



On a Recursively Defined Sequence Involving the Prime Counting Function

Altug Alkan
Graduate School of Science and Engineering
Piri Reis University
Istanbul
Turkey

altug.alkan@pru.edu.tr

Andrew R. Booker
School of Mathematics
University of Bristol
Woodland Road
Bristol, BS8 1UG
United Kingdom

andrew.booker@bristol.ac.uk

Florian Luca
School of Mathematics
University of the Witwatersrand
Private Bag X3, Wits 2050
Johannesburg
South Africa

Research Group in Algebraic Structures and Applications
King Abdulaziz University
Jeddah
Saudi Arabia

Centro de Ciencias Matemáticas
UNAM
Morelia
Mexico

Florian.Luca@wits.ac.za

Abstract

We prove some properties of sequence [A335294](#) from the *On-Line Encyclopedia of Integer Sequences*, defined by

$$a_n = \pi(n) - \pi\left(\sum_{k=1}^{n-1} a_k\right),$$

where $\pi(x)$ is the number of primes $\leq x$. In particular we show that the sequence (a_n) assumes every non-negative integral value infinitely often.

1 Introduction

Let $\pi(x) = \#\{p \text{ prime} : p \leq x\}$ denote the prime counting function. In this paper we consider the sequence $(a_n)_{n \geq 1}$ defined by

$$a_n = \pi(n) - \pi\left(\sum_{k=1}^{n-1} a_k\right) \quad \text{for } n \geq 1.$$

Apart from a difference in the first three terms, this is sequence [A335294](#) in the OEIS,¹ and its initial terms are

0, 1, 2, 0, 1, 1, 1, 1, 0, 0, 1, 1, 2, 1, 1, 0, 1, 1, 2, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 2, 2, 1,
 1, 0, 0, 1, 1, 1, 1, 2, 1, 2, 2, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 0, 0,
 1, 1, 1, 1, 2, 1, 2, 2, 1, 0, 0, 0, 1, 1, 1, 1, 2, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, . . .

A brief examination reveals that the sequence is not monotonic and displays a remarkably slow rate of growth. In this direction, see Table 1, which shows the smallest solutions to $a_n = k$ for each $k \leq 13$. (This is a subsequence of the prime numbers for $k \geq 1$; note that for $k \in \{3, 4, 5, 6, 9, 10, 13\}$ the corresponding n is also the larger of a twin prime pair.)

| | | | | | | | |
|-----|---------|-----------|------------|-------------|--------------|----------------|----------------|
| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| n | 1 | 2 | 3 | 229 | 3259 | 15739 | 449569 |
| k | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| n | 6958841 | 130259903 | 2404517671 | 56014949761 | 538155413969 | 21692297487587 | 21692297487589 |

Table 1: Smallest n satisfying $a_n = k$, cf. [A335337](#)

Let s_n denote the summatory sequence,

$$s_n = \sum_{k=1}^n a_k \quad \text{for } n \geq 0.$$

¹Sequence [A335294](#) imposes the initial condition $a_1 = 1$, whereas we adopt the convention that the empty sum is 0, so that $a_1 = \pi(1) - \pi(0) = 0$. We prefer this convention since it is more elegant and simplifies some proofs. As we show in Section 3, the sequence is eventually stable under any perturbation of the initial terms, so this is not a significant alteration.

Our main result establishes some distributional properties of $(a_n)_{n \geq 1}$ and $(s_n)_{n \geq 0}$. In order to state them, we define $g(x)$ to be the maximum distance between a real number $y \leq x$ and the largest prime $p \leq y$, i.e.,

$$g(x) = \sup_{y \in [2, x]} \min\{y - p : p \leq y\} \quad \text{for } x \geq 2.$$

Note that g is a continuous, piecewise linear, non-decreasing function, and

$$\pi(n) - \pi(n - g(n) - 1) \geq 1 \quad \text{for all integers } n \geq 2.$$

Conjecturally one has $g(x) = O(\log^2 x)$; the best result to date, due to Baker, Harman, and Pintz [2], is that $g(x) \leq x^{21/40}$ for all sufficiently large x .

Theorem 1. *The following conclusions hold:*

- (i) $a_n \geq 0$ and $a_n - \max\{1, 2\pi(a_n)\} \leq a_{n+1} \leq a_n + 1$ for all $n \geq 1$;
- (ii) $a_n = O(\sqrt{g(n)/\log g(n)})$ for all $n \geq 5$;
- (iii) for each $k \geq 0$, there are infinitely many n such that $a_n = k$;
- (iv) $n - g(n) \leq s_n \leq n - 2$ for all $n \geq 9$;
- (v) $s_n = n - g(n)$ for infinitely many n .

Proof. We begin with the upper estimate in (iv). Suppose that $s_n \leq n$ holds for some $n \geq 0$; note that this is the case for $n = 0$. By definition we have

$$s_{n+1} = s_n + a_{n+1} = s_n + \pi(n+1) - \pi(s_n), \tag{1}$$

so that

$$n+1 - s_{n+1} = (n+1 - s_n) - (\pi(n+1) - \pi(s_n)). \tag{2}$$

Since $s_n \leq n$, the right-hand side is non-negative, and in fact it counts the number of non-prime integers in the interval $(s_n, n+1]$. Hence we have $s_{n+1} \leq n+1$. By induction it follows that $s_n \leq n$ for all $n \geq 0$.

Next we improve this to $s_n \leq n - 2$. Suppose $n \geq 9$ is such that $s_{n+i} \leq n+i-2$ for $i \in \{0, 1, 2, 3\}$; we verify this directly for $n = 9$. If $s_{n+4} \geq n+3$ then we have

$$\begin{aligned} n+4 - s_{n+4} \leq 1 &\implies (s_{n+3}, n+4] \text{ contains at most one composite number} \\ &\implies n+2 \text{ and } n+4 \text{ are prime} \\ &\implies n+3, n+1, n \text{ and } n-1 \text{ are composite} \\ &\implies s_{n+3} \geq n+1 \implies n+3 - s_{n+3} \leq 2 \\ &\implies (s_{n+2}, n+3] \text{ contains at most two composite numbers} \end{aligned}$$

$$\begin{aligned}
&\implies s_{n+2} \geq n \implies n + 2 - s_{n+2} \leq 2 \\
&\implies (s_{n+1}, n + 2] \text{ contains at most two composite numbers} \\
&\implies s_{n+1} \geq n - 1 \implies n + 1 - s_{n+1} \leq 2 \\
&\implies (s_n, n + 1] \text{ contains at most two composite numbers} \\
&\implies s_n \geq n - 1.
\end{aligned}$$

This contradicts the assumption that $s_n \leq n - 2$, so we must have $s_{n+4} \leq n + 2$. By induction it follows that $s_n \leq n - 2$ for all $n \geq 9$.

