$

Consequently, $\lim\limits_{n\to+\infty} \dfrac{1}{n} \sum\limits_{k=1}^{n} X_k = 2 - p$ almost surely, and

$$\lim_{n\to+\infty} \frac{|X_1 \ldots X_n|_1}{n} = p \text{ almost surely.}$$

□

To conclude on the case of a directive sequence following a product distribution, let us mention that the previous results can be extended to other alphabets. In particular, Proposition 3.1 is extended as follows.

Proposition 3.6. *Let $a, b \in \mathbb{N}^\star$ with $1 < a < b$, and let $p \in]0,1[$.*

1. *If $\mathcal{A} = \{1, a\}$ and $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_a)^{\otimes \mathbb{N}^\star}$, then*

$$\forall n \ge a, \quad \mathbb{P}(X_n = 1) = p\left(1 - p^{n-a} + p^{n-1}\right)$$

2. *If $\mathcal{A} = \{a, b\}$ and $\mathbb{T} \sim (p\delta_a + (1-p)\delta_b)^{\otimes \mathbb{N}^\star}$, then*

$$\forall n \ge b+1, \quad \mathbb{P}(X_n = a) = p.$$

Proof. 1. We use the same partition as in the proof of Proposition 3.1, but we now distinguish the sets $\mathcal{S}_{n,k}$ for $n - a + 1 \le k \le n$. We have $\mathcal{S}_{n,n} = 1^{n-1} \cdot \{1, a\}^{\mathbb{N}}$ and for $n - a + 1 \le k \le n - 1$, $\mathcal{S}_{n,k} = 1^{k-1} \cdot a \cdot \{1, a\}^{\mathbb{N}}$.

If $n-a+1 \le k \le n-1$, then for the same reason as before, $\mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) = 0$. In all the other cases, $\mathbb{P}(X_n = 1 \,|\, \mathbb{T} \in \mathcal{S}_{n,k}) = p$. Thus,

$$\mathbb{P}(X_n = 1) = p(1 - p^{n-2}(1-p) - \ldots - p^{n-a}(1-p)) = p\left(1 - p^{n-a} + p^{n-1}\right).$$

2. Since $a > 1$, for all $k \in \mathbb{N}^\star$, w_k ends by at least one block of question marks. Thus, except if $n \le b$ (in which case the n^{th} letter might be written during the first step), we are sure that we will know the length of the block containing the n^{th} letter strictly before filling it. As the draws are independent, we deduce that $\mathbb{P}(X_n = a) = p$.

□

4. Sequence directed by a Markov chain

In order to get closer to the deterministic case where a 1 always follows a 2 and *vice versa*, we are now interested in the case of directive sequences which are given by Markov chains.

In the present section, we assume that the directive sequence $\mathbb{T} = (T_n)_{n\in\mathbb{N}^\star}$ is a Markov chain over the alphabet $\{1,2\}$ with initial value $T_1 = 1$ and whose transition probability from 1 to 2 (and from 2 to 1) is $p \in]0,1[$. A large value of p encourages the alternation between 1 and 2. The original case of the Oldenburger–Kolakoski sequence can be viewed as a “limit” case of a Markov chain whose transition probability from 1 to 2 (and from 2 to 1) would be equal to 1.

Theorem 4.1. *Let $p \in]0,1[$ and let $\mathbb{T}$ be a Markov chain over the alphabet $\{1,2\}$ with initial value $T_1 = 1$ and whose transition probability from 1 to 2 (and from 2 to 1) is $p \in]0,1[$. Then*

$$\lim_{n\to+\infty} \mathbb{P}(X_n = 1) = \frac{1}{2}.$$

Proof. Let us first note that for all integers $s > r \ge 1$, $\mathbb{P}(T_s = 1 \mid T_r = 1) = \frac{1}{2}(1 + (1-2p)^{s-r})$ and $\mathbb{P}(T_s = 1 \mid T_r = 2) = \frac{1}{2}(1 - (1-2p)^{s-r})$.

Let $\ell \in \mathbb{N}^\star$ and let $n \ge 8\ell$. Consider the integer $k \in \mathbb{N}^\star$ such that $\mathbb{T} \in \mathcal{S}_{n,k}$. Observe that $k \ge n/4$. Indeed, the longest length of w_k, namely $4k$, is reached for $T_1 \ldots T_k = 2^k$.

