Journal of Integer Sequences, Vol. 14 (2011),
Article 11.2.1

Bounds for the Kolakoski Sequence

Olivier Bordellès
2 allée de la Combe
43000 Aiguilhe
France
borde43@wanadoo.fr

Benoit Cloitre
19 rue Louise Michel
92300 Levallois-Perret
France
benoit7848c@yahoo.fr

Abstract

The Kolakoski sequence (K_n) is perhaps one of the most famous examples of self-describing sequences for which some problems are still open. In particular, one does not know yet whether the density of 1's in this sequence is equal to $\frac{1}{2}$. This work, which does not answer this question, provides explicit bounds for the main sequences related to (K_n). The proofs rest on a new identity involving the partial sums of (K_n) and on Dirichlet's pigeonhole principle which allows us to improve notably on the error-term.

1 Introduction

In 1965, Kolakoski [7] introduced an example of a self-generating sequence by creating the sequence defined in the following way.

Definition 1.

- ⋆ We call *block* all sets of one or more identical digits. The number of digits in a block is the *length* of the block.

⋆ The *Kolakoski sequence* is the sequence (K_n) of blocks of 1's or 2's defined by

$$\begin{cases} K_1 &= 1 \\ K_n &= \text{length of the } n\text{-th block.} \end{cases}$$

⋆ We also define the partial sums

$$S_n = \sum_{j=1}^{n} K_j.$$

⋆ The three following sequences (k_n), (o_n) and (t_n) are related to (K_n) and defined by

$$\begin{cases} k_0 &= 0 \\ k_n &= \min\limits_{1 \le j \le n} \{j : S_j \ge n\} \quad (n \ge 1) \end{cases}$$

and

$$o_n := |\{1 \le j \le n : K_j = 1\}| \quad \text{and} \quad t_n := |\{1 \le j \le n : K_j = 2\}|$$

where $|E|$ means the number of elements of the finite set E.

Remark 2. We have $o_n + t_n = n$ and since

$$S_n = \sum_{j=1}^{n} K_j = \sum_{\substack{j=1 \\ K_j=1}}^{n} 1 + 2 \sum_{\substack{j=1 \\ K_j=2}}^{n} 1 = o_n + 2t_n$$

we infer $S_n = 2n - o_n = n + t_n$, $o_{n+1} - o_n = 2 - K_{n+1}$ and $t_{n+1} - t_n = K_{n+1} - 1$.

The following table shows the first terms of the sequences defined above.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
K_n	1	2	2	1	1	2	1	2	2	1	2	2	1	1	2	1	1
n-th block	1	2		3		4	5	6		7	8		9		10	11	
k_n	1	2		3		4	5	6		7	8		9		10	11	
S_n	1	3	5	6	7	9	10	12	14	15	17	19	20	21	23	24	25
o_n	1	1	1	2	3	3	4	4	4	5	5	5	6	7	7	8	9
t_n	0	1	2	2	2	3	3	4	5	5	6	7	7	7	8	8	8

This sequence, which belongs to the online Encyclopedia of Integer Sequences [10], has been studied by many authors (see [2, 3, 4, 5, 11] for instance), and some conjectures have been made. In particular, no one knows yet whether the density of the 1's is equal to $\frac{1}{2}$, e.g.

$$\lim_{n\to\infty} \frac{o_n}{n} = \frac{1}{2}. \tag{1}$$

This problem seems to be very tricky. It is equivalent to proving that

$$S_n = \frac{3n}{2} + o(n) \quad \text{or} \quad k_n = \frac{2n}{3} + o(n) \quad (n \longrightarrow \infty). \tag{2}$$

These estimates come from the asymptotic formula

$$\sum_{j=1}^{n} (-1)^{k_j} = o(n) \quad (n \longrightarrow \infty) \tag{3}$$

that we conjecture (See Corollary 10). Such sums are frequent in number theory. For instance, the Mertens function $M(n) = \sum_{j=1}^{n} \mu(j)$, where μ is the Möbius function, or the summatory function $L(n) = \sum_{j=1}^{n} \lambda(j)$ of the Liouville λ-function, are strongly related to the prime number theorem. Another interesting example can be found in the paper [9] in which an "almost alternating" sum, related to the Beatty's sequences, is investigated. We did not manage to prove one or the other estimates (1), (2) or (3). The sequence (k_n), with its fractal behavior, seems to prevent the use of ordinary tools to treat such sums (generating functions, convolution identities, periodicity, discrepance theory, specific arithmetic properties, etc). From then on, it could be interesting to provide unconditional effective bounds for the sequences (k_n), (S_n), (t_n) and (o_n). The following estimates, valid for all positive integers n, are obvious (see Lemma 6 for (k_n)).

$$n \le S_n < 2n \quad \text{and} \quad \frac{n}{2} < k_n \le n$$

$$0 \le t_n < n \quad \text{and} \quad 0 < o_n \le n.$$

The aim of this work is to give better bounds than the above trivial inequalities. More precisely, we will first prove the following result.

Theorem 3. *For all positive integers n, we have*

$$\frac{4n}{3} - 1 \le S_n \le \frac{5n}{3} \quad \textit{and} \quad \frac{3n}{5} \le k_n \le \frac{3n}{4} + \frac{3}{2}.$$

$$\frac{n}{3} - 1 \le t_n \le \frac{2n}{3} \quad \textit{and} \quad \frac{n}{3} \le o_n \le \frac{2n}{3} + 1.$$

In a second step, we use Dirichlet's pigeonhole principle to estimate the remainder term in a more subtle way. We will establish the following improvement.

Theorem 4. *Let* $\alpha \doteq 0,6764\ldots$ *be the unique root of the polynomial* $P = X^3 - 26X^2 + 26X - 6$ *in the interval* $\left(\frac{1}{2}, 1\right)$ *and we set*

$$\beta := \frac{\alpha^2 + \alpha - 1}{6(3\alpha - 1)} \doteq 0.021703504\ldots$$

For all positive integers n*, we have*

$$\left(\frac{3}{2} - \beta\right) n - 6 \leq S_n \leq \left(\frac{3}{2} + \beta\right) n + 6 \quad \text{and} \quad \left(\frac{\alpha}{3\alpha - 1}\right) n - 4 \leq k_n \leq \alpha n + 5.$$

$$\left(\frac{1}{2} - \beta\right) n - 6 \leq t_n \leq \left(\frac{1}{2} + \beta\right) n + 6 \quad \text{and} \quad \left(\frac{1}{2} - \beta\right) n - 6 \leq o_n \leq \left(\frac{1}{2} + \beta\right) n + 6.$$

Thus we have unconditionally

$$\left|\frac{o_n}{n} - \frac{1}{2}\right| \leq 0.021703504\ldots + \frac{6}{n}.$$

Although these bounds are still far from (2), they improve on the results established in [8] if n is sufficiently large, where it is proven that, if the limit $\lim_{n\to\infty}(o_n/n)$ exists, then

$$\left|\lim_{n\to\infty} \frac{o_n}{n} - \frac{1}{2}\right| \leq \frac{17}{762} \doteq 0.0223097\ldots$$

and our bounds are even slightly better than their "semi-rigorous" bound

$$\left|\lim_{n\to\infty} \frac{o_n}{n} - \frac{1}{2}\right| \leq \frac{1}{46} \doteq 0.02173913\ldots$$

Nevertheless, it should be mentioned that Theorem 4 does not improve on Chvátal's results (see [4]). The sketch of the proof is the following one.