Next, for all $n \geq 1$ we have

$$a_n = \pi(n) - \pi(s_{n-1}) \geq \pi(n) - \pi(n - 1) \geq 0. \quad (3)$$

It follows that s_n is non-decreasing, and thus

$$a_{n+1} - a_n = (\pi(n + 1) - \pi(n)) - (\pi(s_n) - \pi(s_{n-1})) \leq \pi(n + 1) - \pi(n) \leq 1. \quad (4)$$

Moreover, by Montgomery and Vaughan's upper estimate [8, Corollary 2] for primes in short intervals, we have

$$a_{n+1} \geq a_n - (\pi(s_{n-1} + a_n) - \pi(s_{n-1})) \geq a_n - \max\{1, 2\pi(a_n)\}.$$

This proves (i). We also note that the lower estimate can be improved to $a_{n+1} \geq a_n - \pi(a_n)$ for $n \geq 4$ and $2 \leq a_n \leq 1731$, by work of Gordon and Rodemich [6].

Let n be a natural number satisfying

$$s_n \geq n - g(n). \quad (5)$$

Note that this holds for $n = 3$. If $s_n \geq n + 1 - g(n + 1)$ then

$$s_{n+1} \geq s_n \geq n + 1 - g(n + 1).$$

Otherwise we have $n - g(n) \leq s_n \leq n - g(n + 1)$, so that

$$s_{n+1} = s_n + \pi(n + 1) - \pi(s_n) \geq n - g(n) + \pi(n + 1) - \pi(n - g(n + 1)).$$

By the definition of g we have $\pi(n + 1) - \pi(n - g(n + 1)) \geq 1$, so

$$s_{n+1} \geq n + 1 - g(n) \geq n + 1 - g(n + 1).$$

Thus, in either case, (5) holds with n replaced by $n + 1$. By induction, (5) holds for all $n \geq 3$, and this completes the proof of (iv).

Turning to (ii), let n be a natural number, and suppose that $k = a_n \geq 3$. Applying (4) inductively, we see that

$$a_{n-i} \geq k - (\pi(n) - \pi(n - i)) \quad \text{for all } i < n. \quad (6)$$

Let $h \geq 2$ be the largest integer such that $\pi(h) \leq k/3$. Taking $i = n - 1$ in (6) we see that $\pi(n) \geq k \geq 3\pi(h) > \pi(h)$, whence $h < n$. Moreover, by the prime number theorem we have $h \asymp k \log k$. By [8, Corollary 2], for any non-negative integer $i \leq h$ we have

$$\pi(n) - \pi(n - i) \leq \max\{1, 2\pi(i)\} \leq 2\pi(h) \leq 2k/3,$$

so that $a_{n-i} \geq k/3$. Therefore,

$$s_n - s_{n-h} = \sum_{i=0}^{h-1} a_{n-i} \geq \frac{hk}{3} \gg k^2 \log k.$$

By (iv), $s_n - s_{n-h} = h + O(g(n)) \ll k \log k + g(n)$. Thus, $k^2 \log k \ll k \log k + g(n)$, and (ii) follows.

Next, let p, q be a pair of consecutive odd primes attaining a maximal prime gap. Define sequences $(s'_n)_{n \geq p}$ and $(d_n)_{n \geq p}$ by

$$s'_p = p, \quad s'_{n+1} = s'_n + \pi(n+1) - \pi(s'_n) \quad \text{for } n \geq p, \quad (7)$$

and

$$d_n = s'_n - s_n \quad \text{for } n \geq p.$$

Subtracting (7) and (1), we find that

$$d_{n+1} = d_n - (\pi(s_n + d_n) - \pi(s_n)) \quad \text{for } n \geq p.$$

It follows that $0 \leq d_{n+1} \leq d_n$, so that $s_n \leq s'_n$ for all $n \geq p$. From (7) we see that $s'_{p+1} = s'_{p+2} = \dots = s'_{q-1} = p$. Hence

$$q - 1 - s_{q-1} \geq q - 1 - s'_{q-1} = q - 1 - p = g(q - 1),$$

and by (iv) it follows that $s_{q-1} = q - 1 - g(q - 1)$. This proves (v).

(Moreover we have $s_{q-1} = p$, $s_q = p + 1$, $s_{q+1} = p + 2$, and if $p > 3$ then

$$\max\{s_n : s_n < p\} = p - 1;$$

thus $s_n + h$ is infinitely often prime for each $h \in \{-2, -1, 0, 1\}$. By the estimate of Ford et al. on large prime gaps [4], we also see that $n - s_n \gg \frac{\log n \log \log n \log \log \log n}{\log \log \log n}$ infinitely often.)

Next we prove (iii). First note that if there were only finitely many n with $a_n = 0$ then we would have $s_n \geq n - O(1)$, contradicting (v); hence (iii) is true for $k = 0$. Since a_n can increase by at most 1 at each step and there are infinitely many n with $a_n = 0$, to complete the proof of (iii) it suffices to show that $(a_n)_{n \geq 1}$ is unbounded.

To that end, for a given integer $m \geq 2$ we apply the main result of Banks, Freiberg, and Turnage-Butterbaugh [3] to find a sequence of $m + 1$ consecutive primes with a large gap followed by a relatively dense cluster. Precisely, let $k = k_{m+1}$ in the notation of [3], and set $b_j = -m^{k+1-j}$ for $j = 1, \dots, k$. Then it is easy to see that the polynomial $\prod_{j=1}^k (x + b_j)$ has no fixed prime divisor, so by [3, Theorem 1] there exists a subset $\{h_0, \dots, h_m\} \subseteq \{b_1, \dots, b_k\}$ such that $x + h_0, \dots, x + h_m$ are consecutive primes for infinitely many $x \in \mathbb{Z}$.

