# A Recurrence for the Run Numbers of Kolakoski-$(p, 2p)$

Jon Maiga
Sequence Machine

**Abstract**

The classical Kolakoski sequence self-encodes runs of 1s and 2s. It is unknown whether limiting symbol densities exist. We give a short recurrence that yields the run number at every position and extend it to all Kolakoski-$(p, 2p)$ sequences. This viewpoint ties the elusive odd-$p$ cases (including $p = 1$) to the even-$p$ variants, where symbol densities are known.

# 1 Introduction

# 2 Introduction

The Kolakoski sequence [1, 2], named after the mathematician William Kolakoski, is an infinite sequence that self-encodes its run lengths. It was first discussed by Oldenburger in 1939 [3] before being popularized by Kolakoski. A *run* is a maximal streak of equal terms. The On-Line Encyclopedia of Integer Sequences [4] defines the Kolakoski sequence, which we denote by $k(n)$, as follows: "... $k(n)$ is the length of $n$-th run ...".

$$k(n) = (1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, \dots) = \text{A000002}$$

**Example 1.** The first six runs in the Kolakoski sequence are (1), (2,2), (1,1), (2), (1), and (2,2). According to the definition of the sequence, $k(3) = 2$ indicates that the third run has a length of two: (1, 1). Similarly, for $k(5) = 1$, the fifth run consists of a single element: (1).

There are many interesting questions regarding the properties of the Kolakoski sequence. For instance, it remains unclear whether the sequence contains asymptotically equal numbers of 1s and 2s, as conjectured by Keane [11]. Empirical evidence, for example Nilsson [7], supports this possibility. Previous studies, including [5, 6], have found recurrence relations for $k(n)$ and its companion sequences. In this study, we introduce a recurrence for its run number. Then we generalize it and show how it connects to all Kolakoski-$(p, 2p)$ sequences.

## 2.1 Constructing the Kolakoski sequence

To construct the Kolakoski sequence, we use what we call the *driving term*: at each step, the driving term $k(n)$ specifies the length of the next run to append. We begin with the initial terms $(1, 2, 2)$ and set $n = 3$ (marking the driving term with *), appending $k(n)$ copies of the current symbol (highlighted in bold):

1. At $k(3) = 2$, the third run has length two: append two 1s. (1,2,*2,**1,1**)
2. At $k(4) = 1$, the fourth run has length one: append one 2. (1,2,2,*1,1,**2**)
3. At $k(5) = 1$, the fifth run has length one: append one 1. (1,2,2,1,*1,2,**1**)
4. At $k(6) = 2$, the sixth run has length two: append two 2s. (1,2,2,1,1,*2,1,**2,2**)
5. ...

Note that in each step above, we append a complete run. Consequently, for every $n$ (the index to the driving term in $k$), the appended symbol alternates by run. This construction algorithm leads us to the equivalent mapping that Culik II and Lepistö found [8]:

$$\begin{cases} 1 \mapsto 1,\ 2 \mapsto 11, & \text{if } n \text{ is odd;} \\ 1 \mapsto 2,\ 2 \mapsto 22, & \text{if } n \text{ is even.} \end{cases} \tag{1}$$

# 3 The run number in the Kolakoski sequence

We complement the Kolakoski sequence by creating a sequence that enumerates the runs $r(n)$. That is, every time terms in $k$ change symbol, we increment the run number. The construction of $r$ is analogous to $k$, but instead of appending the symbols, we append the run number.

$$k(n) = (1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, \dots) = \text{A000002},$$
$$r(n) = (1, 2, 2, 3, 3, 4, 5, 6, 6, 7, 8, 8, 9, 9, \dots) = \text{A156253}.$$

The OEIS lists the sequence $r(n)$ as A156253. Because we alternate between odd and even numbers for every appended run in both $k$ and $r$, we can recover $k$ from $r$:

$$k(n) = \begin{cases} 1, & \text{if } r(n) \text{ is odd;} \\ 2, & \text{if } r(n) \text{ is even.} \end{cases} \tag{2}$$

This relation is also noted by Benoit Cloitre in A000002.

