arXiv:2301.01116v1 [cs.DM] 3 Jan 2023

Oldenburger-Kolakoski 序列的概率扩展

Chloé Boisson*、Damien Jamet†和 Irène Marcovici‡

chloe.boisson@ens-lyon.fr, damien.jamet@loria.fr, irene.marcovici@univ-lorraine.fr

2023年1月4日

摘要

Oldenburger-Kolakoski 序列是字母表 $\{1,2\}$ 上唯一一个以 1 开头且与其游程编码相同的无限序列。在本文中，我们从这条广为人知且受到深入研究的序列退后一步，在所写字母的选择中引入一些随机性。由此，我们得以对所得序列中 1 的密度的收敛性给出一些结果。当字母的选择由一个无限的独立同分布随机变量序列或一个马尔可夫链决定时，字母的平均密度会收敛。此外，在独立同分布随机变量的情形下，我们还证明了这些密度几乎处处收敛。

1 引言

由 R. Oldenburger[12] 引入、近期又由 W. Kolakoski [10] 提及的 Oldenburger-Kolakoski 序列 122112122122112... 是字母表 $\{1,2\}$ 上满足 $x_1 = 1$ 且其第 k 个块的长度为 x_k （$k \in \mathbb{N}^\star$）的唯一序列 $x_1x_2x_3\ldots$。

在 [9] 中，M.S. Keane 提出了一个问题：该序列中 1 的密度是否为 $1/2$。在 [7],中，V. Chvátal 证明了 1（分别为 2）的上密度小于 0.50084。M. Rao 稍微改进了这一界限，但 Keane 的问题仍未解决：「Oldenburger-Kolakoski 序列中 1 的密度是否存在且等于 0.5？」

根据定义，Oldenburger-Kolakoski 序列 $\mathcal{O} = (x_n)_{n\in\mathbb{N}^\star}$ 是游程编码算子 Δ 的不动点：

*里昂高等师范学院，15 parvis René Descartes，F-69342，里昂，法国。
†洛林大学，Loria，UMR 7503，旺多厄夫尔莱南西，F-54506，法国。
‡洛林大学，法国国家科学研究中心，法国国家信息与自动化研究院，IECL，F-54000 南锡，法国。

$$\begin{aligned}\mathcal{O} &= \underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\cdots = 1^1 2^2 1^1 2^2 1^1 1^1 2^2 1^1 2^2 1^2 \ldots \quad (1)\\ \Delta(\mathcal{O}) &= \underset{1}{1}\ \underset{2}{2}\ \underset{2}{1}\ \underset{1}{1}\ \underset{1}{2}\ \underset{1}{1}\ \underset{2}{2}\ \underset{2}{1}\ \underset{2}{2}\\ \mathcal{O} &= 1^{x_1}2^{x_2}1^{x_3}2^{x_4}1^{x_5}2^{x_6}1^{x_7}2^{x_8}1^{x_9} = \prod_{n\in\mathbb{N}}(1^{x_{2n+1}}2^{x_{2n+2}}) \quad (2)\end{aligned}$$

在 [12], R. Oldenburger 将字母表 Σ 上的序列称为轨迹，并将序列 $\Delta(w)$ 称为轨迹 w 的指数轨迹。他指出："周期轨迹不同于其指数轨迹"（定理 2, [12]），因此，Oldenburger-Kolakoski 序列不是周期的。

Oldenburger-Kolakoski 序列也与可微词、 C^∞-字和光滑词 [2, 4, 8]有关。字母表 $\Sigma \subset \mathbb{N}^\star$ 上的序列 w 是可微的，当且仅当 $\Delta(w)$ 也定义在同一个字母表 Σ 上。序列 $\Delta(w)$ 称为 w [8]的导数序列 。 C^∞-字，或光滑词，是无限次可微的序列。显然，Oldenburger-Kolakoski 序列是一个 C^∞-字，因为它是游程编码算子 Δ的不动点。

虽然未能完全回答 Keane 的问题，F.M. Dekking 还是确立了 Oldenburger-Kolakoski 序列可能具有的组合性质之间的联系 [8]: 如果 Oldenburger-Kolakoski 序列在取补下封闭（也就是说，如果 w 出现在 $\sim\mathcal{O}$ 中，那么 $\widetilde{w}$ 也会出现，其中 $1=2$ 和 $2=1$），则它是循环的（在 $\mathcal{O}$中出现的任意字都会出现无穷多次）（命题 1, [8]）。此外，Oldenburger-Kolakoski 序列在取补下封闭，当且仅当它包含每个有限的 C^∞-字（命题 2, [8]）。

几年后，A. Carpi 指出，Oldenburger-Kolakoski 序列只包含有限个平方（形如 xx 的字，其中 x 非空），且不包含任何立方（形如 xxx 的字，其中 x 非空） [6]。因此，由于 $\mathcal{O}$ 只包含长度有界的平方，它不可能是非退化态射的不动点：通过这种态射映射平方 $w = xx$ 后，所得仍是一个长度大于 w的平方。

扩展 Oldenburger-Kolakoski 序列的定义有几种方式，具体取决于是否希望保留不动点性质，或遵循构造方案而不要求所得序列是不动点，也就是说，不要求它是不动点的游程编码算子 Δ。例如，可以采用其他字母表，从而按如下方式构造广义 *Oldenburger-Kolakoski* 序列（简称 GOK 序列）：对于任意一对非零自然数 (a, b)，存在唯一的不动点$\mathcal{O}_{a,b}$，它是字母表 $\{a, b\}$ 上的 Δ，且以 a开头。此外，按照这一记号，原始的 Oldenburger-Kolakoski 序列为 $\mathcal{O}_{1,2}$。例如，若 $a = 1$ 且 $b = 3$，则 $\mathcal{O}_{1,3}$ 的前几项为：

$$\mathcal{O}_{1,3} = \underbrace{1}_{1}\underbrace{333}_{3}\underbrace{111}_{3}\underbrace{333}_{3}\underbrace{1}_{1}\underbrace{3}_{1}\underbrace{1}_{1}\underbrace{333}_{3}\cdots = 1^1 3^3 1^3 3^3 1^1 3^1 1^1 3^3 \ldots \quad (3)$$

一个重要结果是，与原始 Oldenburger-Kolakoski 序列的情形不同，$\mathcal{O}_{1,3}$ 和 $\mathcal{O}_{3,1}$ 中 1 的密度已知，约为 0.3972 [1]。