Fix any such x , denote the corresponding primes by p_0, \dots, p_m , and write $h_i = -m^{k+1-j_i}$, where $1 \leq j_0 < \dots < j_m \leq k$. Then

$$\begin{aligned} (m-1)(h_m - h_1) &= (m-1)(m^{k+1-j_1} - m^{k+1-j_m}) \\ &< m^{k+2-j_1} - m^{k+1-j_1} \leq m^{k+1-j_0} - m^{k+1-j_1} = h_1 - h_0, \end{aligned}$$

so that

$$p_1 - p_0 > (m-1)(p_m - p_1) \geq (m-1)p_m - (p_1 + p_2 + \dots + p_{m-1}).$$

As in the proof of (v) above, we define sequences $(s'_n)_{n \geq p_1}$, $(a'_n)_{n > p_1}$, and $(d_n)_{n \geq p_1}$ by

$$\begin{aligned} s'_{p_1} &= p_1, & s'_{n+1} &= s'_n + \pi(n+1) - \pi(s'_n) \quad \text{for } n \geq p_1, \\ a'_n &= s'_n - s'_{n-1} \quad \text{for } n > p_1, \end{aligned}$$

and

$$d_n = s'_n - s_n \quad \text{for } n \geq p_1.$$

As above we find that $0 \leq d_{n+1} \leq d_n$, so

$$a'_n = a_n + d_n - d_{n-1} \leq a_n \quad \text{for } n > p_1.$$

A straightforward inductive argument now shows that

$$\begin{aligned} s'_n &= p_0 && \text{for } p_0 \leq n < p_1, \\ s'_n &= p_0 + n - p_1 + 1 && \text{for } p_1 \leq n < p_2, \\ s'_n &= p_0 + (p_2 - p_1) + 2(n - p_2 + 1) && \text{for } p_2 \leq n < p_3, \\ &\vdots && \\ s'_n &= p_0 + (p_2 - p_1) + 2(p_3 - p_2) + \dots \\ &\quad + (m-2)(p_{m-1} - p_{m-2}) + (m-1)(n - p_{m-1} + 1) && \text{for } p_{m-1} \leq n < p_m. \end{aligned}$$

In particular,

$$s'_{p_{m-1}} = p_0 + (m-1)p_m - (p_1 + p_2 + \dots + p_{m-1}) < p_1,$$

so that

$$a_{p_m} \geq a'_{p_m} = \pi(p_m) - \pi(s'_{p_{m-1}}) = m.$$

Since m was arbitrary, this completes the proof of (iii). \square

2 Some conjectures

It follows from (ii) and (iv) that

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{s_n}{n} = 1. \quad (8)$$

We further conjecture the following.

Conjecture 2.

(A) For any $k \geq 0$, the set $\{n \geq 1 : a_n = k\}$ has a positive density δ_k , satisfying

$$\delta_1 > \delta_0 > \delta_2 > \delta_3 > \delta_4 > \dots$$

(B) $\liminf_{n \rightarrow \infty} (n - s_n) < \infty$.

(C) For any integer $b \geq 2$, the number $A(b) = \sum_{n \geq 1} a_n b^{-n} \in \mathbb{R}$ is transcendental.

Part (A) of the conjecture is based on the numerical data shown in Table 2; note that the convergence is very slow, so it is difficult to measure the densities precisely, and in any case we do not expect them to be recognizable constants. In connection with (B), it seems likely from numerical computations that $s_n = n - 2$ infinitely often; by (iv) this would imply that $\liminf_{n \rightarrow \infty} (n - s_n) = 2$. If $(a_n)_{n \geq 1}$ were an automatic sequence then (C) would follow from the main result of Adamczewski, Bugeaud, and Luca [1].

| | | k | | | | | |
|-----|----------------|----------------|----------------|---------------|--------------|-------------|--|
| i | 0 | 1 | 2 | 3 | 4 | 5 | |
| 1 | 4 | 5 | 1 | 0 | 0 | 0 | |
| 2 | 21 | 65 | 14 | 0 | 0 | 0 | |
| 3 | 219 | 577 | 195 | 9 | 0 | 0 | |
| 4 | 2663 | 4990 | 2065 | 275 | 7 | 0 | |
| 5 | 27671 | 48507 | 20265 | 3287 | 257 | 13 | |
| 6 | 284408 | 475421 | 199765 | 36779 | 3443 | 181 | |
| 7 | 2918543 | 4650175 | 1991476 | 395418 | 41464 | 2800 | |
| 8 | 29607905 | 45960839 | 19809319 | 4108991 | 473258 | 37723 | |
| 9 | 299530722 | 455176760 | 197289962 | 42282008 | 5235205 | 456865 | |
| 10 | 3022594978 | 4517557589 | 1965289965 | 432413509 | 56484650 | 5291355 | |
| 11 | 30450733004 | 44894741076 | 19590459294 | 4400511075 | 599692839 | 59396517 | |
| 12 | 306392386246 | 446604857931 | 195374867235 | 44626996156 | 6295691446 | 652786704 | |
| 13 | 3080065196771 | 4446030725007 | 1949223822125 | 451486351994 | 65543491929 | 7053078276 | |
| 14 | 30940285500711 | 44287714979733 | 19452797930000 | 4559198048883 | 678055064108 | 75277782875 | |

Table 2: Values of $\#\{n \leq 10^i : a_n = k\}$ for $0 \leq k \leq 5$ and $1 \leq i \leq 14$.

In addition to the numerical evidence noted above, we provide the following theoretical evidence in support of the conjectures.

Theorem 3.

- (i) $\#\{n \geq 1 : a_n = k\}$ has positive upper density for at least one $k \in \{0, 1\}$.
- (ii) We have $\#\{n \leq x : a_n = 0\} \gg \frac{x \log \log x}{\log^2 x}$ for all $x \geq 3$ under the hypothesis that $g(x) \ll (\log x)^C$ for some $C > 1$, and $\#\{n \leq x : a_n = 0\} \geq \exp((\log x)^{\frac{1}{4}-o(1)})$ unconditionally.
- (iii) $\liminf_{n \rightarrow \infty} \frac{n-s_n}{\log n} \leq 1$.
- (iv) The number $A(b)$ is irrational.

