A Recurrence Related to the Kolakoski Sequence

Jon Maiga
Sequence Machine

Abstract

The Kolakoski sequence is a self-describing sequence with many interesting properties. In this paper, we introduce a recurrence relation for a companion to the Kolakoski sequence. The recurrence demonstrates a resemblance to a well-known formula for Golomb's sequence. We conclude by raising the question whether the analytical methods employed in studying Golomb's sequence can be adapted to the Kolakoski sequence.

1 Introduction

The Kolakoski sequence [1, 2], named after the mathematician William Kolakoski, is an infinite sequence that self-encodes its run lengths. A *run* is a streak of equal terms. The Online Encyclopedia of Integer Sequences [3] defines the Kolakoski sequence, here denoted by k_n, as: "$\ldots k_n$ is the length of n-th run $\ldots$".

$$k_n = (1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, \ldots) = \text{A000002}$$

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 has an asymptotic equality of the number of ones and twos. Empirical evidence, for example from Nilsson [6], supports this possibility. Another open question asks if there exist a direct formula for the n-th term in the Kolakoski sequence. Previous studies, including [4, 5], have found recurrence relations for k_n and its companion sequences. In this study, we introduce a recurrence that closely resembles a well-known recurrence for Golomb's sequence. We conclude by posing the question of whether the analytical methods applied to Golomb's sequence can be extended to investigate the Kolakoski sequence.

One way to construct the Kolakoski sequence is to start with the terms $(1, 2, 2)$ and $n = 3$ (marked with *) and append the sequence with k_n symbols (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 described above, a complete run is appended. Consequently, for every n, we must alternate the symbol to append. This construction algorithm leads us to the equivalent mapping found by Culik II and Lepistö [7]:

$$\text{If } n \text{ is even then } 1 \to 2 \text{ and } 2 \to 22$$
$$\text{If } n \text{ is odd then } 1 \to 1 \text{ and } 2 \to 11$$

2 A companion sequence

To complement the Kolakoski sequence, we create a companion sequence denoted by a_n. We obtain a_n by appending the index n to a list for each term that is appended to the Kolakoski sequence during the construction process described earlier. We refer to the terms a_n as the *origin* of the corresponding terms k_n.

$$\begin{aligned} k_n &= (1,2,2,1,1,2,1,2,2,1,2,2,1,1,\ldots) = \text{A000002} \\ a_n &= (1,2,2,3,3,4,5,6,6,7,8,8,9,9,\ldots) = \text{A156253} \end{aligned}$$

The companion sequence, a_n, is known as A156253.

Example 2. The terms $k_4 = 1$ and $k_5 = 1$ originated from the 2 at position 3 in k, resulting in $a_4 = 3$ and $a_5 = 3$. The term $k_6 = 2$ originated from the fourth term, resulting in $a_6 = 4$.

In the formula section of A000002, Benoit Cloitre provides a relationship between a_n and k_n:

Lemma 3. *For positive integers n, we have*

$$k_n = \frac{3 + (-1)^{a_n}}{2} = \gcd(a_n, 2)$$

Proof. We recall the mapping by Culik II and Lepistö [7] for the Kolakoski sequence:

$$\text{If } n \text{ is even then } 1 \to 2 \text{ and } 2 \to 22$$
$$\text{If } n \text{ is odd then } 1 \to 1 \text{ and } 2 \to 11$$

This means that ones in the Kolakoski sequence originate from odd values of n, while twos originate from even values of n. By definition, the terms of the sequence a_n represent the origin of the corresponding terms in k_n. Thus, if a_n is odd, then k_n equals one, and if a_n is even, then k_n equals two. We can now substitute odd and even values of a_n into $\gcd(a_n, 2)$ to correctly obtain one and two, respectively. □

Because of the self-describing property of the Kolakoski sequence, the relationship between the sequences k_n and a_n also reveals the length of the run corresponding to the term k_n and a_n.

We will now prove a recurrence relation for a_n that was conjectured by Sequence Machine [8]. The statement of the theorem is as follows:

Theorem 4. *Let $a_1 = 1$. For $n > 1$ we have $a_n = a_{n-\gcd(a_{a_{n-1}},2)} + 1$.*

Proof. First note that $\gcd(a_{a_{n-1}}, 2)$ uses Lemma 3 to determine the length of the run belonging to the previous term by substituting n with a_{n-1}. We will use induction to prove the theorem.

Base case: Let $n = 2$. We know that $a_1 = 1$. Then, we have $a_2 = a_{2-\gcd(a_{a_{2-1}},2)} + 1 = a_{2-\gcd(a_1,2)} + 1 = a_{2-1} + 1 = a_1 + 1 = 2$. The base case holds.

Inductive step: Show that for every positive k, if $a_k = a_{k-\gcd(a_{a_{k-1}},2)} + 1$ holds, then a_{k+1} also holds. For $k > 1$, there are three possible cases:

1. If the run length for a_k is one, $\gcd(a_{a_k}, 2) = 1$, we expect to start a new run: $a_{k+1} = a_k + 1$. By the induction hypothesis we have $a_{k+1} = a_{k+1-\gcd(a_{a_k},2)} + 1 = a_{k+1-1} + 1 = a_k + 1$.