- ⋆ We first establish Proposition 12 which prompts to introduce the set E_n of indexes $j \in \{1, \ldots, n\}$ such that $K_j = 2$ and shows it is sufficient to prove that there are more or less as many even integers as odd integers.
- ⋆ Since we did not manage to prove this assertion, we studied the gaps between consecutive elements of E_n and showed that it is sufficient to establish that there are almost as many even integers as odd integers in the subset $E_n(2)$ of E_n made up of integers $e \in E_n$ such that $e + 2 \in E_n$ (Lemma 17).
- ⋆ Then we studied the gaps between consecutive elements of $E_n(2)$ (Lemmas 15 and 18) and defined two subsets A_n and B_n of $E_n(2)$ (Lemmas 20 and 22) to refine the estimates.

Although they are getting smaller and smaller, the subsets become gradually harder and harder to handle, for they require the knowledge of longer and longer runs of the sequence (K_n) (Lemma 21). The proofs, using disjunction arguments, become also longer and longer, since the number of cases gradually increases. Finally, we use Dirichlet's pigeonhole principle applied to the sets A_n and B_n (Lemma 22) to accurately estimate the remainder term. However, such a strategy does not seem to prove estimates of the form $\ll n^\theta$ with $0 \le \theta < 1$. Finally, it should be mentioned that all the numeric computations have been made using PARI/GP system [12].

2 Some properties of the sequence (k_n)

Steinsky [11] proved the following results.

Lemma 5. *Let $n \ge 2$ be an integer. Then we have*

(i) $S_{k_{n-1}} \in \{n-1, n\}$ and $k_n = k_{n-1} + n - S_{k_{n-1}}$.

(ii) $k_n - k_{n-1} = |K_n - K_{n-1}|$.

(iii) $k_n = n - t_{k_{n-1}}$.

We deduce the following consequences.

Lemma 6. *The sequence (k_n) is non-decreasing and, for all positive integers n, we have*

$$\frac{n+1}{2} \le k_n \le n.$$

Proof. Using Lemma 5 (ii), we get $k_n - k_{n-1} = |K_n - K_{n-1}| \ge 0$. The inequality $k_n \le n$ comes from Lemma 5 (iii). Moreover, we also have $S_{k_n} \in \{n, n+1\}$ by Lemma 5, so that $2k_n > S_{k_n} \ge n$ and hence $k_n > n/2$. Since k_n is an integer, we get the desired lower bound. □

Lemma 7. *For all integers $n \ge 2$, the number of blocks of length equal to 2 between the letters K_1 and K_n is given by $n - k_n$.*

Proof. Using Lemma 5 (ii), for all integers $j \ge 2$, we have

$$(K_{j-1}, K_j) = (1,1) \text{ or } (2,2) \iff k_{j-1} = k_j$$

so that the number in question is

$$\sum_{\substack{j=2 \\ k_{j-1}=k_j}}^{n} 1 = \sum_{j=2}^{n} 1 - \sum_{\substack{j=2 \\ k_{j-1}\ne k_j}}^{n} 1 = n - 1 - \sum_{j=2}^{n} (k_j - k_{j-1}) = n - k_n.$$

This proves Lemma 7. □

3 Some properties of the sequence (K_n)

In what follows, we list some situations which cannot appear in the sequence (K_n). We will call *run* any (finite or not) subsequence of the sequence (K_n). The next lemma lists some of the avoided runs. The easy proof of this fact, left to the reader (see [2, 3] for instance), is based upon the property that the sequence does not contain cubes, that is, runs of the form xxx.

Lemma 8. *The following runs cannot appear in the sequence (K_n).*

(i) $(1, 1, 1, \dots)$ and $(2, 2, 2, \dots)$.

(ii) $(1, 2, 1, 2, 1)$, $(2, 1, 2, 1, 2)$, $(1, 1, 2, 2, 1, 1)$ and $(2, 2, 1, 1, 2, 2)$.

(iii) $(1, 1, 2, 1, 1, 2, 1, 1, 2)$, $(2, 2, 1, 2, 2, 1, 2, 2, 1)$, $(1, 2, 2, 1, 2, 2, 1, 2, 2)$ and $(2, 1, 1, 2, 1, 1, 2, 1, 1)$.

(iv) $(2, 1, 2, 1, 1, 2, 1, 1, 2, 1, 2)$ and $(2, 1, 2, 1, 1, 2, 2, 1, 2, 1, 2)$.

(v) $(2, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2)$ and $(2, 1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2)$.

(vi) $(2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2)$ and $(2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 2)$.

(vii) $(2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2)$.

Proposition 9. *For all positive integers n, we have*

$$K_{S_n} = \frac{3 + (-1)^n}{2} \quad \textit{and} \quad K_{1+S_n} = \frac{3 + (-1)^{n+1}}{2}.$$

Proof. We first notice that, since the first block is $\{1\}$ and the other blocks alternate between 1's and 2's, then the n-th block is either $\{1\}$ or $\{1, 1\}$ (resp., either $\{2\}$ or $\{2, 2\}$) if n is odd (resp., even). The next step is the proof by induction of the following assertion.

For all positive integers n, S_n is the index of the last element of the n-th block. (4)

The assertion (4) is clearly true for $n = 1$. Assume it is true for some $n \geq 1$. By induction hypothesis, we get

$$S_{n+1} = S_n + K_{n+1} = \left\{ \begin{array}{c} S_n + 1 \\ \text{or} \\ S_n + 2 \end{array} \right. = \left\{ \begin{array}{l} \text{index of the unique element of the } (n+1)\text{-th block} \\ \text{or} \\ \text{index of the last element of the } (n+1)\text{-th block} \end{array} \right.$$

since, if $K_{n+1} = 1$ (resp., $K_{n+1} = 2$), then the length of the $(n + 1)$-th block is equal to 1 (resp., 2) and there is only one element (resp., two elements) in this block. This proves (4).

Now we can prove Proposition 9. Indeed, K_{S_n} is the value in the sequence (K_n) of the index of the last element of the n-th block, which is equal to 2 if n is even and to 1 if n is odd, and the asserted result for K_{S_n} follows. We get the formula for K_{1+S_n} in a similar way, since $S_n + 1$ is the index of the first element of the $(n + 1)$-th block. This proves Proposition 9. $\square$

Corollary 10. *For all positive integers n, we have*

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

Proof. By Lemma 5, we have

$$S_{k_n} = \begin{cases} n, & \text{if } k_n \neq k_{n+1}; \\ n+1, & \text{if } k_n = k_{n+1}; \end{cases}$$

so that

$$K_n = K_{S_{k_n}} = \frac{3 + (-1)^{k_n}}{2} \quad \text{if} \quad k_n \neq k_{n+1}$$

and

$$K_{n+1} = K_{S_{k_n}} = \frac{3 + (-1)^{k_n}}{2} = \frac{3 + (-1)^{k_{n+1}}}{2} \quad \text{if} \quad k_n = k_{n+1}$$

which completes the proof. □

Corollary 11. *For all positive integers n, we have*

$$\sum_{j=1}^{n} (-1)^j K_j = 2S_{S_n} - 3S_n.$$

Proof. By induction, the case $n = 1$ being obvious. Assume the equality is true for some positive integer n. Noticing that $[S_n + 1, S_{n+1}] \cap \mathbb{Z} = \{S_n + 1\}$ or $\{S_n + 1, S_{n+1}\}$ according to $K_{n+1} = 1$ or 2 and using Proposition 9 and the induction hypothesis, we get