If $T_1 \ldots T_{\lfloor n/4 \rfloor}$ contains at least 2ℓ occurences of 2, then so do $T_1 \ldots T_k$ and w_k. This implies that w_k contains at least 2ℓ question marks and thus, at least, ℓ blocks of question marks. Let D be the number of blocks of question marks in w_k. It follows that:

$$\mathbb{P}(D < \ell) \le \mathbb{P}\left(\left|T_1 \ldots T_{\lfloor n/4 \rfloor}\right|_2 < 2\ell\right),$$

and the probability on the right tends to 0 as n tends to $+\infty$.

Let i be such that X_n is given by T_i. For $a \in \{1,2\}$, one has:

$$\frac{1}{2}(1 - |1-2p|^\ell) \le \mathbb{P}(T_i = 1 \mid D \ge \ell \cap T_k = a) \le \frac{1}{2}(1 + |1-2p|^\ell).$$

Since

$$\begin{aligned}
\mathbb{P}(X_n = 1 \mid D \ge \ell) &= \mathbb{P}(T_i = 1 \mid D \ge \ell) \\
&= \mathbb{P}(T_i = 1 \mid D \ge \ell \cap T_k = 1) \times \mathbb{P}(T_k = 1 \mid D \ge \ell) \\
&\quad + \mathbb{P}(T_i = 1 \mid D \ge \ell \cap T_k = 2) \times \mathbb{P}(T_k = 2 \mid D \ge \ell)
\end{aligned}$$

and $\mathbb{P}(T_k = 1 \mid D \geq \ell) + \mathbb{P}(T_k = 2 \mid D \geq \ell) = 1$, it follows that:

$$\mathbb{P}(X_n = 1 \mid D \geq \ell) \in \left[\frac{1}{2}(1 - |1-2p|^\ell);\ \frac{1}{2}(1 + |1-2p|^\ell)\right].$$

We deduce that:

$$\limsup_{n\to+\infty} \mathbb{P}(X_n = 1) \leq \frac{1}{2}(1 + |1-2p|^\ell) \quad \text{and} \quad \liminf_{n\to+\infty} \mathbb{P}(X_n = 1) \geq \frac{1}{2}(1 - |1-2p|^\ell).$$

Finally, by letting ℓ tends to $+\infty$, we obtain $\lim_{n\to+\infty} \mathbb{P}(X_n = 1) = \frac{1}{2}$. □

As a direct consequence of Theorem 4.1, we obtain the following result.

Corollary 4.2. *Let $p \in]0,1[$ and let $\mathbb{T}$ be a Markov chain over the alphabet $\{1,2\}$ with initial value $T_1 = 1$ and whose transition probability from 1 to 2 (and from 2 to 1) is $p \in]0,1[$. Then*

$$\lim_{n\to+\infty} \frac{\mathbb{E}(|X_1 \dots X_n|_1)}{n} = \frac{1}{2}$$

We conjecture that the almost sure convergence can be proved in a similar way with some additional technical difficulties.

Conjecture 4.3. *Let $p \in]0,1[$ and let $\mathbb{T}$ be a Markov chain over the alphabet $\{1,2\}$ with initial value $T_1 = 1$ and whose transition probability from 1 to 2 (and from 2 to 1) is $p \in]0,1[$. Then*

$$\lim_{n\to+\infty} \frac{|X_1 \dots X_n|_1}{n} = \frac{1}{2} \text{ almost surely.}$$

Observe that Theorem 4.1 and Corollary 4.2 extend to other alphabets. In particular, one obtains an identical result over the alphabet $\{1,3\}$: if $\mathbb{T}$ is a Markov chain with transition probability $0 < p < 1$ from 1 to 3 (and from 3 to 1), then the average density of 1's is 1/2.

Theorem 4.4. *Let $a \geq 2$ be an integer, let $p \in]0,1[$ and let $\mathbb{T}$ be a Markov chain over the alphabet $\{1,a\}$ with initial value $T_1 = 1$ and whose transition probability from 1 to a (and from a to 1) is $p \in]0,1[$. Then*

$$\lim_{n\to+\infty} \mathbb{P}(X_n = 1) = \frac{1}{2} \text{ and } \lim_{n\to+\infty} \frac{\mathbb{E}(|X_1 \dots X_n|_1)}{n} = \frac{1}{2}.$$

The statement of Theorem 4.4 is somewhat surprising and unexpected since the densities d_1 and d_3 of the letters 1 and 3 in the sequences $\mathcal{O}_{1,3}$ and $\mathcal{O}_{3,1}$ are respectively $d_1 \approx 0.3972$ and $d_3 \approx 0.6028$. [9]. We will come back to this in the discussion of Section 6.