## 3.1 A recursive formula for the Kolakoski run numbers

We now express a recursive formula for $r(n)$. Recall the mapping algorithm (1) for $k(n)$, when we encounter a '1', we append exactly one new symbol, whereas each '2' appends exactly two identical symbols. Hence each '2' pauses the growth of $r(n)$ for one step and therefore we can express the run number as the maximum number of opportunities to start a run minus the number of pauses. We now formalize the recurrence.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k(n)$ (driving terms) | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 |
| $r(n)$ (run numbers) | 1 | 2 | $\underline{2}$ | 3 | $\underline{3}$ | 4 | 5 | 6 | $\underline{6}$ | 7 | 8 | $\underline{8}$ | 9 | $\underline{9}$ | 10 | 11 |

Table 1: Row 2 lists the Kolakoski sequence $k(n)$. Underlined values in row 3 are "extra" symbols: positions where $k(r(n)) = 2$ have produced a duplicate symbol, so the run number $r(n)$ does not increase.

**Theorem 2.** *Let $r(n)$ be the number of runs among the first $n$ terms of the Kolakoski sequence, with $r(1) = 1$. Define*

$$e(i) \;=\; \#\{1 \le j \le r(i) : r(j) \text{ is even}\}.$$

*Recall from (2)*

$$k(j) = 2 \iff r(j) \text{ is even},$$

*so $e(i)$ also counts the number of 2's among the first $r(i)$ terms of $k$.*

*Then for all $n \ge 2$,*

$$r(n) \;=\; n \;-\; e\big(n-1\big). \tag{3}$$

*Proof.* Among the first $n$ symbols of $k$, there are $n$ opportunities to begin a new run, but each time the driving term is a 2 (equivalently for even run numbers in $r$ by (2)) it appends two identical symbols and thus contributes one "extra" symbol that does not start a run. Since there are $e\big(n-1\big)$ such events among the first $r(n-1)$ runs, exactly $e(n-1)$ of the $n$ positions do not start a new run. Hence

$$r(n) \;=\; n \;-\; e\big(n-1\big),$$

which proves the claim. □

## 3.2 A generalization for Kolakoski-(p,2p)

There are variants of the Kolakoski sequence, some of the form Kolakoski-$(p, q)$ where $p$ and $q$ indicate the symbols. When $p$ and $q$ have equal parity (both are odd or both are even) the sequence shows more structure [12]. For example, Kolakoski-(1,2) (A000002) still has an unknown density of 1's and 2's. By contrast, we do know the density of 2s and 4s in Kolakoski-(2,4) (A071928). Jeffrey Shallit comments on the OEIS entry that Kolakoski-(2,4) can be generated by a morphism. Using the theorem-proving software Walnut [9], he proves that $A071928(n)/2 = A157129(n)$ and $\sum_{k=1}^{n} A157129(k) = \frac{3n}{2} + O(1)$.

We now extend the formula $r(n)$ by incorporating, $p$, such that $r_p(n)$ corresponds to the run numbers for Kolakoski-$(p, 2p)$. For $p = 1$ we get the run numbers to the classical Kolakoski-(1,2) sequence, for $p = 2$ we get the run numbers for Kolakoski-(2,4) sequence. For $p = 3$ we get the run numbers to Kolakoski-(3,6), and so on. The proof is analogous to the case for $p = 1$.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k_{2,4}(n)$ (driving terms) | 2 | 2 | 4 | 4 | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 4 | 2 | 2 | 4 | 4 |
| $r_2(n)$ (run numbers) | 1 | 1 | 2 | 2 | 3 | 3 | $\underline{3}$ | $\underline{3}$ | 4 | 4 | $\underline{4}$ | $\underline{4}$ | 5 | 5 | 6 | 6 |

Table 2: Row 2 lists the Kolakoski-(2,4) sequence, $k_{2,4}(n)$. Underlined values in row 3 are "extra" symbols: positions where $k_{2,4}(r_2(n)) = 4$ have produced two "extra" symbols.

Since we use the same construction algorithm for all Kolakoski-$(p, 2p)$ sequences we generalize (2) to

$$k_{p,2p}(n) = \begin{cases} p, & \text{if } r_p(n) \text{ is odd;} \\ 2p, & \text{if } r_p(n) \text{ is even.} \end{cases} \tag{4}$$

**Theorem 3.** *Fix an integer $p \geq 1$. Let $r_p(n)$ be the number of runs in the first $n$ symbols of the Kolakoski sequence whose symbols are $p$ and $2p$, with $r_p(n) = 1$ for $1 \leq n \leq p$. Define*

$$e_p(i) \; = \; \#\big\{ 1 \leq j \leq r_p(i) : r_p(j) \text{ is even} \big\},$$

*Recall from* (4)

$$k_{p,2p}(j) = 2p \iff r_p(j) \text{ is even},$$

*so that $e_p(i)$ counts the number of $2p$-runs among the first $r_p(i)$ runs.*

*Then for every $n > p$,*

$$r_p(n) \; = \; \left\lceil \frac{n}{p} \right\rceil \; - \; e_p\big(n-p\big). \tag{5}$$

*Proof.* The minimum length of each run is $p$. Hence among the first $n$ symbols there are $\lceil n/p \rceil$ potential run-starts.