2.1 Proof of (i) and (ii)

We begin with an overview of the proof. First we show that if a_n has few zero terms then there must exist many long intervals I such that $a_n = 1$ for all $n \in I$. This immediately yields (i), and in turn implies that $c := n - s_n$ is constant for $n+1 \in I$. In view of (4), for any n such that $n, n+1 \in I$ and $n+1$ is prime, s_n must also be prime. Thus, the constellation of primes occurring in I is repeated on the shifted interval $\{n - c : n, n+1 \in I\}$. Now the idea is to use results from the Selberg sieve (in particular [5, Lemma 5.1]) that bound the frequency of such coincidences of the constellations, and this results in a contradiction if a_n has too few zeros. There is a tradeoff between increasing the dimension of the sieve, which reduces the overall frequency asymptotically, and the combinatorial price paid for that increase. Optimizing the dimension then yields (ii).

Next we set some notation to be used in the proof. Let $x > 0$ be a large real number, and let $r, T \in \mathbb{Z}$ be parameters, to be specified in due course, satisfying

$$2 \leq r \leq T \leq r(\log x)^{\frac{r-1}{4r+2}} / \log \log x.$$

We regard r as fixed throughout the proof, so the meaning of \ll, O, o , “sufficiently large”, etc. may depend implicitly on r . Let $K \geq 1$ be a large (absolute) constant, and define

$$H_j = Kj^2(\log x)(\log T) \quad \text{for } 0 \leq j \leq T.$$

Next, set $N = \lfloor x \rfloor$ and

$$N_k = \#\{1 \leq n \leq x : a_n = k\} \quad \text{for } k \geq 0.$$

Then

$$N_0 + N_1 + N_2 + \cdots = N \quad \text{and} \quad N_1 + 2N_2 + \cdots = s_N,$$

so that

$$N_0 - (N_2 + 2N_3 + \cdots) = N - s_N > 0.$$

Thus

$$N_0 + (N_2 + N_3 + \cdots) \leq N_0 + (N_2 + 2N_3 + \cdots) < 2N_0.$$

This also shows that $N_1 + 2N_0 > N$, so that $\max\{N_0, N_1\} > \frac{1}{3}N$. It follows that at least one of the sets $\{n \geq 1 : a_n = k\}$ for $k \in \{0, 1\}$ has upper density $\geq \frac{1}{3}$. This proves (i).

Next, setting $\mathcal{N} = \{1 \leq n \leq x : a_n \neq 1\}$, we have $\#\mathcal{N} < 2N_0$. Let

$$\mathcal{L} = \{\ell \in \mathbb{Z} : 1 \leq \ell \leq x \text{ and } a_n \neq 1 \text{ for some } n \in [\ell - 1, \ell + H_T] \cap \mathbb{Z}_{>0}\}.$$

By [8, Corollary 2], the number of primes contained in \mathcal{L} is at most

$$2\pi(H_T + 2)(\#\mathcal{N} + 1) \leq 4\pi(H_T + 2)N_0.$$

Suppose, for the sake of contradiction, that $4\pi(H_T + 2)N_0 \leq \frac{1}{2}\pi(x)$. From now on we consider primes $p \in [1, x] \setminus \mathcal{L}$, which is at least half of the primes $p \leq x$. These primes have the property that $a_n = 1$ for all integers n satisfying $p - 1 \leq n \leq p + H_T$.

We need two easy facts about primes.

Lemma 4. *Let p_i denote the i th prime. For a suitable choice of the constant K and all sufficiently large x , there are at most $\frac{1}{4}\pi(x)$ primes $p_i \leq x$ for which there exists $j \in \{1, \dots, T\}$ satisfying $p_{i+j} - p_i > H_j$.*

Proof. Fix $j \in \{1, \dots, T\}$. Then, since $j \leq T = o(\pi(x))$, we have

$$\sum_{i=1}^{\pi(x)} (p_{i+j} - p_i) < \sum_{k=\pi(x)+1}^{\pi(x)+j} p_k = (1 + o(1))jx,$$

by the prime number theorem. Therefore, the number of i such that $p_{i+j} - p_i > H_j$ is $\ll jx/H_j = x/(jH_1)$. Summing this over all $j \leq T$, we get a bound of

$$\ll \frac{x}{H_1} \sum_{j \leq T} \frac{1}{j} \ll \frac{x \log T}{H_1} \ll \frac{\pi(x)}{K}.$$

For K sufficiently large this is less than $\frac{1}{4}\pi(x)$. □

Lemma 5. *For a fixed choice of $r \geq 2$, there are at most $o(\pi(x))$ primes $p_i \leq x$ satisfying the following conditions:*

(i) $p_{i+j} - p_i \leq H_j$ for all $j \in \{0, \dots, T\}$;

(ii) there are vectors $(j_1, \dots, j_r), (j'_1, \dots, j'_r) \in \mathbb{Z}^r$ such that

$$0 \leq j_1 < j_2 < \dots < j_r \leq T, \quad 0 \leq j'_1 < j'_2 < \dots < j'_r \leq T,$$

and

$$p_{i+j_1} - p_{i+j'_1} = \dots = p_{i+j_r} - p_{i+j'_r} \neq 0.$$

Proof. Let p_i be such a prime, and set

$$h_j = p_{i+j} - p_i \quad \text{for } j \in \{0, \dots, T\}.$$

Let $(j_1, \dots, j_r), (j'_1, \dots, j'_r)$ be as in (ii), and write

$$\{j_1, \dots, j_r\} \cup \{j'_1, \dots, j'_r\} = \{\ell_1, \dots, \ell_k\},$$

with $\ell_1 < \dots < \ell_k$. From our hypotheses it is clear that $r + 1 \leq k \leq 2r$. Let

$$d = h_{j_1} - h_{j'_1} = \dots = h_{j_r} - h_{j'_r}$$

denote the common difference. Swapping (j_1, \dots, j_r) and (j'_1, \dots, j'_r) if necessary, we may assume without loss of generality that $d > 0$, and it follows that $j_s > j'_s$ for each $s \in \{1, \dots, r\}$.