$$\begin{aligned} 2S_{S_{n+1}} - 3S_{n+1} &= \begin{cases} 2S_{S_n} - 3S_n + 2K_{S_n+1} - 3, & \text{if } K_{n+1} = 1; \\ \\ 2S_{S_n} - 3S_n + 2K_{S_n+1} - 3 + 2K_{S_{n+1}} - 3, & \text{if } K_{n+1} = 2; \end{cases} \\ \\ &= \begin{cases} 2S_{S_n} - 3S_n + (-1)^{n+1}, & \text{if } K_{n+1} = 1; \\ \\ 2S_{S_n} - 3S_n + 2(-1)^{n+1}, & \text{if } K_{n+1} = 2; \end{cases} \\ \\ &= 2S_{S_n} - 3S_n + (-1)^{n+1} K_{n+1} \\ \\ &= \sum_{j=1}^{n} (-1)^j K_j + (-1)^{n+1} K_{n+1} = \sum_{j=1}^{n+1} (-1)^j K_j \end{aligned}$$

as required. The proof is complete. □

4 A useful identity

In what follows, $\lfloor x \rfloor$ is the integer part of $x \in \mathbb{R}$.

Proposition 12. *For all positive integers n, we have*

$$S_n = \frac{3n}{2} + \frac{1}{2}\left(\sum_{\substack{j=1\\K_{2j}=2}}^{\lfloor k_n/2\rfloor} 1 - \sum_{\substack{j=1\\K_{2j+1}=2}}^{\lfloor (k_n-1)/2\rfloor} 1\right) + \frac{(-1)^{k_n}-1}{4} - \frac{(-1)^{k_n}c_n}{2}$$

where $c_n = \begin{cases} 1, & \text{if } K_n = K_{n+1};\\ 0, & \text{if } K_n \neq K_{n+1}.\end{cases}$.

Proof. Using Corollary 11 with k_n instead of n, we get

$$\begin{aligned}
\sum_{j=1}^{k_n}(-1)^j K_j &= 2S_{S_{k_n}} - 3S_{k_n}\\
&= \begin{cases} 2S_{n+1} - 3n - 3, & \text{if } K_n = K_{n+1};\\ 2S_n - 3n, & \text{if } K_n \neq K_{n+1};\end{cases}\\
&= \begin{cases} 2S_n - 3n + 2K_{n+1} - 3, & \text{if } K_n = K_{n+1};\\ 2S_n - 3n, & \text{if } K_n \neq K_{n+1};\end{cases}\\
&= 2S_n - 3n + (-1)^{k_n}c_n,
\end{aligned}$$

where we used Corollary 10. Besides, by Abel summation, we have for all $m \geq 1$

$$\begin{aligned}
\sum_{j=1}^{m}(-1)^j K_j &= (-1)^m \sum_{j=1}^{m} K_j - \sum_{j=1}^{m-1}\left\{(-1)^{j+1} - (-1)^j\right\}\sum_{h=1}^{j} K_h\\
&= (-1)^m S_m + 2\sum_{j=1}^{m-1}(-1)^j S_j = 2\sum_{j=1}^{m}(-1)^j S_j - (-1)^m S_m
\end{aligned}$$

so that

$$S_n = \frac{3n}{2} + \sum_{j=1}^{k_n}(-1)^j S_j - \frac{(-1)^{k_n}}{2}\left(S_{k_n} + c_n\right)$$

The relation $S_j = j + t_j$ implies

$$S_n = \frac{3n}{2} - \frac{(-1)^{k_n}t_{k_n}}{2} + \sum_{j=1}^{k_n}(-1)^j t_j + \frac{(-1)^{k_n}-1}{4} - \frac{(-1)^{k_n}c_n}{2}.$$

We conclude the proof by using the identity

$$\sum_{j=1}^{m}(-1)^j a_j = \frac{(-1)^m a_m - a_1}{2} + \frac{1}{2}\sum_{j=2}^{m}(-1)^j\left(a_j - a_{j-1}\right) \tag{5}$$

which can be viewed as a discrete analogue of Boole's summation formula of order 1 (see [1, 6] for instance) and which implies here that

$$\begin{aligned}\sum_{j=1}^{k_n}(-1)^j t_j &= \frac{(-1)^{k_n} t_{k_n}}{2} + \frac{1}{2}\sum_{j=2}^{k_n}(-1)^j\left(K_j - 1\right)\\ &= \frac{(-1)^{k_n} t_{k_n}}{2} + \frac{1}{2}\sum_{\substack{j=2\\ K_j=2}}^{k_n}(-1)^j\\ &= \frac{(-1)^{k_n} t_{k_n}}{2} + \frac{1}{2}\left(\sum_{\substack{j=1\\ K_{2j}=2}}^{\lfloor k_n/2\rfloor} 1 - \sum_{\substack{j=1\\ K_{2j+1}=2}}^{\lfloor (k_n-1)/2\rfloor} 1\right),\end{aligned}$$

which concludes the proof. □

5 Proof of Theorem 3

We are now in a position to prove Theorem 3, for which one may suppose $n \geq 4$ and whose results come from trivial estimates of the sums of Proposition 12. Considering the four cases $(K_n, K_{n+1}) = (1,1),\ (2,2),\ (1,2)$ and $(2,1)$, the trivial estimate

$$\left|\sum_{\substack{j=1\\ K_{2j}=2}}^{\lfloor n/2\rfloor} 1 - \sum_{\substack{j=1\\ K_{2j+1}=2}}^{\lfloor (n-1)/2\rfloor} 1\right| \leq t_n - 1 \tag{6}$$

applied to Proposition 12 allows us to get the inequalities

$$n + \frac{k_n}{2} - \frac{1}{2} \leq S_n \leq 2n - \frac{k_n}{2}. \tag{7}$$

On the other hand, the inequality

$$\sum_{\substack{j=1\\ K_{2j}=2}}^{\lfloor n/2\rfloor} 1 \leq \frac{n}{2}$$

used in Proposition 12 implies that

$$S_n \leq \frac{3n}{2} + \frac{k_n}{4}$$

and the use of (7) rewritten as $k_n \leq 4n - 2S_n$ gives

$$2S_n \leq 5n - S_n$$

which implies the asserted upper bound for S_n. The lower bound for k_n comes from the inequality $S_{k_n} \geq n$ and from the upper bound S_n formerly established. We get the upper bound for t_n and the lower bound for o_n with the help of the relations $t_n = S_n - n$ and $o_n = n - t_n$. To prove the lower bound for t_n, we introduce the set $E_n := \{j \in \{1, \ldots, n\} : K_j = 2\}$ and show that the gaps between two consecutive elements of E_n can only be equal to 1, 2 or 3. Indeed, suppose there exists $e \in E_n$ such that e and $e + a$ are consecutive in E_n with $a \geq 4$. Then we have $K_{e+1} = K_{e+2} = K_{e+3} = 1$ which is impossible by Lemma 8 (i). We infer that E_n is a subset of $\{1, \ldots, n\}$ whose gaps between two consecutive elements are at most 3, so that

$$n \leq 3\left(|E_n| + 1\right) = 3\left(t_n + 1\right)$$

which implies the asserted lower bound. The relation $S_n = n + t_n$ gives the lower bound for S_n, the relation $o_n = n - t_n$ gives the upper bound for o_n and the inequality $k_n \leq n + 1 - t_{k_n}$ gives the upper bound for k_n. Theorem 3 is completely proven. $\square$