5. Non conservation of the density

In previous sections, we have studied different cases where the directive sequences are random. In all the cases we considered (sequences of independent and identically distributed random variables, Markovian sequences), the densities of letters of the directed sequence obtained are the same as those in the directive sequence, almost surely.

Simulations also suggest that for any (infinite) periodic sequence T, the density of 1's in the directed sequence $\mathcal{O}_T$ is well-defined and is equal to the density of 1's in T, see Figure 1.

On Figure 1, we have chosen to represent only the data on short prefixes of $\mathcal{O}_T$ so that it remains readable, especially to distinguish the densities in the very first terms of the sequence $\mathcal{O}_T$. However, further experiments

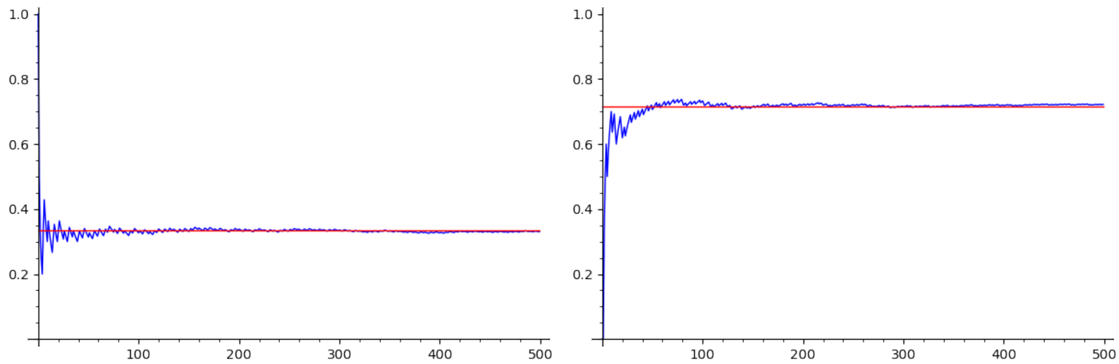

Figure 1: Evolution of the density of 1's in increasingly large prefixes of $\mathcal{O}_{\mathcal{T}}$ for $\mathcal{T} = (122)^\omega$ (left) and $\mathcal{T} = (2112111)^\omega$ (right). The densities seem to converge respectively to 1/3 and to 5/7.

have been carried out on a large number of periodic sequences T and they seem to corroborate our first impression, namely that if the sequence T is periodic, then the densities in $\mathcal{O}_T$ would be the same as those in T. This leads us to state the following conjecture, that extends Keane's conjecture.

Conjecture 5.1. *For any periodic sequence T over the alphabet $\{1, 2\}$, the density of 1's in the directed sequence $\mathcal{O}_T$ is well-defined and is equal to the density of 1's in T.*

Then, a natural question arises: does there exist a directive sequence T over $\{1, 2\}$ for which the density of 1's in T is not conserved in $\mathcal{O}_T$? Obviously, because of Conjecture 5.1, we do not expect to find such a candidate of directive sequence among the periodic ones.

However, we answer this question partially and positively thanks to the fact that the left-to-right reading of $\mathcal{O}_T$ provides the size of the blocks even further to the right (see Sect. 2). In a prospect of building step by step both sequences T and $\mathcal{O}_T$, the knowledge of the length of not yet filled blocks of $\mathcal{O}_T$ could allow us to choose, in a fully arbitrary way, with which letter we will fill them and it could give us the opportunity to force the sequence $\mathcal{O}_T$ to contain relatively more 1's than the sequence T.

More precisely, we construct simultaneously a directive sequence $\mathcal{T} = (\tau_n)_{n\in\mathbb{N}}$ and the sequence $\mathcal{O}_{\mathcal{T}}$ directed by $\mathcal{T}$ as follows: we initialize τ_1 to 2, then $w_1 = \mathcal{O}_{\tau_1} = \llcorner 22 \lrcorner\llcorner ?? \lrcorner$ and from now on, by reading $w_n = \mathcal{O}_{\tau_1 \ldots \tau_n}$ from left to right, we build w_{n+1} by filling the blocks of size 2 with 1's and the blocks of size 1 with 1's and 2's alternatively. The first steps in the simultaneous construction of $\mathcal{T}$ and $\mathcal{O}_{\mathcal{T}}$ are thus as follows (with the notation of Sect. 2):