For a fixed value of k , there are $O(T^k)$ ways of choosing $\{j_1, \dots, j_r\}$ and $\{j'_1, \dots, j'_r\}$ of total cardinality k . If $k = 2r$ then for each choice of indices, there are at most H_T^{r+1} choices for the pair of vectors $v = (h_{j_1}, \dots, h_{j_r}), v' = (h_{j'_1}, \dots, h_{j'_r})$, since v' is determined by v and d . If $k < 2r$ then there are $2r - k$ pairs (s, t) such that $j_s = j'_t$; for each pair we have $d = h_{j_t} - h_{j'_t} = h_{j_t} - h_{j_s}$, so that h_{j_t} is determined by h_{j_s} and d . Hence, in general there are at most H_T^{k+1-r} choices for v, v' for a given choice of indices. Thus, in total we find

$$\ll T^k H_T^{k+1-r} \ll T^{3k+2-2r} ((\log T)(\log x))^{k+1-r}$$

choices for v, v' for our fixed k .

Let us first suppose that $\ell_1 > 0$. Then $n = p_i$ is an integer such that the $k + 1$ distinct linear forms $n, n + h_{\ell_1}, \dots, n + h_{\ell_k}$ are all prime. By [9, Ch. II, Satz 4.2], the number of such $n \leq x$ is

$$\ll_k \frac{x}{(\log x)^{k+1}} \left(\frac{E}{\varphi(E)} \right)^k, \quad \text{where } E = \prod_{1 \leq s \leq k} h_{\ell_s} \cdot \prod_{1 \leq s < t \leq k} (h_{\ell_t} - h_{\ell_s}).$$

Since $h_{\ell_1}, \dots, h_{\ell_k} \leq H_T \ll \log^2 x$, we have $\frac{E}{\varphi(E)} \ll \log \log \log x$. Hence, the number of possibilities for p_i is

$$\begin{aligned} &\ll \frac{T^{3k+2-2r} (\log T)^{k+1-r} x (\log \log \log x)^k}{(\log x)^r} \leq \frac{T^{4r+2} (\log T)^{r+1} x (\log \log \log x)^{2r}}{(\log x)^r} \\ &\ll \frac{x (\log \log \log x)^{2r}}{(\log x) (\log \log x)^{3r+1}} = o(\pi(x)). \end{aligned}$$

If $\ell_1 = 0$ then we lose one linear form, but gain from the fact that j'_1 and $h_{j'_1}$ are fixed at 0. This effectively replaces k by $k - 1$ in the above analysis, so we again find $o(\pi(x))$ possibilities for p_i . Finally, summing over $k \in \{r + 1, \dots, 2r\}$ concludes the proof of the lemma. \square

The following is Lemma 5.1 of Ford, Konyagin, and Luca [5].

Lemma 6. *There is a positive constant δ so that the following holds. Let a_1, \dots, a_k be positive integers, let b_1, \dots, b_k be integers and let $\xi(p)$ be the number of solutions of*

$$\prod_{i=1}^k (a_i n + b_i) \equiv 0 \pmod{p}.$$

If $x \geq 10$, $1 \leq k \leq \delta \frac{\log x}{\log \log x}$, and

$$B := \sum_p \left(\frac{k - \xi(p)}{p} \right) \log p \leq \delta \log x,$$

then the number of integers $n \leq x$ for which $a_1 n + b_1, \dots, a_k n + b_k$ are all prime and $> k$ is

$$\ll \frac{2^k k! \mathfrak{S} x}{(\log x)^k} \exp \left(O \left(\frac{kB + k^2 \log \log x}{\log x} \right) \right), \quad \text{where } \mathfrak{S} = \prod_p \left(1 - \frac{\xi(p)}{p} \right) \left(1 - \frac{1}{p} \right)^{-k}. \quad (9)$$

We are now ready to go. As we said, we work with primes $p_i \leq x$ that are not in \mathcal{L} . The number of them is at least $\frac{1}{2}\pi(x)$. We discard all p_i such that $p_{i+j} - p_i > H_j$ holds for some $j = 1, \dots, T$. By Lemma 4, there are at most $\frac{1}{4}\pi(x)$ such primes. Next, applying Lemma 5, by removing a further $o(\pi(x))$ values of p_i , we may assume that as j and j' range over $\{0, \dots, T\}$, each non-zero difference $p_{i+j} - p_{i+j'}$ occurs with multiplicity at most $r - 1$. After this we are left with at least $(\frac{1}{4} - o(1))\pi(x)$ primes p_i .

Set $c = p_i - s_{p_i}$. By Theorem 1(iv) and the definition of N_0 , we have

$$0 < c \leq M := \min\{g(x), N_0\}.$$

Now consider $p_i, p_{i+1}, \dots, p_{i+T}$. These are of the form $p_{i+j} = p_i + h_j$ for some $h_j \leq H_j$, as in the proof of Lemma 5. On the other hand, $p_{i+T} - p_i \leq H_T$, and since $a_n = 1$ for $p_i - 1 \leq n \leq p_i + H_T$, we have $s_n = n - c$ for $p_i - 2 \leq n \leq p_{i+T}$. Applying (4) with $n = p_{i+j} - 1$, we have

$$\begin{aligned} 0 &= a_{p_{i+j}} - a_{p_{i+j}-1} = (\pi(p_{i+j}) - \pi(p_{i+j}-1)) - (\pi(s_{p_{i+j}-1}) - \pi(s_{p_{i+j}-2})) \\ &= 1 - (\pi(p_{i+j}-1-c) - \pi(p_{i+j}-2-c)). \end{aligned}$$

Hence, $p_{i+j} - 1 - c = p_i + h_j - c - 1$ is prime.

Therefore, $n = p_i$ is such that $n + h_j$ and $n + h_j - c - 1$ are all primes for $j = 0, \dots, T$. This is $2T + 2$ linear forms, but they might not all be distinct. Let m be the cardinality of the intersection

$$\{h_j : 0 \leq j \leq T\} \cap \{h_j - c - 1 : 0 \leq j \leq T\}.$$

Then there exist $j_0 < j_1 < \cdots < j_m$ and $j'_0 < j'_1 < \cdots < j'_m$ with

$$h_{j_1} - h_{j'_1} = \cdots = h_{j_m} - h_{j'_m} = c + 1,$$

so that

$$p_{i+j_1} - p_{i+j'_1} = \cdots = p_{i+j_m} - p_{i+j'_m} > 0.$$

By our construction, we must have $m < r$; in particular, there are at least $2T + 3 - r$ distinct forms among the $n + h_j$ and $n + h_j - c - 1$ for $j = 0, \dots, T$. Hence we may apply Lemma 6 for some $k \in [2T + 3 - r, 2T + 2] \cap \mathbb{Z}$.