广义 Oldenburger-Kolakoski 序列还与任意字母表 $[3, 5]$ 上的光滑词有关。与（广义）Oldenburger-Kolakoski 序列的情形一样，对于字母奇偶性相同的字母表，光滑词的性质更为人所知：例如，尽管字母表 $\{1, 2\}$ 上无限光滑词中字母的频率仍未解决，但在 $[5]$ 中，作者证明了字母表 $\{a, b\}$ 上极值光滑词（按字典序）的字母频率为 0.5，其中 a 和 b 均为偶数。他们还计算了类型为 $\{1, b\}$ 的字母表上极值光滑词的频率，其中 b 为奇数。此外，如果 a 和 b 奇偶性相同，那么字母表 $\{a, b\}$ 上的每个无限光滑词都是循环的 $[5]$。另外，如果 a 和 b 均为奇数，那么每个无限光滑词都关于反转封闭，但不关于取补封闭 $[5]$。另一方面，如果 a 和 b 均为偶数，那么字母表 $\{a, b\}$ 上的极值光滑词既不关于反转封闭，也不关于取补封闭 $[5]$。

关于 Oldenburger-Kolakoski 序列以及在任意双字母字母表上的推广，更详细的综述见 $[13]$。

2 将构造方案推广至任意导向序列

2.1 导向序列的概念。

在广义 Oldenburger-Kolakoski 序列的构造方案中，当 $a = 1$ 和 $b = 2$ 时，$\mathcal{O}_{a,b}$ 的块由 a 和 b 交替组成，如 (1) 所示；当 $a = 1$ 和 $b = 3$ 时，则如 (3) 所示。换言之，若 $t_1 = a$ 且 $t_2 = b$，则 « $\mathcal{O}_{a,b} = (x_i)_{i\in\mathbb{N}^\star}$ 的第 i^{th} 个块长度为 x_i，并填入字母 $t_{i \bmod 2}$ »。

显然，这一构造方案可以推广至 $\{a, b\}$ 上的任意有限序列 $T = (t_1, t_2, \dots)$，如下所示：「$\mathcal{O}_T = (x_i)_{i\in\mathbb{N}^\star}$ 的第 i^{th} 个块长度为 x_i，并填入字母 t_i」（见程序 1）。

```
1 def O( T ) : 2 X = []3 k = 04 for x in T : 5 X += [x] # 将 'x ' 拼接到 X 的末
尾
re
```

程序 1：Python 函数：$\mathcal{O}$ 是一个作用于 $\mathbb{N}^\star$ 上的序列的函数。

我们称序列 $\mathcal{O}_T$ 由序列 T 导向，并称序列 T 是 $\mathcal{O}_T$ 的导向序列。例如，序列 $\mathcal{O}_{a,b}$ 由 $T = (ab)^\omega$ 导向，而 $\mathcal{O}$由 $(12)^\omega$导向。请注意，导向序列 $\mathcal{O}_T$ 可能不再是算子 Δ的不动点。

令我们 现在仔细看看 $\mathcal{O}_T$ 的构造如何提供了更多一些

信息，而不仅仅是序列本身。例如，设 $T=(t_n)_{n\in\mathbb{N}^\star}=21122\ldots$ 是一个取值于 $\{1,2\}$ 的序列:

步骤 1: $\mathcal{O}_{t_1}=22$ 第二个块的长度为 2：因此， 3^{rd} 和 4^{th} 字母位于同一个长度为 2 的块中。记作 $\mathcal{O}_{t_1}=22$?? 。

步骤 2: $\mathcal{O}_{t_1t_2}=22\ 11$?? ： 5^{th} 和 6^{th} 字母分别位于长度为 1 的块中。

步骤 3: $=22\ 11\ 1$?? ： 7^{th} 字母位于长度为 1 的块中。

粗略地说， t_1 给出了包含截至 4^{th} 字母的所有块的长度， t_1t_2给出了包含截至 6^{th}字母的所有块的长度，依此类推.....对于每个 $n\in\mathbb{N}^\star$ ，令 $w_n=\mathcal{O}_{t_1\ldots t_n}$，则 $\sum_{i=1}^{|w_1|}[w_1]_i=4$ 、 $\sum_{i=1}^{|w_2|}[w_2]_i=6$、 $\sum_{i=1}^{|w_3|}[w_3]_i=7$.....

2.2 导向序列集合的划分

现在我们引入导向序列的一些子集，这些子集在后续各节中至关重要。为此，我们根据序列 T 所提供的关于 $\mathcal{O}_T$中各块长度的信息，对这些序列进行分类：设 k 和 n 为两个整数，使得$1\leq k\leq n$ ，并令 $\mathcal{S}_{n,\,k}$ 为满足以下条件的序列 $T=(t_n)_{n\in\mathbb{N}}$ 的集合：读取 t_k 时，包含 $\mathcal{O}_T$ 其 n^{th} 个字母的块的长度已知，而在此之前则未知。形式化地说，若 $w_n=\mathcal{O}_{t_1\ldots t_n}$，则有

$$\mathcal{S}_{n,k}=\Big\{(t_1,\ldots,t_n)\in\{1,2\}^n:\min\{j\in[\![1;n]\!],\sum_{i=1}^{|w_j|}[w_j]_i\geq n\}=k\Big\}. \tag{4}$$

我们举一个简短的例子来说明后一个定义：设 $n=5$ ，并令 $T=(t_1,\ t_2,\ t_3,\ t_4,\ t_5)=(2,\ 1,\ 1,\ 2,\ 2)$。经过第一步并读入 t_1后，我们只知道$t_1=2$ ，仍不知道将包含 $\mathcal{O}_T$ 中 5^{th} 个字母的块的长度。另一方面，在得知 t_2 的值后，我们知道 $\mathcal{O}_T$ 的 5^{th} 个字母将写入一个大小为 1 的块中。因此， $T\in\mathcal{S}_{5,\,2}$。更一般地,

步骤 1: $\mathcal{O}_{t_1}=22$?? 和 $T\in\mathcal{S}_{3,\,1}\cap\mathcal{S}_{4,\,1}$ 因为 t_1 给出了包含 3^{rd} 的块的长度以及 $\mathcal{O}_T$的 4^{th} 。

步骤 2: $\mathcal{O}_{t_1t_2}=22\ 11$? ?; 5^{th} 和字母 6^{th} 分别位于长度为 1 的块中。因此 $T\in\mathcal{S}_{5,\,2}\cap\mathcal{S}_{6,\,2}$。

步骤 3: $\mathcal{O}_{t_1t_2t_3}=22\ 11\ 1$? ? 和 $T\in\mathcal{S}_{7,\,3}$。

步骤 4: $\mathcal{O}_{t_1t_2t_3t_4}=22\ 11\ 1\ 2$? ?? 和 $T\in\mathcal{S}_{8,\,4}\cap\mathcal{S}_{9,\,4}$。

步骤 5: $\mathcal{O}_{t_1t_2t_3t_4t_5}=22\ 11\ 1\ 2\ 2$?? ?? 和 $T\in\mathcal{S}_{10,\,5}\cap\mathcal{S}_{11,\,5}$.

