Some properties of a sequence defined with the aid of prime numbers

Brăduţ Apostol, Laurenţiu Panaitopol, Lucian Petrescu, and László Tóth

Journal of Integer Sequences **18** (2015), Article 15.5.5

Abstract

For every integer $n \geq 1$ let a_n be the smallest positive integer such that $n + a_n$ is prime. We investigate the behavior of the sequence $(a_n)_{n\geq 1}$, and prove asymptotic results for the sums $\sum_{n\leq x} a_n$, $\sum_{n\leq x} 1/a_n$ and $\sum_{n\leq x} \log a_n$.

2010 Mathematics Subject Classification: 11A41, 11N05

Key Words and Phrases: prime numbers, difference of consecutive primes, asymptotic behavior

1 Introduction

For every integer $n \geq 1$ let a_n be the smallest positive integer such that $n+a_n$ is prime. Here $a_1 = 1$, $a_2 = 1$, $a_3 = 2$, $a_4 = 1$, $a_5 = 2$, $a_6 = 1$, $a_7 = 4$, etc. This is sequence A013632 in Sloane's Online Encyclopedia of Integer Sequences [4]. For $n \geq 2$, a_n is the smallest positive integer such that $\gcd(n!, n+a_n) = 1$. In this paper we study the behavior of the sequence $(a_n)_{n\geq 1}$, and prove asymptotic results for the sums $\sum_{n\leq x} a_n$, $\sum_{n\leq x} 1/a_n$ and $\sum_{n\leq x} \log a_n$.

We are going to use the following standard notation:

- $\pi(x)$ is the number of primes $\leq x$,
- $\pi_2(x)$ is the number of twin primes $p, p+2$ such that $p \leq x$,
- p_n is the n-th prime,
- $d_n = p_{n+1} - p_n$,
- $f(x) \ll g(x)$ means that $|f(x)| \leq Cg(x)$, where C is an absolute constant,
- $g(x) \gg f(x)$ means that $f(x) \ll g(x)$,
- $f(x) = F(x) + O(g(x))$ means that $f(x) - F(x) \ll g(x)$,
- $f(x) \asymp g(x)$ means that $cf(x) \leq g(x) \leq Cf(x)$ for some positive absolute constants c and C,
- $f(x) \sim g(x)$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$.

We will apply the following known asymptotic results concerning the distribution of the primes:

$$\pi(x) \sim \frac{x}{\log x}, \quad p_n \sim n \log n \quad \text{(Prime number theorem)},$$

$$\sum_{p_n \leq x} d_n^2 \ll x^{23/18+\varepsilon} \quad \text{for every } \varepsilon > 0 \text{ (unconditional result of Heath-Brown [1])}, \tag{1}$$

$$\sum_{p_n \le x} d_n^2 \ll x(\log x)^3 \quad \text{(assuming the Riemann hypothesis, result of Selberg [3])}, \tag{2}$$

$$\left(\frac{d_2 d_3 \cdots d_n}{(\log 2)(\log 3)\cdots(\log n)}\right)^{1/n} \asymp 1 \quad \text{(due to Panaitopol [2, Prop. 3])}. \tag{3}$$

This research was initiated by Laurenţiu Panaitopol (1940–2008), former professor at the Faculty of Mathematics, University of Bucharest, Romania. The present paper is dedicated to his memory.

2 Equations and identities

By the definition of a_n, for every $n \ge 1$ we have $n + a_n = p_{\pi(n)+1}$, that is

$$a_n = p_{\pi(n)+1} - n. \tag{4}$$

From (4) we deduce that for every $k \ge 1$,

$$a_{p_k} = p_{k+1} - p_k, a_{p_k+1} = p_{k+1} - p_k - 1, \ldots, a_{p_{k+1}-1} = 1. \tag{5}$$

Proposition 1. *For every integer $a \ge 1$ the equation $a_n = a$ has infinitely many solutions.*

Proof. Let $A_k = \{1, 2, \ldots, p_{k+1} - p_k\}$. Since $\limsup_{k\to\infty}(p_{k+1} - p_k) = \infty$, it follows from (5) that for every integer $a \ge 1$ there exist infinitely many integers $k \ge 1$ such that $a \in A_k$, whence the equation $a_n = a$ has infinitely many solutions. $\square$

Now we compute the sum $S_n = \sum_{i=1}^{n} a_i$.