Step 1: We set $\tau_1 = 2$ and then $w_1 = \llcorner 22 \lrcorner\llcorner ?? \lrcorner$

Step 2: (a) The empty block of w_1 of size 2 is filled with 1's: $w_2 = \llcorner 22 \lrcorner\llcorner 11 \lrcorner\llcorner ? \lrcorner\llcorner ? \lrcorner$
(b) We set $\tau_2 = 1$, that is, $\mathcal{T} = 21\ldots$

Step 3: (a) The next block of w_2 of size 1 is filled with 1 : $w_3 = \llcorner 22 \lrcorner\llcorner 11 \lrcorner\llcorner 1 \lrcorner\llcorner ? \lrcorner\llcorner ? \lrcorner$
(b) We set $\tau_3 = 1$, that is, $\mathcal{T} = 211\ldots$

Step 4: (a) The next block of w_3 of size 1 is filled with 2 : $w_4 = \llcorner 22 \lrcorner\llcorner 11 \lrcorner\llcorner 1 \lrcorner\llcorner 2 \lrcorner\llcorner ? \lrcorner\llcorner ?? \lrcorner$
(b) We set $\tau_4 = 2$, that is, $\mathcal{T} = 2112\ldots$

Step 5: and son on...

Program 2 provides a `Python` function for the construction of $\mathcal{O}_{\mathcal{T}}$ and $\mathcal{T}$.
Let $\mathcal{O}_{\mathcal{T}} = (x_i)_{i\in\mathbb{N}}$ with $x_i \in \{1, 2\}$ for all $i \in \mathbb{N}$, then:

$$|\tau_1 \ldots \tau_n|_1 = |x_1 \ldots x_n|_2 - 1 + \frac{1}{2}|x_1 \ldots x_n|_1 + C_n,$$

with $C_n \in \{0,1\}$: indeed, the number of 1's in $\tau_1 \dots \tau_n$ is equal to the sum of the number of blocks of size 2 in w_n (except the first block of w_n because of the initialisation of w_1) and half of the number of blocks of size 1 in w_n. By construction, the number of blocks of size 1 (resp. of size 2) in w_n is equal to the number of 1's (resp. 2's) in $x_1 \dots x_n$. The constant C_n takes into account the cases where $x_n = 1$ and is the first letter of a block of size 2 in w_n.

```
def Sequences(n) :
  T = [2]
  O_T = [2, 2]
  d = 1 # digit to write in the next block of size 1
  for i in range(1, n) :
    if O_T[i] == 2 :
      T += [1]
      O_T += [1]*2
    else :
      T += [d]
      O_T += [d]
      d = 3-d
  return (T, O_T)

```

Program 2: `Python` function for the simultaneous construction of T and $\mathcal{O}_T$.

Theorem 5.2. *If the density $d_1^{\mathcal{T}}$ of 1's in $\mathcal{T}$ exists, then so does the density $d_1^{\mathcal{O}_\mathcal{T}}$ of 1's in $\mathcal{O}_\mathcal{T}$. Moreover, in that case: $d_1^{\mathcal{T}} = \frac{7-\sqrt{17}}{4} = 0.640\dots$, $d_1^{\mathcal{O}_\mathcal{T}} = \frac{1+\sqrt{17}}{8} = 0.719\dots$ and so $d_1^{\mathcal{T}} \neq d_1^{\mathcal{O}_\mathcal{T}}$.*

As a corollary of the above theorem, if the density $d_1^{\mathcal{T}}$ of 1's in $\mathcal{T}$ exists, then the sequences $\mathcal{T}$ and $\mathcal{O}_\mathcal{T}$ are not periodic. Indeed, if the sequences $\mathcal{T}$ and $\mathcal{O}_\mathcal{T}$ were periodic, then their densities of 1's and 2's would be rational, which is not the case.

Proof. For each $n \in \mathbb{N}^\star$, we have:

$$\Big|\, |w_n|_1 - |\tau_1 \dots \tau_n|_2 - 2(|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n|_2) \,\Big| \leq 1$$