We need to check the hypothesis on B and estimate some of the parameters in (9). For B , we partition the primes into $S_1 \cup S_2 \cup S_3$, where

$$S_1 = \{p : p \leq \log^2 x \text{ or } p \mid (c + 1)\}, \quad S_2 = \{p : \xi(p) < k\} \setminus S_1, \quad S_3 = \{p : \xi(p) = k\} \setminus S_1.$$

Since $c \leq x$, $c + 1$ has $O(\log x / \log \log x)$ prime factors exceeding $\log^2 x$. Hence,

$$\begin{aligned} \sum_{p \in S_1} \left(\frac{k - \xi(p)}{p} \right) \log p &\leq k \sum_{p \leq \log^2 x} \frac{\log p}{p} + k \sum_{\substack{p \mid (c+1) \\ p > \log^2 x}} \frac{\log p}{p} \\ &\ll T \log \log x + \frac{T}{\log x} \ll T \log \log x. \end{aligned}$$

For any prime $p \in S_2$, there is a double solution n modulo p to

$$\prod_{0 \leq j \leq T} (n + h_j) \cdot \prod_{\substack{0 \leq j \leq T \\ h_j - c - 1 \notin \{h_0, \dots, h_T\}}} (n + h_j - c - 1) \equiv 0 \pmod{p}.$$

If the double root comes from the forms $n + h_j$ for $j = 0, \dots, T$, we get that p divides $h_{j_2} - h_{j_1}$ for some j_1, j_2 with $0 \leq j_1 < j_2 \leq T$. But this is impossible since $p > \log^2 x$ and $h_j \leq H_T < \log^2 x$ for large x . The same argument shows that the double solution cannot come from two factors of the form $n + h_j - c - 1$ for $j \in \{0, \dots, T\}$. So any double root must appear once from the first set of forms and once from the second, so that p divides $c + 1 + h_{j'} - h_j \neq 0$ for some $j, j' \in \{0, \dots, T\}$. These numbers all lie in the interval $[c + 1 - H_T, c + 1 + H_T]$, and since $c \leq x$, each has $O(\log x / \log \log x)$ prime factors exceeding $\log^2 x$. Thus,

$$\#S_2 \ll \frac{H_T \log x}{\log \log x} \ll \log^3 x.$$

Moreover, writing $m = k - \xi(p)$, there exist $j_1 < \cdots < j_m, j'_1 < \cdots < j'_m$ such that

$$h_{j_1} - h_{j'_1} \equiv \cdots \equiv h_{j_m} - h_{j'_m} \equiv c + 1 \pmod{p}.$$

Since $p \nmid (c + 1)$ and $2H_T + 1 < \log^2 x$ for large x , this implies that

$$h_{j_1} - h_{j'_1} = \cdots = h_{j_m} - h_{j'_m} \neq 0.$$

Thus we have $m < r$, so that $\xi(p) \geq k + 1 - r$. Therefore

$$\sum_{p \in S_2} \left(\frac{k - \xi(p)}{p} \right) \log p \leq (r - 1) \sum_{\log^2 x < p \leq O(\log^3 x)} \frac{\log p}{p} \ll \log \log x.$$

Finally, the primes in S_3 do not contribute to B . Thus, the bound on B holds, and in fact $B = O(T \log \log x)$.

We now estimate (9). Since $B = O(T \log \log x)$, the factor involving \exp tends to 1 as $x \rightarrow \infty$, so it is smaller than 2 for large x . In the expression for \mathfrak{S} , the primes $p \in S_1$ contribute at most

$$\begin{aligned} \left(\frac{c+1}{\varphi(c+1)} \right)^k \prod_{p \leq \log^2 x} \left(1 - \frac{1}{p} \right)^{-k} &= O(\log \log x)^k \exp \left(k \sum_{p \leq \log^2 x} \frac{O(1)}{p} \right) \\ &= \exp(O(T \log \log \log x)). \end{aligned}$$

The contribution from $p \in S_2$ is at most

$$\begin{aligned} \prod_{p \in S_2} \left(1 - \frac{k+1-r}{p} \right) \left(1 - \frac{1}{p} \right)^{-k} &= \prod_{p \in S_2} \left(1 - \frac{k+1-r}{p} \right) \left(1 + \frac{k}{p} + O\left(\frac{k^2}{p^2}\right) \right) \\ &= \prod_{p \in S_2} \left(1 + O\left(\frac{1}{p}\right) \right) = \exp \left(\sum_{p \in S_2} \frac{O(1)}{p} \right) = e^{O(1)}. \end{aligned}$$

Similarly, from $p \in S_3$ we get a contribution of

$$\prod_{p \in S_3} \left(1 - \frac{k}{p} \right) \left(1 - \frac{1}{p} \right)^k = \exp \left(k \sum_{p > \log^2 x} \frac{O(1)}{p^2} \right) = \exp \left(O\left(\frac{k}{\log x}\right) \right) = e^{O(1)}.$$

Thus, in total we have

$$\mathfrak{S} = \exp(O(T \log \log \log x)).$$

Applying Lemma 6, the number of $n \leq x$ of this form is

$$\ll \frac{2^k k! x}{(\log x)^k} \exp(O(T \log \log \log x)).$$

Since $2k \leq 4T + 4 < \log x$ for large x , this is largest when $k = 2T + 3 - r$. Using also that

$$2^{2T+3-r} (2T+3-r)! = T^{2T+8-4r} e^{O(r \log T) + O(T)} = T^{2T+8-4r} e^{O(T)},$$

we obtain

$$\ll \frac{T^{2T+8-4r} \pi(x)}{(\log x)^{2T+2-r}} \exp(O(T \log \log \log x)).$$

This is for fixed c, h_1, \dots, h_T . The number of choices for these parameters is at most

$$MH_1 \cdots H_T = M(T!)^2 H_1^T \leq MT^{2T} (\log x)^T \exp(O(T \log \log \log x)).$$

Thus, in total the number of possibilities is

$$\ll \frac{M\pi(x)}{\exp\left((T+2-r)\log\left(\frac{\log x}{T^4}\right)\right)} \exp(O(T \log \log \log x)).$$

This must account for at least $(\frac{1}{4} - o(1))\pi(x)$ primes, so for sufficiently large x we have

$$M = \min\{g(x), N_0\} \gg \exp\left((T+2-r)\log\left(\frac{\log x}{T^4}\right) - O(T \log \log \log x)\right).$$

If $g(x) \ll (\log x)^C$ for some $C > 1$, then taking $r = 2$ and $T = \lfloor C \rfloor + 1$ results in a contradiction for sufficiently large x . Hence, our hypothesis that $4\pi(H_T + 2)N_0 \leq \frac{1}{2}\pi(x)$ must be false, and it follows that $N_0 \gg x(\log \log x)/\log^2 x$.