Proposition 2. *For every prime $n \ge 3$ we have*

$$S_n = \frac{1}{2}\left(2p_{\pi(n)+1} - p_{\pi(n)} + \sum_{k=1}^{\pi(n)-1} d_k^2\right), \tag{6}$$

and for every composite number $n \ge 4$,

$$S_n = \frac{1}{2}\left(p_{\pi(n)}^2 + 2(n + 1 - p_{\pi(n)})p_{\pi(n)+1} + \sum_{k=1}^{\pi(n)-1} d_k^2 - n^2 - n\right). \tag{7}$$

Proof. If $n \geq 3$ is a prime, then $n = p_m$ for some $m \geq 2$. By using (4),

$$\begin{aligned}
S_n &= \sum_{i=1}^{n} \left(p_{\pi(i)+1} - i\right) \\
&= 2 + 3 + (5+5) + \cdots + (p_m - p_{m-1})p_m + p_{m+1} - \frac{n(n+1)}{2} \\
&= 2 + \sum_{k=2}^{m} p_k(p_k - p_{k-1}) + p_{m+1} - \frac{n(n+1)}{2} \\
&= \frac{1}{2}\left(p_1^2 + 2\sum_{k=2}^{m} p_k^2 - 2\sum_{k=2}^{m} p_k p_{k-1} + 2p_{m+1} - n^2 - n\right) \\
&= \frac{1}{2}\left(2p_{m+1} - n + \sum_{k=1}^{m-1} (p_{k+1} - p_k)^2\right)
\end{aligned}$$

and (6) follows by using that $m = \pi(n)$.

Now let $t \geq 4$ be composite. Let $m \geq 2$ be such that $p_m < t < p_{m+1}$. By applying (6) for $n = p_m$, where $m = \pi(n) = \pi(t)$, we deduce

$$\begin{aligned}
S_t &= S_n + \sum_{i=n+1}^{t} a_i = S_n + \sum_{i=n+1}^{t} \left(p_{\pi(i)+1} - i\right) \\
&= \frac{1}{2}\left(2p_{\pi(t)+1} - p_{\pi(t)} + \sum_{k=1}^{\pi(t)-1} (p_{k+1} - p_k)^2\right) + \frac{(2p_{\pi(t)+1} - n - t - 1)(t-n)}{2} \\
&= \frac{1}{2}\left(2p_{\pi(t)+1} - p_{\pi(t)} + \sum_{k=1}^{\pi(t)-1} (p_{k+1} - p_k)^2 + 2p_{\pi(t)+1}(t-n) - t^2 - t + n^2 + n\right) \\
&= \frac{1}{2}\left(p_{\pi(t)}^2 + 2(t + 1 - p_{\pi(t)})p_{\pi(t)+1} + \sum_{k=1}^{\pi(t)-1} (p_{k+1} - p_k)^2 - t^2 - t\right)
\end{aligned}$$

and (7) is proved. □

Remark 3. *If n is prime, then (7) reduces to (6). Therefore, the identity (7) holds for every integer $n \geq 3$.*

Next we compute the product $P_n = \prod_{i=1}^{n} a_i$.

Proposition 4. *For every prime $n \geq 3$ we have*

$$P_{n-1} = \prod_{k=1}^{\pi(n)-1} d_k!, \tag{8}$$

and for every composite number $n \geq 4$,

$$P_{n-1} = \prod_{k=1}^{\pi(n)-1} d_k! \prod_{k=1}^{n-p_{\pi(n)}} (p_{\pi(n)+1} - p_{\pi(n)} - k + 1). \tag{9}$$

Proof. Let $n = p_m \geq 3$ be a prime. By using (5),

$$P_{n-1} = \prod_{i=2}^{m}(p_i - p_{i-1})! = \prod_{i=1}^{m-1}(p_{i+1} - p_i)!,$$

which proves (8).