Indeed, to within one unit, each digit 2 of $\tau_1 \dots \tau_n$ gives rise to a single 2 in w_n, and a same quantity $|\tau_1 \dots \tau_n|_2$ of 1's gives rise to a single 1 in w_n, while the rest of them (so $|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n|_2$) give rise to two 1's in w_n. Moreover, the first 2 of $\tau_1 \dots \tau_n$ is the only one to be written twice in w_n, so that we always have exactly $|w_n|_2 = |\tau_1 \dots \tau_n|_2 + 1$. It follows that:

$$\begin{aligned}
\frac{|w_n|_1}{|w_n|} &= \frac{|w_n|_1}{|w_n|_1 + |w_n|_2} \\
&= \frac{|\tau_1 \dots \tau_n|_2 + 2(|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n|_2) + o(n)}{|\tau_1 \dots \tau_n|_2 + 2(|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n|_2) + |\tau_1 \dots \tau_n|_2 + o(n)} \\
&= \frac{2|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n|_2 + o(n)}{2|\tau_1 \dots \tau_n|_1 + o(n)} \\
&= \frac{3|\tau_1 \dots \tau_n|_1 - |\tau_1 \dots \tau_n| + o(n)}{2|\tau_1 \dots \tau_n|_1 + o(n)} \underset{n \to +\infty}{\longrightarrow} \frac{3d_1^{\mathcal{T}} - 1}{2d_1^{\mathcal{T}}},
\end{aligned}$$

assuming that the density $d_1^{\mathcal{T}}$ of 1's in $\mathcal{T}$ exists. We conclude that the density of 1's (resp. of 2's) in $\mathcal{O}_\mathcal{T}$ then also exists, and satisfies

$$d_1^{\mathcal{O}_\mathcal{T}} = \frac{3d_1^{\mathcal{T}} - 1}{2d_1^{\mathcal{T}}}. \tag{5.1}$$

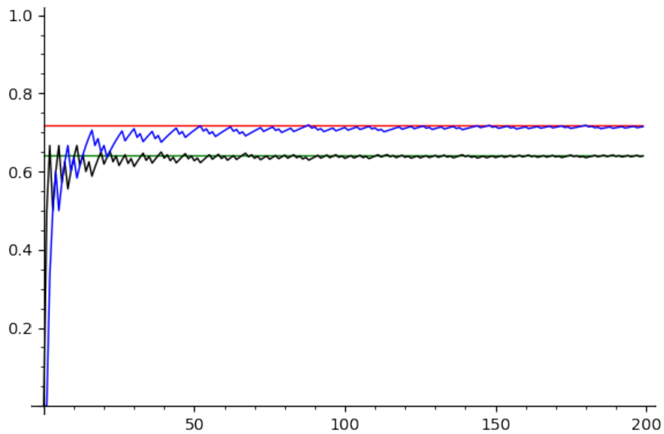

Figure 2: Evolution of the densities of 1's in $\mathcal{O}_T$ (blue) and T (black), where the two sequences are defined by Program 2.

We noticed above that, for each $n \in \mathbb{N}^\star$, $|\tau_1 \dots \tau_n|_1 = |x_1 \dots x_n|_2 - 1 + \frac{1}{2}|x_1 \dots x_n|_1 + C_n$, with $C_n \in \{0, 1\}$. Moreover, we have shown that if $d_1^{\mathcal{T}}$ exists then so do $d_1^{\mathcal{O}_\mathcal{T}}$ and $d_2^{\mathcal{O}_\mathcal{T}}$ an. By tending n towards infinity, we obtain:

$$d_1^{\mathcal{T}} = d_2^{\mathcal{O}_\mathcal{T}} + \frac{1}{2} d_1^{\mathcal{O}_\mathcal{T}} \tag{5.2}$$

By putting together equations (5.1) and (5.2), we deduce $d_1^{\mathcal{T}} = \frac{1+\sqrt{17}}{8}$ and $d_1^{\mathcal{O}_\mathcal{T}} = \frac{7-\sqrt{17}}{4}$.

□

Simulations suggest that the densities are indeed converging to these values, see Figure 2.

6. Conclusion and discussion

Over the alphabet $\{1, a\}$, with $a \in \{2, 3\}$, we have shown that in almost all the sequences directed by an infinite sequence $\mathbb{T} = (T_n)_{n \in \mathbb{N}^\star}$ of i.i.d. random variables with $\mathbb{P}(T_n = 1) = p \in]0, 1[$ and $\mathbb{P}(T_n = a) = 1 - p$, the density of 1's is equal to p. We have also shown that the average density of 1's among all sequences directed by a Markov chain with transition probability $p \in]0, 1[$ from 1 to a and from a to 1 is equal to $1/2$.

Keane's conjecture [3] states that this result can be extended to the deterministic case, namely when $p = 1$, over the alphabet $\{1, 2\}$. On the other hand, over the alphabet $\{1, 3\}$ this result is not extendable to the deterministic case since the density of 1's in $\mathcal{O}_{1,3}$ is close to 0.3972. [9].