On the other hand, assuming that $N_0 \ll x/\log^2 x$ (and making no hypothesis on $g(x)$), we can take $T = \lfloor r(\log x)^{\frac{r-1}{4r+2}}/\log \log x \rfloor$, and we conclude that $N_0 \geq \exp((\log x)^{\frac{r-1}{4r+2}})$ for all sufficiently large x . Since this is true for every $r \geq 2$, we have $N_0 \geq \exp((\log x)^{\frac{1}{4}-o(1)})$.

2.2 Proof of (iii)

Consider positive integers $M < N$, and let $h = \min\{n - s_n : M \leq n < N\}$. Then

$$\begin{aligned} s_N - s_M &= \sum_{n=M}^{N-1} a_{n+1} = \sum_{n=M}^{N-1} (\pi(n+1) - \pi(s_n)) \geq \sum_{n=M}^{N-1} (\pi(n+1) - \pi(n-h)) \\ &= \sum_{i=0}^h (\pi(N-i) - \pi(M-i)) \geq (h+1)(\pi(N-h) - \pi(M)). \end{aligned}$$

By Theorem 1(iv) and the result of Baker, Harman, and Pintz [2], we have

$$h \leq g(M) \leq M^{21/40} \quad \text{for sufficiently large } M.$$

Choosing $N = \lceil M + M^{7/12} \rceil$ and applying Heath-Brown's asymptotic [7] for primes in short intervals, we have

$$\pi(N-h) - \pi(M) = (1+o(1)) \frac{N-M}{\log M} \quad \text{as } M \rightarrow \infty.$$

On the other hand,

$$s_N - s_M \leq N - (M - g(M)) \leq (1+o(1))(N-M),$$

so that $h \leq (1+o(1))\log M$. Thus, every sufficiently large interval $[M, M + M^{7/12})$ contains an integer n with $n - s_n \leq (1+o(1))\log n$.

2.3 Proof of (iv)

Let N be a large natural number, and write $\{1, \dots, N\}$ as a disjoint union $I_1 \cup \dots \cup I_J$ of intervals I_j such that a_n is constant on each I_j and J is as small as possible. Setting $m_j = \max I_j$ for $j \leq J$, we have either $m_j = N$ or $a_{m_j+1} \neq a_{m_j}$. From (4) we see that if $a_{n+1} \neq a_n$ then either $\pi(n+1) \neq \pi(n)$ or $\pi(s_n) \neq \pi(s_{n-1})$. Since both sequences $\pi(n)$ and $\pi(s_{n-1})$ are non-decreasing and $s_n \leq n$ for all n , it follows that

$$J \leq 1 + \#\{n < N : a_{n+1} \neq a_n\} \leq 1 + 2\pi(N).$$

Thus, for at least one of the intervals, say $I_j = \{n_1, \dots, n_2\}$, we have

$$\#I_j = n_2 - n_1 + 1 \geq \frac{N}{1 + 2\pi(N)}.$$

By the prime number theorem, for any fixed $\varepsilon > 0$ this exceeds $(\frac{1}{2} - \varepsilon) \log N$ for all sufficiently large N .

Suppose $A(b) = u/v$ is rational. Then, multiplying by $v(b-1)b^{n_1-1}$, we obtain

$$u(b-1)b^{n_1-1} = v(b-1)b^{n_1-1} \sum_{n=1}^{\infty} \frac{a_n}{b^n}.$$

Let $c = a_{n_1}$. Since a_n is constant for $n_1 \leq n \leq n_2$, we have

$$\begin{aligned} u(b-1)b^{n_1-1} &= v(b-1) \sum_{n=1}^{n_1-1} a_n b^{n_1-1-n} + v(b-1)c \sum_{n=n_1}^{n_2} b^{n_1-1-n} + v(b-1) \sum_{n \geq n_2+1} a_n b^{n_1-1-n} \\ &= v(b-1) \sum_{n=1}^{n_1-1} a_n b^{n_1-1-n} + vc(1 - b^{n_1-n_2-1}) + v(b-1)b^{n_1-n_2-1} \sum_{m \geq 1} \frac{a_{n_2+m}}{b^m}. \end{aligned}$$

Hence,

$$vb^{n_1-n_2-1} \left((b-1) \sum_{m \geq 1} \frac{a_{n_2+m}}{b^m} - c \right) = u(b-1)b^{n_1-1} - v(b-1) \sum_{n=1}^{n_1-1} a_n b^{n_1-1-n} - vc \quad (10)$$

is an integer.

On the other hand, since $0 \leq a_{n_2+m} \leq c + m$ for $m \geq 1$, we have

$$-c \leq (b-1) \sum_{m \geq 1} \frac{a_{n_2+m}}{b^m} - c \leq (b-1) \sum_{m \geq 1} \frac{c+m}{b^m} - c = \frac{b}{b-1}.$$

Hence the left-hand side of (10) is bounded in modulus by

$$\frac{(c+2)v}{b^{n_2-n_1+1}} \leq \frac{(c+2)v}{N^{(\frac{1}{2}-\varepsilon)\log b}} \ll \frac{v\sqrt{g(N)}}{N^{(\frac{1}{2}-\varepsilon)\log b}}.$$

From Baker, Harman, and Pintz [2], we have $g(N) \leq N^{21/40}$ for sufficiently large N . Since $\log b \geq \log 2 > \frac{21}{40}$, for small enough ε this expression tends to 0 as $N \rightarrow \infty$. Since it has to be an integer, it must be 0 for all sufficiently large N .