Now let $t \geq 4$ be composite such that $p_m < t < p_{m+1}$. By applying (8) for $n = p_m$, where $m = \pi(n) = \pi(t)$, we deduce

$$\begin{aligned} P_{t-1} &= P_{n-1}\prod_{i=n}^{t-1} a_i = P_{n-1}\prod_{i=n}^{t-1}\left(p_{\pi(i)+1} - i\right) \\ &= \prod_{k=1}^{\pi(t)-1} d_k! \prod_{j=1}^{t-p_m}(p_{m+1} - p_m - j + 1) \\ &= \prod_{k=1}^{\pi(t)-1} d_k! \prod_{k=1}^{t-p_{\pi(t)}}\left(p_{\pi(t)+1} - p_{\pi(t)} - k + 1\right) \end{aligned}$$

and (9) is proved. □

Remark 5. *If n is prime, then the second product in (9) is empty and (9) reduces to (8). Hence the identity (9) holds for every integer $n \geq 3$.*

3 Asymptotic results

Theorem 6. *For every $\varepsilon > 0$,*

$$x\log x \ll \sum_{n\leq x} a_n \ll x^{23/18+\varepsilon}, \tag{10}$$

where $23/18 \doteq 1.277$. If the Riemann hypothesis is true, then the upper bound in (10) is $x(\log x)^3$.

Proof. Let $x \geq 2$ and let $p_k \leq x < p_{k+1}$. By using (6) for $n = p_{k+1}$,

$$\begin{aligned} \sum_{n\leq x} a_n \leq \sum_{i=1}^{p_{k+1}} a_i &= \frac{1}{2}\left(2p_{k+2} - p_{k+1} + \sum_{i=1}^{k} d_i^2\right) \\ &\ll p_{k+2} + \sum_{p_i \leq x} d_i^2. \end{aligned}$$

Taking into account the estimate (1) due to Heath-Brown, and the fact that $p_{k+2} \sim p_k \leq x$ we get the unconditional upper bound in (10). If the Riemann hypothesis is true, then by using Selberg's result (2) we obtain the upper bound $x(\log x)^3$.

Now, for the lower bound we use the trivial estimate

$$\sum_{p_n\leq x} d_n^2 \gg x\log x,$$

which follows from the inequality between the arithmetic and quadratic means. We deduce that

$$\sum_{n\le x} a_n \ge \sum_{i=1}^{p_k} a_i = \frac{1}{2}\left(2p_{k+1} - p_k + \sum_{i=1}^{k-1} d_i^2\right)$$
$$\gg \sum_{p_i \le p_{k-1}} d_i^2 \gg p_{k-1}\log p_{k-1} \sim x\log x,$$

since $p_{k-1} \sim k\log k$ and $k = \pi(x) \sim x/\log x$, $\log k \sim \log x$. □

To prove our next result we need the following

Lemma 7. *We have*

$$\sum_{2\le n\le x} \log d_n = x\log\log x + O(x). \tag{11}$$

Proof. The inequalities (3) can be written as

$$cn < \sum_{i=2}^{n} \log d_i - \sum_{i=2}^{n} \log\log i < Cn$$

for some positive absolute constants c and C. Now (11) emerges by applying the well known asymptotic formula

$$\sum_{2\le n\le x} \log\log n = x\log\log x + O(x).$$

□

Theorem 8. *We have*

$$\sum_{n\le x} \frac{1}{a_n} = \frac{x\log\log x}{\log x} + O\left(\frac{x}{\log x}\right). \tag{12}$$

Proof. For $x = p_m - 1$ $(m \ge 2)$ we have by (5),

$$\sum_{n\le p_m - 1} \frac{1}{a_n} = 1 + \sum_{i=2}^{m}\left(1 + \frac{1}{2} + \cdots + \frac{1}{p_i - p_{i-1}}\right).$$

For an arbitrary $x \ge 3$ let p_k $(k \ge 2)$ be the prime such that $p_k \le x < p_{k+1}$. Using the familiar inequalities

$$\log m < 1 + \frac{1}{2} + \cdots + \frac{1}{m} \le 1 + \log m \quad (m \ge 1)$$

we deduce

$$\log(p_i - p_{i-1}) < 1 + \frac{1}{2} + \cdots + \frac{1}{p_i - p_{i-1}} \le 1 + \log(p_i - p_{i-1}) \quad (i \ge 2)$$

and

$$1 + \sum_{i=2}^{k} \log(p_i - p_{i-1}) + \frac{1}{d_k} < \sum_{n\le p_k - 1} \frac{1}{a_n} + \frac{1}{a_{p_k}}$$

$$\leq \sum_{n\leq x} \frac{1}{a_n} \leq \sum_{n\leq p_{k+1}-1} \frac{1}{a_n} \leq 1 + k + \sum_{i=2}^{k+1} \log(p_i - p_{i-1}).$$