集合 $\{\mathcal{S}_{n,k}:\ k\in[\![1;\ n]\!]\}$ 是 $\{1,2\}^n$ 的一个划分。事实上，字 w_n 的长度至少为 n，因为在每一步中，读入 t_i 后，我们至少写下 $\mathcal{O}_T$ 的一个字母。并且每个字母都是 1 或 2，因此 $\sum_{i=1}^{|w_n|}[w_n]_i\geq n$。

注意到 $\mathcal{S}_{n,n}=\{(1,\ldots,1)\}$。事实上，$w_{n-1}$ 的长度至少为 $n-1$，且当且仅当写出的字母全为 1 时，长度恰好为 $n-1$。此外 $\mathcal{S}_{n,n-1}=\{(1,\ldots,1,2,1),(1,\ldots,1,2,2)\}$。事实上，很容易验证 $(1,\ldots,2,1),(1,\ldots,2,2)\in\mathcal{S}_{n,n-1}$。反过来，若 $(t_1,\ldots,t_n)\in\mathcal{S}_{n,n-1}$，则存在 $i\in[\![1;\ n-1]\!]$ 使得 $t_i=2$。注意，T 中的第一个 2 写了两次。因此，若在 $[\![1;\ n-2]\!]$ 中存在这样的 i，我们就会有 $\sum_{i=1}^{|w_{n-2}|}[w_{n-2}]_i\geq n$。

2.3 将定义推广到无限指引序列

既然我们已经开始研究导向序列和被指引序列的概念，一个自然的问题便随之产生：「能否将 $\mathcal{O}_T$ 序列的定义推广到无限（可能非周期的）序列 T? 」设 $\mathcal{A}\subseteq\mathbb{N}^\star$ 为一个字母表，并设 $T=(t_n)_{n\in\mathbb{N}^\star}$ 是 $\mathcal{A}$上的一个无限序列。根据构造，对每个 $n\in\mathbb{N}^\star$，$\mathcal{O}_{t_1\ldots t_n}$ 都是$\mathcal{O}_{t_1\ldots t_{n+1}}$的前缀。因此，将 $\mathcal{O}_T$ 定义为当 n 趋于无穷时 $\mathcal{O}_{t_1\ldots t_n}$ 的极限：$\mathcal{O}_T=\lim\limits_{n\to\infty}\mathcal{O}_{t_1\ldots t_n}$。

本文研究当 $T=(t_n)_{n\in\mathbb{N}^\star}$ 是 $\mathcal{A}$上的一个无限序列时，$\mathcal{O}_T$ 中字母的密度。这些密度是否存在? 如果存在，它们的值是多少?

我们尤其关注导向序列为随机序列的情形。设$\mathbb{T}=(T_n)_{n\in\mathbb{N}^\star}$ 为一列随机变量。根据定义，由$\mathbb{T}$ 指引的序列是随机序列 $\mathbb{X}=(X_n)_{n\in\mathbb{N}^\star}$，其定义为 $\mathbb{X}=\mathcal{O}_{\mathbb{T}}$。

本文的结构安排如下。在第3节中，我们考察导向序列由取值于二字母字母表上的独立同分布（i.i.d.）随机变量构成的情形。在第4节中，我们讨论随机序列 $\mathbb{X}$ 由马尔可夫链导向的情形。

3 由独立随机变量导向的序列

本节中，$\mathbb{T}=(T_n)_{n\in\mathbb{N}^\star}$ 是一个取值于二字母字母表 $\mathcal{A}=\{1,2\}$上的独立同分布随机变量序列（简称 i.i.d.），其中 $\mathbb{P}(T_n=1)=p$ 且 $\mathbb{P}(T_n=2)=1-p$ 对每个 $n\in\mathbb{N}^\star$都成立，p 是 $]0,1[$中的一个固定参数。因此，序列 $\mathbb{T}$ 服从乘积分布 $(p\delta_1+(1-p)\delta_2)^{\otimes\mathbb{N}^\star}$。我们记作 $\mathbb{T}\sim((p\delta_1+(1-p)\delta_2)^{\otimes\mathbb{N}^\star}$。

设 $\mathbb{X}=(X_n)_{n\in\mathbb{N}^\star}$ 是由 $\mathbb{T}$ 指定的序列。序列 $\mathbb{X}$ 是一个分布先验 未知的随机序列。假设要计算某个较大的整数 n 对应的 $\mathbb{X}$ 的第 n^{th} 个字母 X_n。除非序列 $\mathbb{T}$ 以一长串 1 开头（这是一个发生概率很低的事件），否则只需读取 $\mathbb{T}$ 的前几项 T_1、……、T_k，直到确定包含 X_n 的块的长度和位置，然后以概率 p 用 1 填充该块，或以概率 $1-p$ 用 2 填充该块。这样得到的 X_n 值将具有所需的分布。由此可知，

$\lim_{n\to\infty}\mathbb{P}(X_n = 1) = p$，我们甚至可以利用这一观察更精确地计算 $\mathbb{P}(X_n = 1)$ 的值，如下述命题所示。

命题 1. 若 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes\mathbb{N}^\star}$，其中 $p \in]0,1[$，则对任意 $n \geq 2$,

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

证明. 我们将根据上一节引入的 $\{1,2\}^n$ 的划分 $\{\mathcal{S}_{n,k} : k \in [\![1;n]\!]\}$，对事件 $\{X_n = 1\}$ 进行分解，并使用观察结果 (??) 和 (??)。以下两个特殊情形显而易见:

$$\mathbb{P}(X_n = 1 \,|\, (T_1,...,T_n) \in \mathcal{S}_{n,n}) = p, \quad \text{and} \quad \mathbb{P}(X_n = 1 \,|\, (T_1,...,T_n) \in \mathcal{S}_{n,n-1}) = 0.$$

现在考虑某个 $(T_1,...,T_n) \in \mathcal{S}_{n,k}$，其中 $k < n-1$。令 k' 为满足 $|w_{k'-1}| < n$ 和 $|w_{k'}| \geq n$ 的唯一整数。具体而言，字母 X_n 是在第 k'^{th} 步写下的，且它等于 $T_{k'}$。根据 $\mathcal{S}_{n,k}$ 的定义，我们有 $k' \geq k$。我们来证明$k' \neq k$。我们通过反证法，假设 $k' = k$。现在来看第 $k-1^{\text{th}}$ 步结束时我们已知的内容。

- 根据 k' 的定义， X_n 将写入 $w_{k'-}$ 右侧的下一个块中 $_{1.}$
- 根据 k 的定义，我们还不知道这个块的长度。

因此，我们不知道 $w_{k'-1}$ 右侧任何（空）块的长度。这直接意味着 $w_{k'-1}$ 由 **1** 组成（因此， $w_{k'-1} = 1^{k'-1}$）。由此可得$n \leq |w_{k'}| \leq |w_{k'-1}| + 2 = k' + 1$。由于 $k' = k < n-1$, 我们得到一个矛盾，这意味着 $k < k'$。