6 Proof of Theorem 4

The proof of Theorem 3 shows that we may take advantage of the specificities of the sequence (K_n). In this section, we study these specificities in a more accurate way. We first check the validity of the inequalities for all $n \in \{1, \ldots, 99\}$, and we may suppose $n \geq 100$. We use the set E_n introduced in the proof of Theorem 3 and define the two following subsets.

$$F_n := \{2j : 1 \leq j \leq \lfloor n/2 \rfloor \text{ and } K_{2j} = 2\}$$

$$G_n := \{2j + 1 : 1 \leq j \leq \lfloor (n-1)/2 \rfloor \text{ and } K_{2j+1} = 2\}$$

and, for all sets of integers E and all positive integers a and b, we set

$$E\left(a\right) := \{e \in E : e \text{ and } e + a \text{ are consecutive in } E\} \quad \text{and} \quad E(a, b) := (E(a))(b).$$

It is obvious that the elements of F_n (resp., $F_n(2)$) are the even indexes of E_n (resp., $E_n(2)$), and the elements of G_n (resp., $G_n(2)$) are the odd indexes of E_n (resp., $E_n(2)$).

Example 13. Here are the main sets used in the proof ($n = 100$).

- $\star$ $E_{100} = \{2, 3, 6, 8, 9, 11, 12, 15, 18, 19, 21, 24, 26, 27, 30, 33, 35, 36, 38, 39, 42, 44, 45, 47, 50, 53, 54, 56, 57, 60, 62, 63, 65, 66, 69, 72, 74, 75, 77, 80, 81, 83, 84, 87, 89, 90, 92, 93, 96, 99, 100\}$.

- $\star$ $F_{100} = \{2, 6, 8, 12, 18, 24, 26, 30, 36, 38, 42, 44, 50, 54, 56, 60, 62, 66, 72, 74, 80, 84, 90, 92, 96, 100\}$.

- $\star$ $G_{100} = \{3, 9, 11, 15, 19, 21, 27, 33, 35, 39, 45, 47, 53, 57, 63, 65, 69, 75, 77, 81, 83, 87, 89, 93, 99\}$.

- $\star$ $E_{100}(1) \cup E_{100}(3) = \{2, 3, 8, 11, 12, 15, 18, 21, 26, 27, 30, 35, 38, 39, 44, 47, 50, 53, 56, 57, 62, 65, 66, 69, 74, 77, 80, 83, 84, 89, 92, 93, 96, 99\}$.

- $\star$ $E_{100}(2) = \{6, 9, 19, 24, 33, 36, 42, 45, 54, 60, 63, 72, 75, 81, 87, 90\}$.

- $\star$ $F_{100}(2) = \{6, 24, 36, 42, 54, 60, 72, 90\}$ and $G_{100}(2) = \{9, 19, 33, 45, 63, 75, 81, 87\}$.

- $\star$ $A_{100} = \{9, 36, 54, 75, 81\}$ and $B_{100} = \{6, 19, 24, 33, 42, 45, 60, 63, 72, 87\}$.

Set	E_{100}	F_{100}	G_{100}	$E_{100}(1) \cup E_{100}(3)$	$E_{100}(2)$	$F_{100}(2)$	$G_{100}(2)$	A_{100}	B_{100}
Cardinal	$t_{100} = 51$	26	25	$100 - k_{100} = 34$	16	8	8	5	10

The main purpose of this section is to provide better estimates of $\Big||F_n| - |G_n|\Big|$ than the trivial bound (6) used in the proof of Theorem 3. To this end, we first study some properties of the sets $E_n(1) \cup E_n(3)$ and $E_n(2)$.

Lemma 14. *We have* $|E_n(1) \cup E_n(3)| = n - k_n + \varepsilon_n$ *where* $\varepsilon_n \in \{-1, 0, 1\}$.

Proof. It has been seen above that the gaps between consecutive elements of E_n are at most equal to 3. In particular, we get

$$\begin{aligned} &e \in E_n(1) \iff K_e = K_{e+1} = 2. \\ &e \in E_n(3) \iff K_e = K_{e+3} = 2 \text{ and } K_{e+1} = K_{e+2} = 1. \end{aligned}$$

We infer that $|E_n(1) \cup E_n(3)|$ counts the number of blocks of length 2 between K_1 and K_n, possibly plus or minus 1 block, so that

$$|E_n(1) \cup E_n(3)| = n - k_n + \varepsilon_n$$

by Lemma 7. $\square$

Lemma 15. *The gaps between two consecutive elements of $E_n(2)$ can only be equal to* $3,\ 5,\ 6,\ 9$ *or* 10.

Proof.

1. We first prove that, in $E_n(2)$, there do not exist consecutive elements with gaps equal to 1, 2, 4, 7 or 8. Indeed, if a gap between two consecutive elements of $E_n(2)$ is :

⋆ equal to 1, then there exists $e \in E_n$ such that $e, e+1, e+2, e+3 \in E_n$ and therefore we have

$$(K_e, \dots, K_{e+3}) = (2, 2, 2, 2)$$

which is impossible by Lemma 8 (i).

⋆ equal to 2, then there exists $e \in E_n$ such that $e, e+2, e+4 \in E_n$ and we have

$$(K_e, \dots, K_{e+4}) = (2, 1, 2, 1, 2)$$

which is impossible by Lemma 8 (ii).

⋆ equal to 4, then there exists $e \in E_n$ such that $e, e+2, e+4, e+6 \in E_n$ and we have

$$(K_e, \dots, K_{e+6}) = (2, 1, 2, 1, 2, 1, 2)$$

which is impossible by Lemma 8 (ii).

⋆ equal to 7, then there exists $e \in E_n$ such that

$$(K_{e+4}, \dots, K_{e+8}) = (1, 2, 1, 2, 1).$$

Indeed, we have $K_{e+2} = K_{e+7} = 2$ by hypothesis, and $K_{e+4} = K_{e+6} = 1$ otherwise we have $e+2, e+4 \in E_n(2)$. We infer that $K_{e+5} = 2$, otherwise $(K_{e+4}, K_{e+5}, K_{e+6}) = (1, 1, 1)$ and $K_{e+8} = 1$, for $K_{e+9} = 2$. This run is impossible by Lemma 8 (ii).

⋆ equal to 8, then there exists $e \in E_n$ such that

$$(K_e, \dots, K_{e+10}) = (2, 1, 2, 1, 1, 2, 1, 1, 2, 1, 2).$$

Indeed, we have $K_e = K_{e+2} = K_{e+8} = K_{e+10} = 2$, hence $K_{e+1} = K_{e+9} = 1$, but also $K_{e+4} = 1$ otherwise $e+2 \in E_n(2)$, and similarly $K_{e+6} = 1$ otherwise $e+4 \in E_n(2)$, and therefore $K_{e+5} = 2$ otherwise there exist three consecutive 1's, and hence $K_{e+3} = K_{e+7} = 1$ otherwise $e+3,\ e+5 \in E_n(2)$. This run is impossible by Lemma 8 (iv).

2. Now it only remains to be shown that gaps between two consecutive elements of $E_n(2)$ are at most equal to 10. Suppose that e and $e+a$ are consecutive in $E_n(2)$ with $a \geq 11$. We then have

$$K_e = K_{e+2} = 2 \quad \text{and} \quad K_{e+1} = K_{e+4} = 1$$

for if $K_{e+4} = 2$, then $e+2 \in E_n(2)$ contrary to the hypothesis. Let us treat two cases.