   Example 5. $(4, 5, 6) = (a_{k-1}, a_k, a_{k+1})$

2. If the run length for a_k is two, $\gcd(a_{a_k}, 2) = 2$, and is the first term in its run, we expect to extend the run: $a_{k+1} = a_k$. By the induction hypothesis we have $a_{k+1} = a_{k+1-\gcd(a_{a_k},2)} + 1 = a_{k+1-2} + 1 = a_{k-1} + 1 = a_k$.

   Example 6. $(5, 6, 6) = (a_{k-1}, a_k, a_{k+1})$

3. If the run length for a_k is two, $\gcd(a_{a_k}, 2) = 2$, and is the second term in its run, we expect to start a new run: $a_{k+1} = a_k + 1$. By the induction hypothesis we have $a_{k+1} = a_{k+1-\gcd(a_{a_k},2)} + 1 = a_{k+1-2} + 1 = a_{k-1} + 1 = a_k + 1$.

   Example 7. $(6, 6, 7) = (a_{k-1}, a_k, a_{k+1})$

In all cases, we have shown that the theorem holds for $n = k + 1$. By induction, the theorem holds for all positive integers n. □

We now have a formula for a_n and indirectly for k_n (indirectly, in the sense that $k_n = \gcd(a_n, 2)$). In the formula section of A156253, there is a conjectured formula:

$$a(n) = (a(a(n-1)) \bmod 2) + a(n-2) + 1$$

A proof for this variation of a_n can be constructed in a similar way.

3 Similarity to Golomb's sequence

In A156253, N. J. A. Sloane made the following comments:

> This seems to be A001462 rewritten so the run lengths are given by A000002.
> ...
>
> Note that the Kolakoski sequence A000002 and Golomb's sequence A001462 have very similar definitions, although the asymptotic behavior of A001462 is well-understood, while that of A000002 is a mystery.

The Online Encyclopedia of Integer Sequences [3] definition of Golomb's sequence, here denoted by g_n, is similar to Kolakoski's: "... g(n) is the number of times n occurs ..."

$g(n) = (1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, \dots) =$ A001462

The entry for A001462 has a formula by Colin Mallows:

$$g_n = g_{n-g_{g_{n-1}}} + 1$$

The formula g_n indeed resembles of a_n:

$$a_n = a_{n-\gcd(a_{a_{n-1}},2)} + 1$$

The only difference between a_n and g_n is the additional greatest common divisor operation. This operation limits the lengths of the runs to either one or two, depending on the parity of $a_{a_{n-1}}$. This relationship between a_n and g_n confirms N. J. A. Sloane's first observation.

The formula for Golomb's sequence, g_n works similarly to a_n. We can compute the term, g_n, by looking at the length of the previous term's run, $g_{g_{n-1}}$. We demonstrate the calculation with two examples:

Example 8. For n = 15 (highlighted in bold), $(1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, \mathbf{6})$. We examine the run length of the previous term: $g_{g_{14}} = g_6 = 4$ indicating it should be four consecutive sixes. We take $n - 4 = 15 - 4 = 11$, which is the last term in the run of fives. Substituting $g_{11} = 5$ into g_n extends the run with another six, $g_{15} = 5 + 1 = 6$.

Example 9. For n = 16 (highlighted in bold), $(1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, \mathbf{7})$. We examine the run length of the previous term $g_{g_{15}} = g_6 = 4$, indicating it should be four consecutive sixes. We take $n - 4 = 16 - 4 = 12$, which is on the first six in the run of sixes. Substituting $g_{12} = 6$ into g_n starts the next run of sevens, $g_{16} = 6 + 1 = 7$.

Marcus and Fine [9] showed that Golomb's sequence can be approximated. Let ϕ be the Golden Ratio and $E(n)$ an error term, then $g_n \approx \phi^{2-\phi} n^{\phi-1} + E(n)$. For more details, refer to the formula section for A001462.

4 Conclusion

In this paper, we have introduced an indirect recurrence relation for the Kolakoski sequence. The recurrence bears a close resemblance to a formula for Golomb's sequence. We are interested in exploring whether the analytical techniques applied Golomb's sequence can be extended to the new recurrence. However, the added complexity of the greatest common divisor operation may make it impossible. Nonetheless, studying the reasons behind this difficulty could provide valuable insights.

5 Acknowledgment

I extend my full gratitude to the dedicated editors and volunteers who tirelessly contribute to the unparalleled Online Encyclopedia of Integer Sequences. Without their efforts the discoveries made by Sequence Machine would not have been possible.

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

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

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

[6] 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

[7] 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.

[8] J. Maiga, Sequence Machine, Computer-generated formulas for A156253 https://sequencedb.net/s/A156253.

[9] D. Marcus and N. J. Fine, Solutions to problem 5407, *Amer. Math. Monthly*, Vol. **74** (1967). http://www.jstor.org/stable/2314296

2020 *Mathematics Subject Classification*: Primary 68R15.
Keywords: Kolakoski, Golomb

(Concerned with sequences A000002, A001462, A156253.)