$(T_1, ..., T_n)$ 属于 $\mathcal{S}_{n,\,k}$ 这一事实仅取决于序列的开头部分 $(T_1, ..., T_k)$。回想一下，$\mathbb{T}$ 具有乘积分布 $(p\delta_1 + (1-p)\delta_2)^{\otimes\mathbb{N}^\star}$. 由于 $k' > k$，且 $X_n = T_{k'}$，我们推得

$$\mathbb{P}(X_n = 1 \,|\, (T_1,...,T_n) \in \mathcal{S}_{n,k}) = p.$$

最后，我们使用全概率公式:

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

□

作为推论，我们得到 $\mathbb{X}$ 中 **1** 的比例的如下收敛结果。

推论 1。若 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$ 满足 $p \in]0,\ 1[$, 则

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

证明。由于 $\mathbb{X}$ 的取值属于 $\{1,2\}$，我们有 $|X_1 \dots X_n|_1 = 2n - (X_1 + \dots + X_n)$。由此可得

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

根据命题 1，我们有 $\lim_{n\to\infty} \mathbb{P}(X_n = 1) = p$ 且 $\lim_{n\to\infty} \mathbb{P}(X_n = 2) = 1 - p$。由此可推得 $\lim_{n\to\infty} \mathbb{E}(X_n) = 2 - p$。根据 Cesàro 引理，我们得到：

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

□

每次我们用 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$进行模拟时，$\mathbb{X}$ 中 1 的频率似乎都收敛到 p。因此，我们预计序列 $|X_1 \dots X_n|_1/n$ 几乎处处收敛到 p，而不仅仅是在期望意义下收敛。由于随机变量 $(X_n)_{n\in\mathbb{N}^\star}$ 彼此相关，我们无法直接应用强大数定律（SLLN）来证明 $(X_1 + \dots + X_n)/n$几乎处处收敛。不过，由于这些相关性足够弱，我们可以应用以下更强版本的强大数定律。

定理 1 (Lyons [11])。设$(Y_n)_{n\in\mathbb{N}^\star}$ 是一个实值随机变量序列，使得对于所有 $n \in \mathbb{N}^\star$、$|Y_n| \le 1$ 以及

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

那么$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n Y_k = 0$ 几乎处处成立。

为了应用定理 1，我们先证明下面的引理。

引理 1. 若 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$，其中 $p \in]0、\ 1[$，则对任意 $m \ge 1$ 和任意 $n \ge m+2$,

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

证明。由于 $\mathcal{S}_{n,\ n} \cap (X_m = 2) = \emptyset$ 且 $\mathcal{S}_{n,\ n-1} \cap (X_m = 2) = \emptyset$，我们有

$$\mathbb{P}(X_m = 2 \cap X_n = 1) = \sum_{k=1}^{n-2} \mathbb{P}(X_m = 2 \cap X_n = 1 \cap (T_1, \dots, T_n) \in \mathcal{S}_{n,k}).$$

由此可得

$$\mathbb{P}(X_m = 2 \cap X_n = 1) = \sum_{k=1}^{n-2} \mathbb{P}((T_1, ..., T_n) \in \mathcal{S}_{n,k}) \times \mathbb{P}(X_n = 1 \mid (T_1, ..., T_n) \in \mathcal{S}_{n,k}) \times \mathbb{P}(X_m = 2 \mid X_n = 1 \cap (T_1, ..., T_n) \in \mathcal{S}_{n,k})$$

首先，注意到对于 $k \leq n-2$, $\mathbb{P}(X_n = 1 \mid (T_1, \ldots\ldots, T_n) \in \mathcal{S}_{n,\,k}) = p$。现在我们来证明

$$\mathbb{P}(X_m = 2 \mid X_n = 1 \cap (T_1, ..., T_n) \in \mathcal{S}_{n,k}) = \mathbb{P}(X_m = 2 \mid (T_1, ..., T_n) \in \mathcal{S}_{n,k}).$$

这等价于证明：当 $X_m = 2$ 和 $(T_1, \ldots\ldots, T_n) \in \mathcal{S}_{n,\,k}$ 不相容时，

$$\mathbb{P}(X_n = 1 \mid X_m = 2 \cap (T_1, ..., T_n) \in \mathcal{S}_{n,k}) = \mathbb{P}(X_n = 1 \mid (T_1, ..., T_n) \in \mathcal{S}_{n,k}).$$

令 i 为满足字母 X_m 由 T_i给出的整数。我们可以将事件 $(T_1, \ldots\ldots, T_n) \in \mathcal{S}_{n,\,k}$ 分解为以下两种情况。

1. 若 $i > k$，那么读完 $(T_1, \ldots\ldots, T_k)$ 后，我们知道包含X_m 和 X_n 的块的大小，但不知道它们的内容：

$$\underbrace{\times\times\times\times\times\times\times\times}_{w_k}\,\underline{??}\,\underline{?}\,\underbrace{2}_{X_m}\,\underline{??}\,\underline{??}\,\underbrace{?}_{X_n}.$$

在这种情况下，给出 X_m 和 X_n 取值的变量相互独立，因此，关于 $X_m = 2$ 的额外信息不会影响 $X_n = 1$发生的概率。

2. 若 $i \leq k$，那么读完 $(T_1, \ldots\ldots, T_k)$ 后，我们已经知道 $X_m = 2$是否成立，但无法得知包含 X_n的块的内容；该内容是独立抽取的：

$$\underbrace{\times\times\times\times\times\times 2\times}_{w_k \text{ contains } X_m}\,\underline{??}\,\underline{?}\,\underline{?}\,\underline{??}\,\underline{??}\,\underbrace{?}_{X_n}.$$

在所有情况下，我们都有

$$\mathbb{P}(X_n = 1 \mid X_m = 2 \cap (T_1, ..., T_n) \in \mathcal{S}_{n,k}) = \mathbb{P}(X_n = 1 \mid (T_1, ..., T_n) \in \mathcal{S}_{n,k}) = p.$$

由此可得

$$\begin{aligned}\mathbb{P}(X_m = 2 \cap X_n = 1) &= \sum_{k=1}^{n-2} \mathbb{P}(T_1, ..., T_n) \in \mathcal{S}_{n,k}) \times p \times \mathbb{P}(X_m = 2 \mid (T_1, ..., T_n) \in \mathcal{S}_{n,k}) \\ &= p \times \mathbb{P}(X_m = 2).\end{aligned}$$

□

现在我们可以陈述以下定理。

定理 2. 若 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_2)^{\otimes \mathbb{N}^\star}$ 且 $p \in]0,1[$，则

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

证明. 为了应用定理 1，我们需要对随机变量 $(X_n)_{n\in\mathbb{N}^\star}$ 进行中心化。因此，对于 $n \in \mathbb{N}^\star$，我们引入随机变量 $\tilde{X}_n = X_n - (2-p)$，以满足 $|\tilde{X}_n| \leq 1$ 且 $\lim_{n\to\infty} \mathbb{E}(\tilde{X}_n) = 0$。现在，我们利用引理 1 计算 $\mathbb{E}(\tilde{X}_m \tilde{X}_n)$，其中 $n \geq m+2$。我们有