⋆ **1st case** : $K_{e+3} = 1$. Then $K_{e+5} = 2$.

(i) If $K_{e+6} = 1$, then we have

$$(K_e, \dots, K_{e+10}) = (2,1,2,1,1,2,1,1,2,2,1).$$

Indeed, if $K_{e+7} = 2$, then we have $e+5 \in E_n(2)$. We infer that $K_{e+8} = 2$. Moreover, we have $K_{e+9} = 2$, otherwise we get a run $(1,1,2,1,1,2,1,1)$ which is impossible. Also, $K_{e+10} = 1$, otherwise $e+8 \in E_n(2)$.

We then observe in this case that K_{e+11} can neither be equal to 1, otherwise we have three consecutive blocks of length 2, nor be equal to 2, otherwise $e+9 \in E_n(2)$. This case is then impossible.

(ii) If $K_{e+6} = 2$, then we have

$$(K_e, \dots, K_{e+10}) = (2,1,2,1,1,2,2,1,1,2,1).$$

Indeed, $K_{e+8} = 1$ otherwise $e+6 \in E_n(2)$, and $K_{e+10} = 1$ otherwise we have three consecutive blocks of length 2. Such a run is impossible, since it has three consecutive blocks of length 2.

$\star$ **2**$^{\text{nd}}$ **case** : $K_{e+3} = 2$. Then $K_{e+5} = 1$ otherwise $e+3 \in E_n(2)$. We also have

$$(K_e, \dots, K_{e+10}) = (2,1,2,2,1,1,2,1,1,2,1 \text{ or } 2).$$

Indeed, in view of K_{e+4} and K_{e+5}, we must have $K_{e+6} = 2$, hence $K_{e+7} = 1$ otherwise we have three consecutive blocks of length 2, and $K_{e+8} = 1$ otherwise $e+6 \in E_n(2)$, and then $K_{e+9} = 2$.

(i) If $K_{e+10} = 1$, then K_{e+11} can neither be equal to 1, otherwise we have a run $(1,1,2,1,1,2,1,1)$ which is impossible by Lemma 8 (iii), nor be equal to 2 otherwise $e+9 \in E_n(2)$.

(ii) If $K_{e+10} = 2$, then we have $K_{e+11} = 1$, and K_{e+12} can neither be equal to 1, otherwise we have three consecutive blocks of length 2, nor be equal to 2 otherwise $e+10 \in E_n(2)$.

The proof is complete. $\square$

Lemma 16. *The gaps between two consecutive elements of* $E_n(1) \cup E_n(3)$ *can only be equal to* 1, 3 *or* 5.

Proof. They cannot be equal to 2 for, if e and $e+2$ are consecutive in $E_n(1) \cup E_n(3)$, then we have $e \in E_n(2)$. They cannot be equal to 4 either, for if e and $e+4$ are consecutive $E_n(1) \cup E_n(3)$, then e can neither belong to $E_n(1)$ otherwise $e+1 \in E_n(3)$ and then e and $e+1$ are consecutive in $E_n(1) \cup E_n(3)$, nor belong to $E_n(3)$ otherwise $e+3 \in E_n(1)$ and then e and $e+3$ are consecutive in $E_n(1) \cup E_n(3)$, which is a contradiction. Now it only remains to be shown that gaps between two consecutive elements of $E_n(1) \cup E_n(3)$ are at most equal to 5. Suppose that e and $e+a$ are consecutive in $E_n(1) \cup E_n(3)$ with $a \geq 6$.

- ⋆ If $e \in E_n(1)$, then $K_{e+2} = 1$. Moreover, $K_{e+4} = 1$ otherwise $e+1 \in E_n(3)$. Hence $K_{e+3} = 2$. Similarly, $K_{e+6} = 1$ otherwise $e+3 \in E_n(3)$. Therefore $K_{e+5} = 2$. Hence we have
$$(K_{e+2}, \dots, K_{e+6}) = (1, 2, 1, 2, 1)$$
which is impossible by Lemma 8 (ii).

- ⋆ If $e \in E_n(3)$, then $K_{e+1} = K_{e+2} = 1$ and $K_{e+4} = 1$ otherwise $e+3 \in E_n(1)$. Similarly, $K_{e+6} = 1$ otherwise $e+3 \in E_n(3)$. Hence $K_{e+5} = 2$ and we have
$$(K_{e+2}, \dots, K_{e+6}) = (1, 2, 1, 2, 1)$$
which is impossible by Lemma 8 (ii).

The proof is complete. □

Lemma 17. *We have*
$$\Big||F_n| - |G_n|\Big| \leq \Big||F_n(2)| - |G_n(2)|\Big| + 3.$$

Proof. By Lemma 16, the set $E_n(1) \cup E_n(3)$ is alternatively made up of even and odd numbers, and since 2 is the first element of this set, the difference between the number of even and odd integers of $E_n(1) \cup E_n(3)$ is equal to 0 or 1, i.e.
$$|F_n \cap (E_n(1) \cup E_n(3))| = |G_n \cap (E_n(1) \cup E_n(3))| + 0 \text{ or } 1$$
and therefore
$$\begin{aligned} |F_n| &\leq |F_n \cap E_n(2)| + |F_n \cap (E_n(1) \cup E_n(3))| + 1 \\ &= |F_n(2)| + |F_n \cap (E_n(1) \cup E_n(3))| + 1 \\ &\leq |F_n(2)| + |G_n \cap (E_n(1) \cup E_n(3))| + 2 \\ &\leq |G_n| + |F_n(2)| - |G_n(2)| + 3 \end{aligned}$$
and similarly we have $|G_n| \leq |F_n| + |G_n(2)| - |F_n(2)| + 3$ which completes the proof of Lemma 17. □

Lemma 18. *There do not exist four consecutive numbers with the same parity in $E_n(2)$.*

Proof. 1. If e and $e+6$ are consecutive in $E_n(2)$, then we have
$$(K_e, \dots, K_{e+6}) = \begin{cases} (2, 1, 2, 2, 1, 1, 2) & \text{if } e+2 \in E_n(1); \\ \\ (2, 1, 2, 1, 1, 2, 2) & \text{if } e+2 \in E_n(3). \end{cases} \tag{8}$$
Indeed, suppose $e+2 \in E_n(1)$. Then $(K_e, \dots, K_{e+3}) = (2, 1, 2, 2)$, and hence $K_{e+4} = 1$, and $K_{e+5} = 1$ otherwise either $e+3 \in E_n(2)$ or $e+2 \in E_n(3)$. A similar argument applies if $e+2 \in E_n(3)$.

If e and $e+10$ are consecutive in $E_n(2)$, then if $e+2 \in E_n(1)$, a similar argument as above proves that

$$(K_e, \ldots, K_{e+10}) = (2,1,2,2,1,1,2,1,1,2,2).$$

Besides, we have $e+2 \notin E_n(3)$. Indeed, if $e+2 \in E_n(3)$, then we get

$$(K_e, \ldots, K_{e+10}) = (2,1,2,1,1,2,1,1,2,1,2) \quad \text{or} \quad (2,1,2,1,1,2,2,1,2,1,2)$$

and these two runs are impossible by Lemma 8 (iv).

2. Now let us prove that, if three consecutive numbers have the same parity in $E_n(2)$, then the gap between the first two and between the last two can only be equal to 6.