Therefore,

$$\sum_{m \geq 1} \frac{a_{n_2+m}}{b^m} = \frac{c}{b-1} = \sum_{m \geq 1} \frac{c}{b^m}.$$

By Theorem 1(iii) there exists $n_3 > n_2$ such that $a_{n_3} = 0$. Thus, we have

$$\sum_{j=1}^{\infty} \frac{a_{n_3+j}}{b^{n_3+j}} = \sum_{m \geq 1} \frac{c}{b^m} - \sum_{m=1}^{n_3-n_2-1} \frac{a_{n_2+m}}{b^m} = \sum_{m=1}^{n_3-n_2-1} \frac{c - a_{n_2+m}}{b^m} + \frac{cb^{n_2-n_3+1}}{b-1}.$$

Multiplying both sides by $(b-1)b^{n_3-1}$, we see that the right-hand side is an integer, so

$$(b-1) \sum_{j \geq 1} \frac{a_{n_3+j}}{b^{j+1}} \in \mathbb{Z}.$$

On the other hand, we have $0 \leq a_{n_3+j} \leq j$, and by Theorem 1(iii) both inequalities are strict for infinitely many j . Hence,

$$0 < (b-1) \sum_{j \geq 1} \frac{a_{n_3+j}}{b^{j+1}} < (b-1) \sum_{j \geq 1} \frac{j}{b^{j+1}} = 1.$$

This is a contradiction, so $A(b)$ must be irrational.

3 Generalizations and suggestions for further work

The sequence a_n admits a vast generalization via sequences of the form

$$a_f(n) = \pi(f(n)) - \pi\left(\sum_{k=1}^{n-1} a_f(k)\right)$$

for various functions f . For instance, choosing $f(n) = tn$ for a fixed integer $t > 0$, our proof of (8) can be generalized to show that

$$\lim_{n \rightarrow \infty} \frac{a_f(n)}{n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{s_f(n)}{n} = t,$$

where $s_f(n) = \sum_{k=1}^n a_f(k)$ denotes the summatory function. One can pose many of the same questions and conjectures for these sequences.

Another possible generalization is to consider the same recurrence formula with different initial conditions. However, it turns out that this offers no increase in generality, in the sense that if $(a'_n)_{n \geq 1}$ is any sequence satisfying

$$a'_n = \pi(n) - \pi\left(\sum_{k=1}^{n-1} a'_k\right) \quad \text{for } n > n_0$$

for some $n_0 \geq 0$, then $a'_n = a_n$ for all sufficiently large n . (The same proof shows that taking $f(n) = n + c$ for some $c \in \mathbb{Z}$ in the above, we have $a_f(n) = a_{n+c}$ for sufficiently large n .) To see this, let s'_n be the summatory sequence of a'_n , and set $d_n = s'_n - s_n$. Then, as in the proof of Theorem 1(v), we find that $|d_{n+1}| \leq |d_n|$. From that proof we also see that each of the sequences $(s_n)_{n \geq 1}$ and $(s_n + 1)_{n \geq 1}$ assumes infinitely many prime values, and it follows that d_n eventually reaches 0. Thus $s'_n = s_n$ and $a'_n = a_n$ for sufficiently large n .

At the same time, there are several possible avenues for further research on (a_n) . We conclude with a few speculative suggestions.

1. Assuming Cramér's conjecture, by Theorem 1(ii) we have

$$\#\{n \leq x : a_n \neq 0\} \geq \frac{\sum_{n \leq x} a_n}{\max_{n \leq x} a_n} \gg \frac{x \sqrt{\log \log x}}{\log x}.$$

This could be improved with some information on higher moment statistics of a_n . For instance, can one give a non-trivial upper bound for $\sum_{n \leq x} a_n^2$?

2. It is easy to see that the difference sequence $a_{n+1} - a_n$ is almost always 0, so a_n has many long constant runs. (This idea was used in the proof of Theorem 3(iv).) Assuming either Conjecture 2(A) or Dickson's conjecture, one can see that for any $k \geq 0$ there are arbitrarily long runs of n with $a_n = k$. Unconditionally, by Theorem 3(i) this holds for at least one $k \in \{0, 1\}$, and from the proof of Theorem 1(iii) we get arbitrarily long runs on which a_n is both constant and arbitrarily large. Can one give an unconditional proof of long constant runs for a specific value of k ?
3. The previous question admits many generalizations. For instance, assuming Dickson's conjecture, one can see that there are arbitrarily long arithmetic progressions $n, n + d, \dots, n + kd$ such that $a_{n+jd} = j$ for $j = 0, \dots, k$. Can this be proved unconditionally?

4 Acknowledgments

Altug Alkan would like to thank Robert Israel, Remy Sigrist and Giovanni Resta for their valuable computational assistance regarding OEIS contributions [A335294](#) and [A335337](#).

References

- [1] B. Adamczewski, Y. Bugeaud, and F. Luca, Sur la complexité des nombres algébriques, *C. R. Math. Acad. Sci. Paris* **339** (2004), 11–14.
- [2] R. C. Baker, G. Harman, and J. Pintz, The difference between consecutive primes. II, *Proc. London Math. Soc.* **83** (2001), 532–562.

- [3] W. D. Banks, T. Freiberg, and C. L. Turnage-Butterbaugh, Consecutive primes in tuples, *Acta Arith.* **167** (2015), 261–266.
- [4] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, Long gaps between primes, *J. Amer. Math. Soc.* **31** (2018), 65–105.
- [5] K. Ford, S. V. Konyagin, and F. Luca, Prime chains and Pratt trees, *Geom. Funct. Anal.* **20** (2010), 1231–1258.
- [6] D. M. Gordon and G. Rodemich, Dense admissible sets. In *Algorithmic number theory (Portland, OR, 1998)*, Vol. 1423 of *Lecture Notes in Comput. Sci.*, pp. 216–225. Springer, 1998.
- [7] D. R. Heath-Brown, The number of primes in a short interval, *J. Reine Angew. Math.* **389** (1988), 22–63.
- [8] H. L. Montgomery and R. C. Vaughan, The large sieve, *Mathematika* **20** (1973), 119–134.
- [9] K. Prachar, *Primzahlverteilung*, Springer-Verlag, 1957.

2020 *Mathematics Subject Classification*: Primary 11A41; Secondary 11B83.

Keywords: prime-counting function.

(Concerned with sequences [A335294](#) and [A335337](#).)

Received June 12 2020; revised version received January 27 2021. Published in *Journal of Integer Sequences*, January 31 2021.

Return to [Journal of Integer Sequences home page](#).