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

汇总这些值并利用命题 1，我们得到

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

定义函数 $\Phi : \mathbb{N}^\star \to \mathbb{R}$，使得 $\Phi(0) = \Phi(1) = 1$，且对所有 $k \geq 2$，$\Phi(k) = 0$。于是，对所有 $m,\ n \in \mathbb{N}^\star$，均有 $\mathbb{E}(X_m X_n) \leq \Phi(|n-m|)$，并且 Φ 显然满足 $\Phi \geq 0$ 和 $\sum_{n\geq 1} \Phi(n)\, n < \infty$。由定理 1，我们推得

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

因此，$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = 2 - p$ 几乎必然，并且

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

□

最后，关于导向序列服从乘积分布的情形，值得一提的是，前面的结果可以推广到其他字母表。特别是，命题 1 可推广如下。

命题 2。 设 $a, b \in \mathbb{N}^\star$ 满足 $1 < a < b$，并设 $p \in]0,1[$。

1. 若 $\mathcal{A} = \{1, a\}$ 且 $\mathbb{T} \sim (p\delta_1 + (1-p)\delta_a)^{\otimes \mathbb{N}^\star}$，则

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

2. 若 $\mathcal{A}=\{a,b\}$ 且 $\mathbb{T}\sim(p\delta_a+(1-p)\delta_b)^{\otimes\mathbb{N}^\star}$，则

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

证明。 1. 我们使用与命题 1 的证明中相同的划分，但现在区分集合 $\mathcal{S}_{n,k}$ for $n-a+1\leq k\leq n$。我们有 $\mathcal{S}_{n,n}=\{(t_1,\dots,t_n)\in\{1,a\}^n:(t_1,\dots,t_{n-1})=(1,\dots,1)\}$，并且对于 $n-a+1\leq k\leq n-1$，

$$\mathcal{S}_{n,k}=\{(t_1,\dots,t_n)\in\{1,a\}^n:(t_1,\dots,t_k)=(1,\dots,1,a)\}.$$

如果 $n-a+1\leq k\leq n-1$，那么出于与之前相同的理由，$\mathbb{P}(X_n=1\,|\,(T_1,\dots,T_n)\in\mathcal{S}_{n,k})=0$。在所有其他情况下，$\mathbb{P}(X_n=1\,|\,(T_1,\dots,T_n)\in\mathcal{S}_{n,k})=p$。因此，

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

2. 由于 $a>1$，对于所有 $k\in\mathbb{N}^\star$，我们都有 $\sum_{j=1}^{|w_k|}[w_k]_j-|w_k|\geq(a-1)|w_k|>0$。这意味着在 k^{th} 步之后，我们总能知道至少一个空块的长度。因此，除非 $n\leq b$（此时 n^{th} 字母可能在第一步中写出），否则我们确定能在填充包含 n^{th} 字母的块之前，知道该块的长度。由于 T_j 相互独立，我们推得 $\mathbb{P}(X_n=a)=p$。

□

4 由马尔可夫链导向的序列

为了更接近确定性情形（即 1 后面总是跟着 2，反之亦然），我们现在关注由马尔可夫链给出的导向序列。

本节中，我们假设导向序列 $\mathbb{T}=(T_n)_{n\in\mathbb{N}^\star}$ 是字母表 $\{1,2\}$ 上的马尔可夫链，其初始值为 $T_1=1$，且从 1 到 2（以及从 2 到 1）的转移概率为 $p\in]0,\ 1[$。p 的值越大，越倾向于在 1 和 2 之间交替。原始的 Oldenburger-Kolakoski 序列情形可以看作是一个马尔可夫链的“极限”情形，其从 1 到 2（以及从 2 到 1）的转移概率等于 1。

定理3。 设 $p\in]0、\ 1[$，且 $\mathbb{T}$ 是字母表 $\{1,2\}$ 上的马尔可夫链，其初始值为 $T_1=1$，从 *1* 到 *2*（以及从 *2* 到 *1*）的转移概率为 $p\in]0,\ 1[$。则

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

证明。 首先注意，对于所有整数 $s>r\geq1$，$\mathbb{P}(T_s=1\,|\,T_r=1)=\frac{1}{2}(1+(1-2p)^{s-r})$ 和 $\mathbb{P}(T_s=1\,|\,T_r=2)=\frac{1}{2}(1-(1-2p)^{s-r})$。

设 $\ell\in\mathbb{N}^\star$ 且令 $n\geq2\ell$。考虑整数 $k\in\mathbb{N}^\star$，使得 $(T_1、\dots\dots、T_n)\in\mathcal{S}_{n,\,k}$。如果在 $T_1、\dots\dots、T_{\lfloor n/2\rfloor}$ 中至少出现 2ℓ 次 2，那么在 $(T_1、\dots\dots、T_k)$ 中至少出现 2ℓ 次 2，因此在 w_k 中至少出现 2ℓ 次 2。这意味着

$n - |w_k| \geq 2\ell$，这意味着，当写出 w_k 时，我们知道位置 $|w_k|$ 和位置 n 之间至少有 ℓ 个空块的长度。与命题 1 的证明类似，考虑整数 k'，使得 X_n 由 $T_{k'}$ 给出。我们还引入 $D = k - k'$，如下图所示（根据定义， $X_{|w_k|} = T_k$ 和 $X_n = T_{k'}$）：

$$\underbrace{\times\times\times\times\times\times\times}_{}\ \underbrace{\times}_{X_{|w_k|}}\ \underbrace{??\ ?\ ??\ ??}_{D-1 \text{ blocks}}\ \underbrace{?}_{X_n}.$$

根据上述观察，我们有

$$\mathbb{P}(D < \ell) \leq \mathbb{P}(\{ \text{ less that } 2\ell \text{ occurences of 2 among } T_1, \ldots, T_{\lfloor n/2 \rfloor} \}),$$

并且，右侧的概率在 n 趋于 ∞时趋于 0。此外，

$$\begin{aligned}
\mathbb{P}(X_n = 1 \mid D \geq \ell) &= \mathbb{P}(T_{k'} = 1 \mid D \geq \ell) \\
&= \mathbb{P}(T_{k'} = 1 \mid D \geq \ell \cap T_k = 1) \times \mathbb{P}(T_k = 1 \mid D \geq \ell) \\
&\quad + \mathbb{P}(T_{k'} = 1 \mid D \geq \ell \cap T_k = 2) \times \mathbb{P}(T_k = 2 \mid D \geq \ell) \\
&\in \left[\frac{1}{2}(1 - |1-2p|^\ell);\ \frac{1}{2}(1 + |1-2p|^\ell) \right]
\end{aligned}$$

这得益于证明开头的评述。由此可得

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

然后令 ℓ 趋于无穷大，得到 $\lim_{n\to\infty} \mathbb{P}(X_n = 1) = \frac{1}{2}$。 - □

由定理3可直接得到以下结果。

推论2. 设 $p \in]0,1[$，且 $\mathbb{T}$ 是取值于字母表 $\{1,2\}$ 上的马尔可夫链，初始值为 $T_1 = 1$，从 *1* 转移到 *2*（以及从 *2* 转移到 *1*）的转移概率为 $p \in]0,1[$。则