- ⋆ **1st case :** Suppose that e, $e+6$ and $e+16$ are consecutive in $E_n(2)$. By the arguments above, we have

  $$(K_e, \ldots, K_{e+16}) = (2,1,2,2,1,1,2,1,2,2,1,1,2,1,1,2,2)$$

  or

  $$(K_e, \ldots, K_{e+16}) = (2,1,2,1,1,2,2,1,2,2,1,1,2,1,1,2,2)$$

  and these two runs are impossible by Lemma 8 (v).

- ⋆ **2nd case :** Suppose that e, $e+10$ and $e+16$ are consecutive in $E_n(2)$. By the arguments above, we have

  $$(K_e, \ldots, K_{e+16}) = (2,1,2,2,1,1,2,1,1,2,2,1,2,2,1,1,2)$$

  or

  $$(K_e, \ldots, K_{e+16}) = (2,1,2,2,1,1,2,1,1,2,2,1,2,1,1,2,2)$$

  and these two runs are impossible by Lemma 8 (vi).

- ⋆ **3rd case :** Suppose that e, $e+10$ and $e+20$ are consecutive in $E_n(2)$. By the arguments above, we have

  $$(K_e, \ldots, K_{e+20}) = (2,1,2,2,1,1,2,1,1,2,2,1,2,2,1,1,2,1,1,2,2)$$

  and this run is impossible by Lemma 8 (vii).

Since the numbers have the same parity and they are consecutive in $E_n(2)$, there is no other possibility according to Lemma 15.

3. We are now able to prove Lemma 18. According to the above arguments, if four numbers with the same parity are consecutive in $E_n(2)$, one can only have $e,\ e+6,\ e+12,\ e+18$ consecutive in $E_n(2)$. The (tedious) examination of the eight possibilities induced by (8) shows that each case leads to a run appearing in Lemma 8 (ii), which concludes the proof. □

Remark 19. The result of Lemma 18 is optimal since, for instance, the numbers 75, 81 and 87 are consecutive in $E_n(2)$ for all integers $n \geq 90$ (see Example 13).

In what follows, we introduce the following subsets of $E_n(2)$.

$$A_n := E_n(2,6) \cup E_n(2,10)$$
and
$$B_n := E_n(2,3) \cup E_n(2,5) \cup E_n(2,9).$$

Lemma 20. *We have* $|F_n(2) \cap B_n| = |G_n(2) \cap B_n| + 0$ *or* 1.

Proof. The proof rests on the following assertion.

$$\textit{The gap between two consecutive elements of } B_n \textit{ is odd.} \tag{9}$$

Indeed, if statement (9) is true, then the elements of B_n are alternatively even and odd, and the result follows by noticing that 6 is the first element of this set.

The rest of the text is devoted to the proof of (9). Let $e \in B_n$.

1. If e, $e+a$ and $e+a+b$ are consecutive in $E_n(2)$ with $(a,b) \in \{3,5,9\}^2$, then e and $e+a$ are consecutive in B_n with a odd.

2. If e, $e+a$ and $e+a+b$ are consecutive in $E_n(2)$ with $(a,b) \in \{3,5,9\} \times \{6,10\}$, then $e+a \in A_n$. Let $e+a+b+c$ be the successor of $e+a+b$ in $E_n(2)$.

(i) If $c \in \{3,5,9\}$, then e and $e+a+b$ are consecutive in B_n with $a+b$ odd.

(ii) If $c = 6$, then $e+a+b \in A_n$. Let $e+a+b+6+d$ be the successor of $e+a+b+6$ in $E_n(2)$.

- ⋆ If $d \in \{3,5,9\}$, then e and $e+a+b+6$ are consecutive in B_n with $a+b+6$ odd.
- ⋆ If $d = 6$, then $e+a+b+6 \in A_n$. By Lemmas 15 and 18, the successor of $e+a+b+12$ in $E_n(2)$ can only be the number $e+a+b+12+f$ with $f \in \{3,5,9\}$, so that e and $e+a+b+12$ are consecutive in B_n with $a+b+12$ odd.

(iii) If $c = 10$, then $e+a+b \in A_n$. Let $e+a+b+10+g$ be the successor of $e+a+b+10$ in $E_n(2)$. Again, By Lemmas 15 and 18, we have $g \in \{3,5,9\}$, so that e and $e+a+b+10$ are consecutive in B_n with $a+b+10$ odd.

There is no other possibility by Lemmas 15 and 18, which completes the proof. ☐

Lemma 21. *If* $(e, e+6) \in (E_n(2,6))^2$*, then* $(K_e, \ldots, K_{e+27})$ *has a configuration of the form*

$$\begin{cases} (2,1,2,1,1,2,2,1,2,2,1,1,2,1,2,2,1,2,2,1,1,2,1,1,2,1,2,2), & \textit{if } e+25 \notin E_n; \\ \\ (2,1,2,1,1,2,2,1,2,2,1,1,2,1,2,2,1,2,2,1,1,2,1,1,2,2,1,2), & \textit{if } e+25 \in E_n. \end{cases}$$

Proof.

1. We first notice that, according to (8), if $(e, e+6) \in (E_n(2,6))^2$, one can only have $e+2 \in E_n(3)$ and $e+8 \in E_n(1)$. We infer that

$$(e, e+6) \in (E_n(2,6))^2 \iff (K_e, \ldots, K_{e+14}) = (2,1,2,1,1,2,2,1,2,2,1,1,2,1,2).$$

2. We then examine all the possibilities offered by the sequence (K_n).

   ⋆ Let us show that $e+15 \in E_n(2)$. Indeed, we necessarily have $e+15 \in E_n$, otherwise $(K_{e+11}, \ldots, K_{e+15}) = (1,2,1,2,1)$ which is impossible by Lemma 8 (ii). We deduce that $K_{e+16} = 1$. If $e+15 \notin E_n(2)$, then $K_{e+17} = 1$ and we have

   $$(K_{e+8}, \ldots, K_{e+17}) = (2,2,1,1,2,1,2,2,1,1)$$

   and the lengths give the run $(2,2,1,1,2,2)$, which is impossible by Lemma 8 (ii).

   ⋆ Thus, $e+15 \in E_n(2)$, hence $K_{e+17} = 2$ and $K_{e+19} = 1$ otherwise $e+15$ and $e+17$ are consecutive in $E_n(2)$ which is impossible by Lemma 15.

   ⋆ We also have $e+18 \notin E_n(2)$ otherwise the lengths of the run $(K_{e+13}, \ldots, K_{e+19})$ generate the run $(1,2,1,2,1)$. Similarly, we necessarily have $e+18 \in E_n$ otherwise the lengths of the run $(K_{e+1}, \ldots, K_{e+20})$ give the sequence $(1,1,2,2,1,2,2,1,1,2,1,1,2,2)$ which is impossible.

   ⋆ We infer that $K_{e+21} = 2$ otherwise $(K_{e+19}, \ldots, K_{e+21}) = (1,1,1)$. One can also prove that $e+21 \notin E_n(2)$ otherwise the lengths of the run $(K_{e+8}, \ldots, K_{e+22})$ give $(2,2,1,1,2,1,2,2,1,1)$ which is impossible.

   ⋆ We have $K_{e+22} = 1$ otherwise $(K_{e+17}, \ldots, K_{e+22}) = (2,2,1,1,2,2)$, which is impossible by Lemma 8 (ii).