Whenever the driving term is $2p$ (equivalently for even run numbers in $r_p$ by (4)) the appended run has length $2p$. Its second half therefore prevents one of the $\lceil n/p \rceil$ potential run-starts from occurring. Conversely, a $p$-run occupies only a single run of length $p$ and creates no such suppressed start.

Among the runs that begin no later than the first $r_p(n-p)$ runs, exactly $e_p\big(n-p\big)$ are $2p$-runs, and each of them accounts for one lost run-start inside the first $n$ symbols. Subtracting these losses from the $\lceil n/p \rceil$ potential run starts gives

$$r_p(n) \; = \; \left\lceil \frac{n}{p} \right\rceil \; - \; e_p\big(n-p\big),$$

which proves the claim. □

# 4 Conclusion

We have shown that the seemingly intricate pattern of runs in the Kolakoski sequence admits a remarkably simple description: subtracting the number of "extra" symbols that even run numbers generate from the potential run starts yields the exact run index at every position. The same idea extends to all Kolakoski-$(p, 2p)$ sequences and therefore bridges the gap between the classical, still mysterious case $p = 1$ and the even-$p$ cases which we understand better. Several avenues remain open, notably extending the recurrence to arbitrary $(p, q)$ variants and formalizing the self-regulating dynamics of $r(n)$.

## 5 Acknowledgment

I thank the editors and volunteers who tirelessly maintain the On-Line Encyclopedia of Integer Sequences; their dedication has enabled Sequence Machine [10] to make these discoveries. I am also thankful to Jeffrey Shallit who took time to give me feedback on this article.

## References

[1] W. Kolakoski, Problem 5304: Self Generating Runs, *Amer. Math. Monthly*, Vol **72** (1965), 674. https://www.jstor.org/stable/2313883

[2] W. Kolakoski, Problem 5304, *Amer. Math. Monthly*, Vol. **73** (1966), 681–682. https://www.jstor.org/stable/2314839

[3] R. Oldenburger, Exponent trajectories in symbolic dynamics, *Transactions of the American Mathematical Society (3) 46*, 1939, pp. 453–466.

[4] N. J. A. Sloane et al., The On-Line Encyclopedia of Integer Sequences, https://oeis.org.

[5] B. Steinsky, A recursive formula for the Kolakoski sequence A000002, *J. Integer Seq.*, Vol. **9** (2006), Article 06.3.7.

[6] J-M. Fédou and G. Fici, Some remarks on differentiable sequences and recursivity, *J. Integer Seq.*, Vol. **13** (2010), Article 10.3.2.

[7] J. Nilsson, Letter frequencies in the Kolakoski sequence, *Acta Physics Polonica A*, Vol. **126** (2014). http://przyrbwn.icm.edu.pl/APP/PDF/126/a126z2p31.pdf

[8] K. Culik II, J. Karhumäki, and A. Lepistö, Alternating iteration of morphisms and the Kolakovski [sic] sequence, in G. Rozenberg and A. Salomaa, eds., *Lindenmayer Systems*, Springer-Verlag, 1992, pp. 93–103.

[9] H. Mousavi. Automatic theorem proving in Walnut. arXiv preprint arXiv:1603.06017 [cs.FL], 2016, http://arxiv.org/abs/1603.06017.

[10] J. Maiga, Sequence Machine https://sequencedb.net/.

[11] M. S. Keane et al. Ergodic theory and subshifts of finite type. *Ergodic theory, symbolic dynamics and hyperbolic spaces*, 1991, pp. 57–66.

[12] B. Sing, Kolakoski-($2m, 2n$) are limit-periodic model sets, *J. Math. Phys.*, 2003, pp. 899–912.

---

2020 *Mathematics Subject Classification*: Primary 68R15.
*Keywords:* kolakoski, Kolakoski-(p,q), run-length encoding, self-describing sequence

---

(Concerned with sequences A000002, A071928, A156253, A157129.)

---