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

我们猜想，收敛也几乎处处成立，但迄今我们仍无法证明这一点，因为在马尔可夫情形下，相关性的计算要复杂得多。

猜想1。 设 $p \in]0、1[$，并设 $\mathbb{T}$ 是一个取值于字母表 $\{1,2\}$ 的马尔可夫链，其初始值为 $T_1 = 1$，从 *1* 到 *2*（以及从 *2* 到 *1*）的转移概率为$p \in]0,1[$。那么

$$\lim_{n\to\infty} \frac{|X_1 \ldots X_n|_1}{n} = \frac{1}{2} \ \textit{almost surely.}$$

注意，定理3和推论2很容易推广到其他字母表。特别是，在字母表 $\{1,3\}$ 上也能得到相同的结果：如果 $\mathbb{T}$ 是转移概率为 $0 < p < 1$ 的马尔可夫链，即从 1 到 3（以及从 3 到 1）的转移概率均为 $0 < p < 1$，那么 1 的平均密度为 $1/2$。

定理 4。 设 $a \geq 2$ 为整数，设 $p \in]0、1[$，并设 $\mathbb{T}$ 是字母表 $\{1, a\}$ 上的马尔可夫链，其初始值为 $T_1 = 1$，且从 1 到 a （以及从 a 到 1）的转移概率为 $p \in]0、1[$。那么

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

定理 4 的陈述有些令人惊讶，也出乎意料，因为我们知道，序列 $\mathcal{O}_{1,\ 3}$ 和 $\mathcal{O}_{3,\ 1}$ 中字母 1 和 3 的密度 d_1 和 d_3 分别是$d_1 \approx 0$，40 和 $d_3 \approx 0$，60。 [1]。我们将在第 6 节的讨论中再谈这一点。

5 密度不守恒

在前几节中，我们研究了导向序列为随机序列的不同情形。在我们考察的所有情形中（独立同分布随机变量序列、马尔可夫序列），所得导向序列中字母的密度几乎处处与导向序列中的密度相同。

模拟还表明，对于任意（无限）周期序列 T，导向序列 $\mathcal{O}_T$ 中 1 的密度是良定义的，并且等于 T中 1 的密度，见图 1。

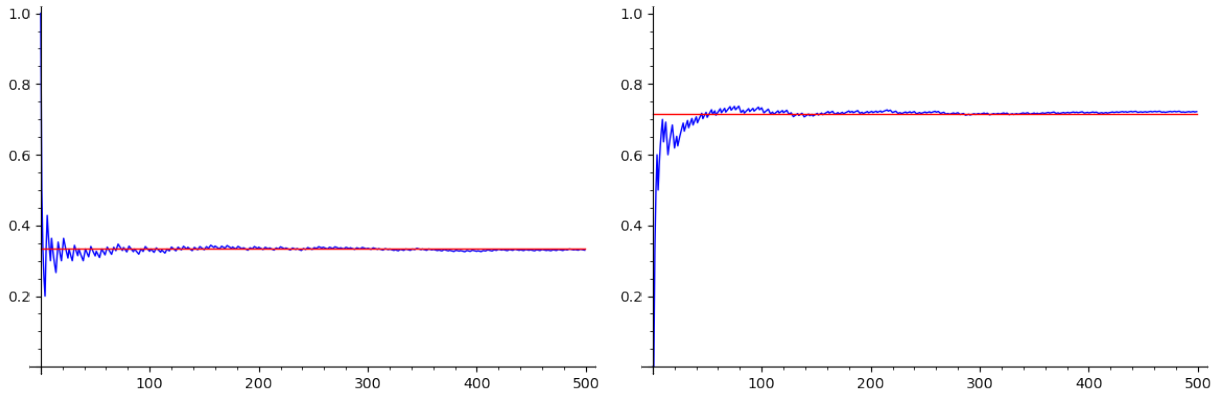


图 1：对于 T =(122)$^\omega$（左）和 T =(2112111)$^\omega$ （右），$\mathcal{O}_T$ 越来越长的前缀中 1 的密度变化情况。密度似乎分别收敛于 $1/3$ 和 $5/7$。

在图 1 中，我们只展示了 $\mathcal{O}_T$ 较短前缀的数据，以便图表仍然便于使用，尤其是能够区分序列 $\mathcal{O}_T$ 最初几项中的密度。不过，我们还对大量周期序列 T 进行了进一步实验，结果似乎印证了我们的初步印象，即如果序列 T 是周期的，那么 $\mathcal{O}_T$ 中的密度将与 T 中的相同。这使我们提出以下猜想，它推广了 Keane 猜想。

猜想 2. 对于字母表 $\{1,2\}$ 上的任意周期序列 T，导向序列 $\mathcal{O}_T$ 中 1 的密度是良定义的，并且等于 T 中 1 的密度。

于是，一个自然的问题随之产生：是否存在一个导向序列 T over $\{1,2\}$，使得 T 中 1 的密度在 $\mathcal{O}_T$ 中不守恒？显然，鉴于猜想 2，我们不指望在周期序列中找到这样的导向序列候选对象。

不过，得益于从左到右读取 $\mathcal{O}_T$ 可以提供更靠右块的大小这一事实（见第 2 节），我们对这个问题给出了部分肯定的回答。在逐步构造序列 T 和 $\mathcal{O}_T$ 的过程中，了解 $\mathcal{O}_T$ 中尚未填充的块的长度，可以让我们完全任意地选择用哪个字母填充它们，从而有机会迫使序列 $\mathcal{O}_T$ 中的 1 相对多于序列 T。

我们对 T 和 $\mathcal{O}_T$ 的同步构造方案的主要思路可概括如下：先将 T_1 初始化为 2，然后 $\mathcal{O}_T = 22??$；此后，从左到右读取 $\mathcal{O}_T$，将其大小为 2 的块填入 1，并交替将其大小为 1 的块填入 1 和 2。因此，同时构造 T 和 $\mathcal{O}_T$ 的最初几步如下（采用第 2 节的记号）：

步骤 1： 我们设定 $\mathrm{T}^{(1)} = (2)$，然后 $\mathcal{O}^{(1)} = \underline{22}\,\underline{??}$

步骤 2： (a) $\mathcal{O}^{(1)}$ 中大小为 2 的空块必须填入 1：$\mathcal{O}^{(2)} = 22\ 11\ ??$ (b) 我们设定 $\underline{\mathrm{T}}\,\underline{\ }\,\underline{\ }$ $^{(2)} = (2,1)$

步骤 3： (a) 我们用 1 填入 $\mathcal{O}_T^{(2)}$ 中下一个大小为 1 的块：$\mathcal{O}^{(3)} = \underline{22}\,\underline{11}\,\underline{1}\,\underline{?}\,\underline{?}\,\underline{\ }$
(b) 因此 $\mathrm{T}^{(3)} = (2,1,1)$

步骤 4： (a) 我们用 2 填入 $\mathcal{O}_T^{(3)}$ 中下一个大小为 1 的块：$\mathcal{O}^{(4)} = \underline{22}\,\underline{11}\,\underline{1}\,\underline{2}\,\underline{?}\,\underline{??}$
(b) 因此 $\mathrm{T}^{(4)} = (2,1,1,2)$

步骤 5： 等等.....