   ⋆ We have $K_{e+23} = 1$ for $e+21 \notin E_n(2)$. Thus $K_{e+24} = 2$ otherwise $(K_{e+22}, \ldots, K_{e+24}) = (1,1,1)$.

   ⋆ $K_{e+25} = 1$ or 2.

   If $K_{e+25} = 1$, then $K_{e+26} = 2$ otherwise $(K_{e+19} \ldots, K_{e+26}) = (1,1,2,1,1,2,1,1)$. In particular, we have $e+24 \in E_n(2)$. Moreover, we then have $K_{e+27} = 2$ otherwise $(K_{e+23}, \ldots, K_{e+27}) = (1,2,1,2,1)$.

   If $K_{e+25} = 2$, then $K_{e+26} = 1$ otherwise $(K_{e+24}, \ldots, K_{e+26}) = (2,2,2)$, and $K_{e+27} = 2$ otherwise $(K_{e+22}, \ldots, K_{e+27}) = (1,1,2,2,1,1)$, which is impossible by Lemma 8 (ii), so that $e+25 \in E_n(2)$.

The proof is complete. □

Lemma 22. *We have* $|F_n(2) \cap A_n| \leq |F_n(2) \cap B_n|$ *and* $|G_n(2) \cap A_n| \leq |G_n(2) \cap B_n|$.

Proof. By definition of A_n and Lemmas 15 and 18, between two consecutive elements e and e' of A_n, there is always at least an element of B_n having the same parity as e in the following cases

⋆ $e \in E_n(2,10)$, for then $e+10 \in B_n$ by Lemma 18.

⋆ $e \in E_n(2,6)$ and $e+6 \in B_n$.

It remains to be studied the case where e, $e+6$, $e+12$ are consecutive in $E_n(2)$, for then e and $e+6$ are consecutive in $E_n(2,6)$.

In what follows, we consider two pairs $(e, e+6)$ and $(e+a, e+6+a)$ consecutive in $(E_n(2,6))^2$ where a is a positive integer. By Dirichlet's pigeonhole principle, Lemma 22 follows if we show that between these two pairs, there is always at least two elements of B_n having the same parity as e. To prove that, we will use the configurations of Lemma 21.

1. By Lemma 18, we have $e+12 \in B_n$.

2. By Lemma 21, if $e+25 \notin E_n$, then $a \geq 27$

3. Let us show that, if $e+25 \in E_n$, then $a \geq 33$.

If $e+25 \in E_n$, then $K_{e+29} = 1$ otherwise $e+25$ and $e+27$ are consecutive in $E_n(2)$ which is impossible by Lemma 15. If $e+28 \in E_n$, then the lengths of the sequence $(K_{e+14}, \dots, K_{e+28})$ generate the run $(2,1,2,2,1,2,2,1,2)$ whose lengths give the run $(1,2,1,2,1)$ which is impossible by Lemma 8 (ii). Thus, we have $K_{e+28} = 1$ and hence $K_{e+30} = 2$. We notice that $e+30 \notin E_n(2,6)$ otherwise $(K_{e+29}, \dots, K_{e+33}) = (1,2,1,2,1)$. We also have $K_{e+31} = 1$ otherwise the lengths of the sequence $(K_{e+22}, \dots, K_{e+31})$ give the impossible run $(2,2,1,1,2,2)$. Finally, we have $K_{e+32} = 2$ otherwise $K_{e+33} = 2$ and the lengths of the sequence $(K_{e+26}, \dots, K_{e+33})$ then give the impossible run $(1,2,1,2,1)$. But we also have $e+32 \notin E_n(2,6)$ otherwise $(K_{e+31}, \dots, K_{e+35}) = (1,2,1,2,1)$.

4. Now we are in a position to conclude the proof of Lemma 22.

Let $(e, e+6)$ and $(e+a, e+6+a)$ be two consecutive pairs of $(E_n(2,6))^2$. According to the above arguments, we have

- ⋆ If $e+25 \notin E_n$, then $e+12$ and $e+24$ are two elements of B_n comprised between these two pairs.
- ⋆ If $e+25 \in E_n$, then $e+12$ and $e+30$ are two elements of B_n comprised between these two pairs.

The proof is complete. □

Lemma 23. *We have*

$$\Big| |F_n(2)| - |G_n(2)| \Big| \leq \frac{t_n + k_n - n + 7}{3}.$$

Proof. By Lemmas 20 and 22, we have

$$\begin{aligned}
|F_n(2)| &\leq |F_n(2) \cap A_n| + |F_n(2) \cap B_n| + 1 \\
&\leq 2|F_n(2) \cap B_n| + 1 \\
&\leq 2|G_n(2) \cap B_n| + 3 \\
&\leq 2|G_n(2)| + 3,
\end{aligned}$$

and similarly we get $|G_n(2)| \leq 2|F_n(2)| + 3$. The relation $|E_n(2)| = |F_n(2)| + |G_n(2)|$ implies that

$$\frac{|E_n(2)|}{3} - 1 \leq |F_n(2)| \leq \frac{2|E_n(2)|}{3} + 1$$

and similar inequalities for $|G_n(2)|$ so that

$$\Big||F_n(2)| - |G_n(2)|\Big| \leq \frac{|E_n(2)|}{3} + 2$$

and we conclude the proof by using Lemma 14. ☐

Lemmas 17 and 23, used in Proposition 12, give at once the following estimate.

Corollary 24. *We have*

$$\left|S_n - \frac{3n}{2}\right| \leq \frac{t_{k_n} + k_{k_n} - k_n}{6} + 4.$$

We now are able to prove Theorem 4.

Proof of Theorem 4. Using Corollary 24 and the inequalities $t_{k_n} \leq n+1-k_n$ from Lemma 5 and $k_{k_n} \leq 3k_n/4 + 3/2$ from Theorem 3, we get

$$S_n \leq \frac{3n}{2} + \frac{t_{k_n} + k_{k_n} - k_n}{6} + 4 \leq \frac{5n}{3} - \frac{5k_n}{24} + \frac{53}{12}$$

and the lower bound $k_n \geq 3n/5$ from Theorem 3 implies that

$$S_n \leq \frac{37n}{24} + \frac{53}{12},$$

and since $S_{k_n} \geq n$, we infer

$$k_n \geq \frac{24n}{37} - 3.$$

The same method, now using the lower bound

$$S_n \geq \frac{3n}{2} - \frac{t_{k_n} + k_{k_n} - k_n}{6} - 4 \geq \frac{4n}{3} + \frac{5k_n}{24} - \frac{53}{12}$$

and the inequality $S_{k_n} \leq n+1$, leads to

$$k_n \leq \frac{24n}{35} + 4.$$

At the end of this first step, we then have

$$\frac{24n}{37} - 3 \leq k_n \leq \frac{24n}{35} + 4 \quad (n \geq 1). \tag{10}$$

We repeat this process by defining the sequences $(u_m)_{m\geq 1}$ and $(v_m)_{m\geq 1}$ by

$$\begin{cases} u_1 = \dfrac{24}{35} \\ u_{m+1} = \dfrac{6(1-3u_m)}{u_m^2 - 26u_m + 8} \end{cases} \quad \text{and} \quad v_m = \frac{u_m}{3u_m - 1}.$$