When $\mathbb{T}$ is a Markov chain, the closer its transition probability p is to 1, the more likely the sequence $\mathcal{O}_T$ is to share a long prefix with $\mathcal{O}_{1,3}$. Therefore, the closer the transition probability p is to 1, the closer the density of 1's in the sequence $\mathcal{O}_T$ is to that in the sequence $\mathcal{O}_{1,3}$ on a long prefix. However, computer experiments suggest that when the first perturbations in the alternation of 1's and 3's appear in $\mathbb{T}$, the density of 1's in the prefix of $\mathcal{O}_\mathbb{T}$ eventually approaches 0.5 as this prefix gets longer. See Figure 3 for an illustration with $p = 0.99$.

This implies it seems difficult to derive information about the original Oldenburger–Kolakoski sequence $\mathcal{O}_{1,2}$ by letting p tend to 1 in the Markovian case over the alphabet $\{1, 2\}$.

Finally, the study of sequences directed by random sequences on alphabets of more than 2 letters or by random sequences constructed from other distributions is also largely open.

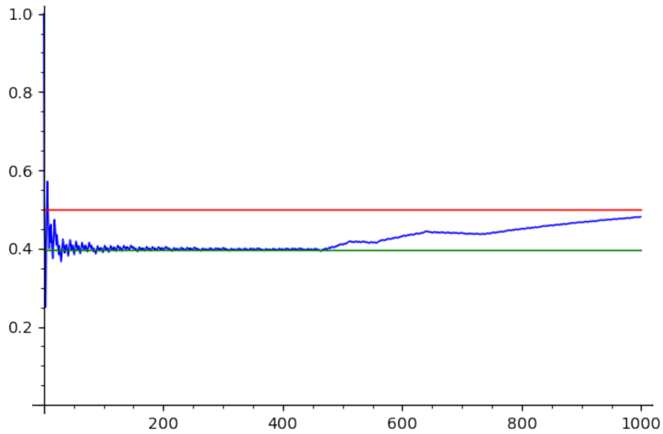

Figure 3: Evolution of the frequency of 1's for a markovian directive sequence on the alphabet $\{1, 3\}$ of parameter $p = 0.99$: the frequency is first close to the one of $\mathcal{O}_{1,3}$ then moves away from it to converge to 1/2.

References

[1] R. Oldenburger, Exponent trajectories in symbolic dynamics. *Trans. Am. Math. Soc.* **46** (1939) 453–466.

[2] W. Kolakoski, Self-generating runs, problem 5304. *Am. Math. Monthly* **73** (1966) 681–682.

[3] M.S. Keane, Ergodic theory and subshifts of finite type, in Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces. Lectures given at the Workshop "Hyperbolic Geometry and Ergodic Theory", held at the International Centre for Theoretical Physics in Trieste, Italy, 17–28 April, 1989. Oxford University Press, Oxford (1991) 35–70.

[4] V. Chvátal, Notes on the Kolakoski sequence. Technical report, DIMACS Technical Report 93-84, December 1993.

[5] V. Berthé, S. Brlek and P. Choquette, Smooth words over arbitrary alphabets. *Theor. Comput. Sci.* **341** (2005) 293–310.

[6] S. Brlek, S. Dulucq, A. Ladouceur and L. Vuillon, Combinatorial properties of smooth infinite words. *Theor. Comput. Sci.* **352** (2006) 306–317.

[7] F.M. Dekking, On the structure of selfgenerating sequences. *Sémin. Théor. Nombres Bordeaux* (1980) 1–6.

[8] A. Carpi, Repetitions in the Kolakovski sequence. *Bull. EATCS* **50** (1993) 194–197.

[9] M. Baake and B. Sing, Kolakoski-(3, 1) is a (deformed) model set. *Can. Math. Bull.* **47** (2004) 168–190.

[10] V. Berthé, S. Brlek and P. Choquette, Smooth words over arbitrary alphabets. *Theor. Comput. Sci.* **341** (2005) 293–310.

[11] S. Brlek, D. Jamet and G. Paquin, Smooth words on 2-letter alphabets having same parity. *Theor. Comput. Sci.* **393** (2008) 166–181.

[12] B. Sing, More Kolakoski sequences. *Integers* **11B** (2011) A14.

[13] R. Lyons, Strong laws of large numbers for weakly correlated random variables. *Michigan Math. J.* **35** (1988) 353–359.