对于每个 $n \in \mathbb{N}^\star$，都有 $\mathcal{O}^{(n)} = \mathcal{O}_{\mathrm{T}^{(n)}}$。此外，$\mathrm{T}^{(n)}$ 是 $\mathrm{T}^{(n+1)}$ 的前缀，而 $\mathcal{O}^{(n)}$ 是 $\mathcal{O}^{(n+1)}$ 的前缀，因此我们设 $T = \lim_{n\to\infty} \mathrm{T}^{(n)}$。由此可得 $\mathcal{O}_T = \lim_{n\to\infty} \mathcal{O}^{(n)}$。

对于所有 $i \in \mathbb{N}$，我们用 t_i，$x_i \in \{1,2\}$ 表示 $T = (t_i)_{i\in\mathbb{N}}$ 和 $\mathcal{O}_T = (x_i)_{i\in\mathbb{N}}$，那么：

1. $\mathrm{T}^{(n)} = (t_1、\ .....、\ t_n)$ 以及 $|\mathrm{T}^{(n)}| = n$。
2. $\mathcal{O}^{(n)} = (t_i^{x_i})_{i\in[\![1,n]\!]}$ 和 $\mathcal{O}_T = (t_i^{x_i})_{i\in\mathbb{N}^\star}$。
3. $|\mathrm{T}^{(n)}|_1 = |x_1 \dots x_n|_2 - 1 + \frac{1}{2}|x_1 \dots x_n|_1 + C_n$，其中 $C_n \in \{0,1\}$：事实上，$\mathrm{T}^{(n)}$ 中 1 的个数等于 $\mathcal{O}^{(n)}$ 中大小为 2 的块的个数（不包括 $\mathcal{O}^{(n)}$ 的第一个块，因为 $\mathcal{O}^{(1)}$的初始化所致）与 $\mathcal{O}^{(n)}$中大小为 1 的块的个数的一半之和。根据构造，$\mathcal{O}^{(n)}$ 中大小为 1 的块（分别为大小为 2 的块）的个数等于 $x_1 \dots x_n$中 1（分别为 2）的个数。常数 C_n考虑了 $x_n = 1$ 是 $\mathcal{O}^{(n)}$中大小为 2 的块的首字母的情形。

程序 2 提供了一个用于构造 $\mathcal{O}_T$ 和 T的 Python 函数。

```
1 定义 序列（n）： 2 T = [2] 3 O_T = [2, 2] 4 d = 1 # 要写入下一个大小为 1 的块中的数字 5 对于 i 在
范围（1，n）内： 6 如果 O_T[i] == 2： 7 T += [1] 8 O_T += [1] * 2 9 否则： 10 T += [d] 11 O_
T
```

程序 2：用于同时构造 T 和 $\mathcal{O}_T$的 Python 函数。

定理5。设 $T = \text{极限}_{n\to\infty}\mathrm{T}^{(n)}$。以下性质成立：

1. 如果 中 ’ 的密度 d_1^T 1 T $d_1^T = \frac{1+\sqrt{17}}{8} = 0.640\ldots$ $d_1^{\mathcal{O}} = \frac{7-\sqrt{17}}{4} =$ 存在，那么，$0.719\ldots$ 因此 $d_1^T \neq d_1^{\mathcal{O}}$。

2. 如果 T 中 1 的密度 d_1^T 存在，那么序列 T 和 $\mathcal{O}_T$ 不是周期的。

证明。 1. 对于每个 $n \in \mathbb{N}^\star$，我们有：

$$\Big|\, |\mathcal{O}^{(n)}|_1 - |\mathrm{T}^{(n)}|_2 - 2(|\mathrm{T}^{(n)}|_1 - |\mathrm{T}^{(n)}|_2) \,\Big| \leq 1$$

事实上，精确到一个单位，$\mathrm{T}^{(n)}$ 的每个数字 2 都会在 $\mathcal{O}^{(n)}$ 中产生一个 2；同样数量的 $|\mathrm{T}^{(n)}|_2$ 个 1 会产生一个 1，而其余的（即 $|\mathrm{T}^{(n)}|_1 - |\mathrm{T}^{(n)}|_2$）则会在 $\mathcal{O}^{(n)}$ 中产生两个 1。此外，$\mathrm{T}^{(n)}$ 中第一个 2 是唯一一个在 $\mathcal{O}^{(n)}$ 中被写两次的，因此我们总是恰好有 $|\mathcal{O}^{(n)}|_2 = |\mathrm{T}^{(n)}|_2 + 1$。于是，

$$\begin{aligned}\frac{|\mathcal{O}^{(n)}|_1}{|\mathcal{O}^{(n)}|} &= \frac{|\mathcal{O}^{(n)}|_1}{|\mathcal{O}^{(n)}|_1 + |\mathcal{O}^{(n)}|_2} = \frac{2|T^{(n)}|_1 - |T^{(n)}|_2 + o(n)}{-1 + |T^{(n)}|_2 + 2(|T^{(n)}|_1 - |T^{(n)}|_2) + |T^{(n)}|_2 + 1 + o(n))} \\ &= \frac{3|T^{(n)}|_1 - |T^{(n)}| + o(n)}{2|T^{(n)}|_1 + o(n)} \xrightarrow[n\to\infty]{} \frac{3d_1^T - 1}{2d_1^T}\end{aligned}$$

由此可得

$$d_1^{\mathcal{O}} = \frac{3d_1^T - 1}{2d_1^T} \tag{5}$$

且 $\mathcal{O}_T$ 中 1 的密度（分别地，2 的密度）存在。

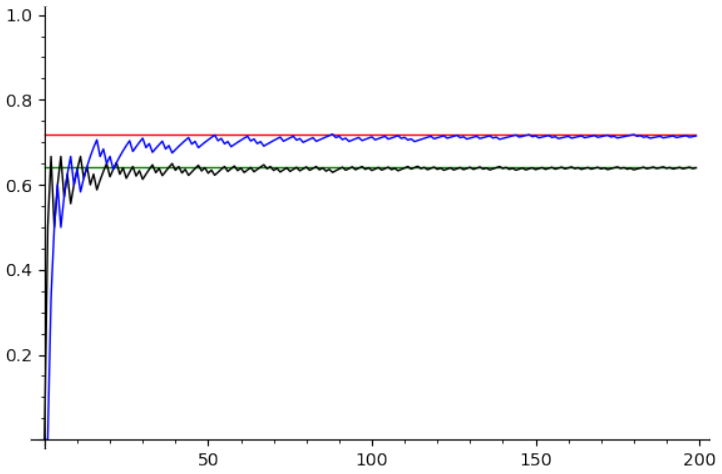


图 2： $\mathcal{O}_T$ （蓝色）和 T （黑色）中 1 的密度变化，这两个序列由程序 2 定义。

我们在上文注意到，对于每个 $n \in \mathbb{N}^\star$， $|\mathbf{T}^{(n)}|_1 = |x_1 \dots x_n|_2 - 1 + \frac{1}{2}|x_1 \dots x_n|_1 + C_n$，其中 $C_n \in \{0, 1\}$。此外，若 d_1^T 存在，那么 d_1^O 和 d_2^O 也存在；并且令 n 趋于无穷大，可得：