It is easy to check that, for all positive integers m, we have $\frac{1}{2} < u_m < 1$ and $\frac{1}{2} < v_m < 1$ so that these two sequences are well-defined. Let us show by induction that, for all positive integers m, we have

$$nv_m - 4 \leq k_n \leq nu_m + 5 \quad (n \geq 1) \tag{11}$$

the case $m = 1$ being a consequence of (10). Assume estimate (11) is true for some m. Then, using induction hypothesis rewritten in the form $k_{k_n} \leq k_n u_m + 5$, we get

$$S_n \leq \frac{3n}{2} + \frac{t_{k_n} + k_{k_n} - k_n}{6} + 4 \leq \frac{5n}{3} - \frac{k_n\,(2 - u_m)}{6} + 5$$

and induction hypothesis used in the form $k_n \geq nv_m - 4$ implies that

$$S_n \leq \frac{n}{6}\,(u_m v_m - 2v_m + 10) - \frac{2u_m - 19}{3} = \frac{n\,(u_m^2 + 28u_m - 10)}{6\,(3u_m - 1)} - \frac{2u_m - 19}{3} = \frac{n}{v_{m+1}} - \frac{2u_m - 19}{3}$$

and the lower bound $S_{k_n} \geq n$ gives

$$k_n \geq nv_{m+1} + \frac{v_{m+1}\,(2u_m - 19)}{3} = nv_{m+1} - \frac{2\,(19 - 2u_m)\,(3u_m - 1)}{u_m^2 + 28u_m - 10} \geq nv_{m+1} - 4.$$

Similarly, the lower bound

$$S_n \geq \frac{3n}{2} - \frac{t_{k_n} + k_{k_n} - k_n}{6} - 4 \geq \frac{4n}{3} + \frac{k_n\,(2 - u_m)}{6} - 5$$

and the inequality $S_{k_n} \leq n + 1$, lead to

$$k_n \leq nu_{m+1} + \frac{2u_{m+1}\,(11 - u_m)}{3} = nu_{m+1} + \frac{4(11 - u_m)(3u_m - 1)}{u_m^2 - 26u_m + 8} \leq nu_{m+1} + 5$$

which completes the proof of (11). We conclude by noticing that the sequence (u_m) converges to α.

Finally, estimates for t_n and o_n follow from relations $t_n = S_n - n$ and $o_n = n - t_n$. The proof is complete. □

7 Concluding remarks

Remark 25. There exist other identites analogue to Proposition 12. For instance, we were able to prove the following results.

(i) *For all positive integers n, we have*

$$k_n = \frac{2}{3}\left(n + \sum_{\substack{j=1 \\ (K_{2j-1}, K_{2j}) = (1,1)}}^{\lfloor k_n/2 \rfloor} 1 - \sum_{\substack{j=1 \\ (K_{2j-1}, K_{2j}) = (2,2)}}^{\lfloor k_n/2 \rfloor} 1 \right) + r_n$$

where $r_1 = 1/3$ *and, for all integers* $n \geq 2$,

$$r_n = \begin{cases} -1/3, & \text{if } (K_{n-1}, K_n, K_{n+1}) = (1,1,2); \\ 0, & \text{if } (K_n, K_{n+1}) = (2,1) \\ 1/3, & \text{if } (K_n, K_{n+1}) = (1,1) \text{ or } (K_{n-1}, K_n, K_{n+1}) = (2,1,2) \\ 2/3, & \text{if } (K_n, K_{n+1}) = (2,2). \end{cases}$$

(ii) *For all positive integers* n, *we have*

$$t_n = \frac{k_n}{2} + \sum_{\substack{j=2 \\ (K_{j-1},K_j)=(2,2)}}^{n} 1 + \frac{K_n}{2} - 1.$$

Although these identities are equivalent to Proposition 12, they seemed to be less handy to get fine estimates of the remainder term.

Remark 26. Like the continuous case and the usual summation formulas (Euler, Euler–MacLaurin, Boole, etc), it might be useful to derive an identity of order two or more to improve on the error-term. In the discrete case, if we treat the sum in (5) by Abel summation, we get

$$\sum_{j=1}^{m} (-1)^j a_j = \frac{(-1)^m a_m - a_1}{2} + \frac{(-1)^m (a_m - a_{m-1})}{4} + \frac{a_2 - a_1}{4} - \frac{1}{4} \sum_{j=2}^{m-1} (-1)^j \Delta^2 a_j$$

where we set $\Delta^2 a_j := a_{j+1} - 2a_j + a_{j-1}$ ($j \geq 2$). Applied to the sequence (t_j) noticing that $\Delta^2 t_j = K_{j+1} - K_j$, this gives

$$\sum_{j=1}^{k_n} (-1)^j t_j = \frac{(-1)^{k_n} t_{k_n}}{2} + \frac{(-1)^{k_n} (K_{k_n} - 1)}{4} + \frac{1}{4} - \frac{1}{4} \sum_{j=2}^{k_n - 1} (-1)^j (K_{j+1} - K_j)$$

for all integers $n \geqslant 2$ and, following the proof of Proposition 12, we infer that

$$S_n = \frac{3n}{2} - \frac{1}{4} \sum_{j=2}^{k_n - 1} (-1)^j (K_{j+1} - K_j) + \frac{(-1)^{k_n}}{4} (K_{k_n} - 2c_n)$$

where c_n is defined in Proposition 12. The trivial estimate of the sum gives

$$\left| \sum_{j=2}^{k_n - 1} (-1)^j (K_{j+1} - K_j) \right| \leq \sum_{j=2}^{k_n - 1} |K_{j+1} - K_j| = \sum_{j=2}^{k_n - 1} (k_{j+1} - k_j) = k_{k_n} - 2$$

where we used Lemma 5. Compared to (6), the saving is not significant.

8 Acknowledgments

We express our gratitude to the referee for his careful reading of the manuscript and the many valuable suggestions he provided.

References

[1] B. C. Berndt and L. Schoenfeld, Periodic analogues of the Euler–MacLaurin and Poisson summation formulas with applications to number theory, *Acta Arith.* **28** (1975), 23–68.

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

[3] A. Carpi, On repeated factors in C^∞-words, *Inf. Process. Lett.* **52** (1994), 289–294.

[4] V. Chvátal, Notes on the Kolakoski sequence, DIMACS Technical report, 1993.

[5] F. M. Dekking, What is the long range order in the Kolakoski sequence?, *The Mathematics of Long–range Aperiodic Order*, NATO Adv. Sci. Inst. Ser. C Maths. Phys. Sci., **489**, Kluwer Acad. Publ., Dordrecht (1997), 115–125.

[6] R. Johnsonbaugh, Summing an alternating series, *Amer. Math. Monthly* **86** (1979), 637–648.

[7] W. Kolakoski, Problem 5304: self generating runs, *Amer. Math. Monthly* **72** (1965), 674.

[8] E. J. Kupin and E. S. Rowland, Bounds on the frequency of 1 in the Kolakoski word, preprint (2008).

[9] K. O'Bryant, B. Reznick, and M. Serbinowska, Almost alternating sums, *Amer. Math. Monthly* **113** (2006), 673–688.

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

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

[12] PARI/GP, available at ftp://megrez.math.u-bordeaux.fr/pub/pari.

2010 *Mathematics Subject Classification*: Primary 11B37; Secondary 11B83, 11B85.
Keywords: Kolakoski sequence, Dirichlet's pigeonhole principle.

(Concerned with sequence A000002.)

Received November 7 2010; revised version received January 28 2011. Published in *Journal of Integer Sequences*, February 9 2011.

Return to Journal of Integer Sequences home page.