By (11) we obtain

$$\sum_{n\leq x} \frac{1}{a_n} = k \log\log k + O(k),$$

Here $k = \pi(x) \sim x/\log x$, $\log k \sim \log x$ and we deduce (12). □

Theorem 9. *One has*

$$x \ll \sum_{n\leq x} \log a_n \ll x \log x.$$

Proof. For an arbitrary $x \geq 3$ let p_k $(k \geq 2)$ be the prime such that $p_k \leq x < p_{k+1}$. Using the elementary inequalities

$$m \log m - m + 1 \leq \log m! \leq m \log m \quad (m \geq 1)$$

we deduce by applying (8) that

$$\sum_{n\leq x} \log a_n \leq \sum_{n\leq p_{k+1}-1} \log a_n = \sum_{i=1}^{k} \log d_i! \leq \sum_{i=1}^{k} d_i \log d_i$$

$$< \sum_{i=1}^{k} d_i \log p_i < (\log p_k) \sum_{i=1}^{k} d_i < (\log p_k) p_{k+1},$$

where we also used that $d_i = p_{i+1} - p_i < p_i$ by Chebyshev's theorem. Here

$$p_k \sim k \log k, \quad k = \pi(x) \sim x/\log x, \quad \log k \sim \log x, \tag{13}$$

and we obtain the upper bound $x \log x$.

On the other hand,

$$\sum_{n\leq x} \log a_n > \sum_{n\leq p_k - 1} \log a_n = \sum_{i=1}^{k-1} \log d_i!$$

$$> \sum_{i=1}^{k-1} (d_i \log d_i - d_i + 1) = \sum_{i=2}^{k-1} d_i \log d_i - p_k + k + 1.$$

Here

$$\begin{aligned}
\sum_{i=2}^{k-1} d_i \log d_i &= \sum_{\substack{i=2\\ d_i\ge 3}}^{k-1} d_i \log d_i + 2\log 2 \sum_{\substack{i=2\\ d_i=2}}^{k-1} 1 \\
&\ge (\log 3) \sum_{\substack{i=2\\ d_i\ge 3}}^{k-1} d_i + (2\log 2)\pi_2(k-1) \\
&= (\log 3) \left(\sum_{i=2}^{k-1} d_i - \sum_{\substack{i=2\\ d_i=2}}^{k-1} d_i \right) + (2\log 2)\pi_2(k-1) \\
&= (\log 3)\,(p_k - p_2 - 2\pi_2(k-1)) + (2\log 2)\pi_2(k-1) \\
&= (\log 3)p_k - 2\log(3/2)\pi_2(k-1) - 3\log 3 \\
&> (\log 3)p_k - 2\log(3/2)k - 3\log 3,
\end{aligned}$$

where it is sufficient to use the obvious estimate $\pi_2(k-1) < k$. Note that $\log 3 \doteq 1.09$, $2\log(3/2) \doteq 0.81$, $3\log 3 \doteq 3.29$.

We deduce that

$$\sum_{n\le x} \log a_n > 0.09 p_k - 3.$$

Now, (13) gives the lower bound x. □

References

[1] D. R. Heath–Brown, The differences between consecutive primes, III, *J. London Math. Soc.* (2) **20** (1979), 177–178.

[2] L. Panaitopol, Properties of the series of differences of prime numbers, *Publ. Centre de Rech. Math. Pures Neuchâtel (Serie 1)* **31** (2000), 21–28.

[3] A. Selberg, On the normal density of primes in small intervals, and the difference between consecutive primes, *Arch. Math. Naturvid.* **47** (1943), No. 6, 87–105.

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

B. Apostol
Pedagogic High School ”Spiru Haret”, Str. Timotei Cipariu 5, RO–620004 Focşani, Romania
E-mail: apo_brad@yahoo.com

L. Petrescu
Technical College ”Henri Coandă”, Str. Tineretului 2, RO–820235 Tilcea, Romania
E-mail: petrescuandreea@yahoo.com

L. Tóth
Department of Mathematics, University of Pécs
Ifjúság útja 6, H–7624 Pécs, Hungary
E-mail: ltoth@gamma.ttk.pte.hu