$$d_1^T = d_2^O + \frac{1}{2} d_1^O \tag{6}$$

结合方程 (5) 和 (6)，我们推导出 $d_1^T = \frac{1+\sqrt{17}}{8}$ 和 $d_1^O = \frac{7-\sqrt{17}}{4}$。

2. 如果序列 T 和 $\mathcal{O}_T$ 是周期的，那么它们的 1 和 2 的密度就会是有理数，但事实并非如此。

□

模拟结果表明，这些密度确实正在趋近于这些值，见图 2。

6 结论与讨论

在字母表 $\{1, a\}$ 上，取 $a \in \{2, 3\}$，我们证明，对于几乎所有由无限序列 $\mathbb{T} = (T_n)_{n \in \mathbb{N}^\star}$ 所指示的序列（其中随机变量独立同分布，且具有 $\mathbb{P}(T_n = 1) = p \in]0,\ 1[$ 和 $\mathbb{P}(T_n = a) = 1 - p$），1 的密度等于 p。我们还证明，由转移概率为 $p \in]0,\ 1[$ （从 1 转移到 a，以及从 a 转移到 1）的马尔可夫链所指示的所有序列中，1 的平均密度等于 $1/2$。

Keane 猜想 [9] 指出，这一结果可以推广到确定性情形，即当 $p = 1$ 时，在字母表 $\{1, 2\}$ 上成立。另一方面，在字母表 $\{1, 3\}$

这一结果无法推广到确定性情形，因为 $\mathcal{O}_{1,\ 3}$ 中 1 的密度接近 0.3972。 [1]。

当 $\mathbb{T}$ 是马尔可夫链时，其转移概率 p 越接近 1，序列 $\mathcal{O}_T$ 就越可能与$\mathcal{O}_{1,\ 3}$共享较长的前缀。因此，转移概率 p 越接近 1，序列 $\mathcal{O}_T$ 中 1 的密度在较长前缀上就越接近序列$\mathcal{O}_{1,\ 3}$ 中的 1 的密度。然而，计算机实验表明，当 $\mathbb{T}$ 中 1 和 3 交替排列首次出现扰动时，随着前缀变长， $\mathcal{O}_{\mathbb{T}}$ 前缀中 1 的密度最终趋近于 0.5。图3展示了 $p = 0.99$ 的示例。

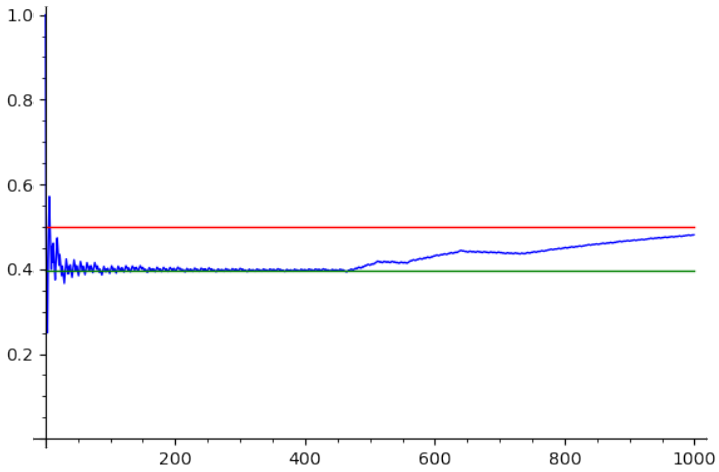


图3：字母表为 $\{1, 3\}$ 、参数为 $p = 0.99$ 的马尔可夫指导序列中 1 的频率变化：其频率起初接近 $\mathcal{O}_{1,\ 3}$，随后逐渐远离该值并趋近于 $1/2$。

这意味着，似乎很难在字母表$\{1, 2\}$上的马尔可夫情形中令 p 趋近于 1，从而推导出有关原始 Oldenburger-Kolakoski 序列$\mathcal{O}_{1,\ 2}$ 的信息。

最后，研究由随机序列驱动、定义在含有 2 个以上字母的字母表上的序列，或由其他分布构造的随机序列，似乎也很有趣。

参考文献

[1] Michael Baake 和 Bernd Sing。Kolakoski-(3, 1) 是一个（变形的）模型集。《加拿大数学通报》，47(2):168–190，2004。

[2] Valérie Berthé、Srecko Brlek 和 Philippe Choquette。任意字母表上的光滑词。理论计算机科学，341(1-3):293–310，2005年。

[3] Valérie Berthé、Srecko Brlek 和 Philippe Choquette。任意字母表上的光滑词。理论计算机科学，341(1-3):293–310，2005年。[4] Srecko Brlek、Serge Dulucq、A. Ladouceur 和 Laurent Vuillon。光滑无限词的组合性质。理论计算机科学，352(1-3):306–317，2006年。[5] Srecko Brlek、Damien Jamet 和 Geneviève Paquin。具有相同奇偶性的双字母表上的光滑词。理论计算机科学，393(1-3):166–181，2008年。[6] Arturo Carpi。Kolakovski序列中的重复。欧洲理论计算机科学协会公报，50:194–197，1993年。[7] Vašek Chvátal。关于Kolakoski序列的注记。技术报告，DIMACS技术报告93-84，1993年12月。[8] F. M. Dekking。论自生成序列的结构。波尔多数论研讨会，第1–6页，1980年。[9] Michael S. Keane。遍历理论与有限型子移位。载于《遍历理论、符号动力系统与双曲空间》。在“《双曲几何与遍历理论》”研讨会上所作讲座，该研讨会于*1989*年*4*月*17*日至*28*日在意大利的里雅斯特国际理论物理中心举行，第35–70页。牛津等：牛津大学出版社，1991年。[10] William Kolakoski。自生成游程，问题5304。《美国数学月刊》，73(6):681–682，1966年。[11] Russell Lyons。弱相关随机变量的强大数定律。《密歇根数学杂志》，35(3):353 – 359，1988 年。[12] Rufus Oldenburger。符号动力系统中的指数轨迹。《美国数学会汇刊》，46(3):453–466，1939年。[13] Bernd Sing。更多Kolakoski序列。《整数》，11B:A14，2011年。