

SEARCHING FOR CONSECUTIVE INTEGERS DIVISIBLE BY A POWER OF THEIR LARGEST PRIME FACTOR

Jean-Marie De Koninck (Québec, Canada)

Matthieu Moineau (Paris, France)

Communicated by Péter Burcsi

(Received 13 February 2026; accepted 6 May 2026)

Abstract. Given an integer $n \geq 2$, let $P(n)$ stand for its largest prime factor. Given integers $k \geq 2$ and $\ell \geq 2$, consider the set $E_{k,\ell}$ of those integers $n \geq 2$ for which $P(n+i)^\ell \mid n+i$ for $i = 0, 1, \dots, k-1$. These sets are very thin. For instance, the smallest element of $E_{3,2}$ is 1 294 298, whereas the smallest known element of $E_{3,3}$ has 77 digits. The study of the sets $E_{k,\ell}$ originated in 2009 and was later pursued by others. Here, we present a survey of the results obtained so far and provide new ones. In the process, given an arbitrary integer $\ell \geq 2$ and setting $G_\ell := \{n \in \mathbb{N} : P(n)^\ell \mid n\}$, we introduce two polynomial-time algorithms each providing the list of all elements of G_ℓ below a given bound. We conclude by raising various open problems.

1. Introduction

In a book of the first author [2] published in 2009, thousands of numbers with interesting arithmetic properties are listed. One of these “fascinating numbers” is the integer 1 294 298, which has the particular property of being the smallest positive integer n for which each of the numbers $n, n+1, n+2$ is divisible by the square of its largest prime factor. Indeed, one can check that

$$\begin{aligned} 1\,294\,298 &= 2 \cdot 61 \cdot 103^2, \\ 1\,294\,299 &= 3^4 \cdot 19 \cdot 29^2, \\ 1\,294\,300 &= 2^2 \cdot 5^2 \cdot 7 \cdot 43^2. \end{aligned}$$

Key words and phrases: Factorization, polynomials, friable numbers, Dickman’s function, Mason’s theorem.

2010 Mathematics Subject Classification: 11A05, 11A51, 11C08, 11N25, 11N32, 11Y16.

<https://doi.org/10.71352/ac.59.060526>

The next six integers with this same property are

$$9\,841\,094, 158\,385\,500, 1\,947\,793\,550, 5\,833\,093\,013, 11\,587\,121\,710, \\ 20\,944\,167\,840.$$

How many more are there? What about three consecutive integers each divisible by the cube of their largest prime factor? Or four consecutive? This motivates the introduction of the following family of sets. Given fixed integers $k \geq 2$ and $\ell \geq 2$, let

$$E_{k,\ell} := \{n \in \mathbb{N} : P(n+i)^\ell \mid n+i \text{ for each } i = 0, 1, \dots, k-1\},$$

where $P(m)$ stands for the largest prime factor of m , so that in particular $1\,294\,298 \in E_{3,2}$.

Of course, given an arbitrary integer $\ell \geq 2$ and distinct primes p_0, \dots, p_{k-1} with $k \geq 2$, the Chinese remainder theorem guarantees the existence of infinitely many integers n such that $p_i^\ell \mid n+i$ for $i = 0, 1, \dots, k-1$. However, this theorem does not guarantee that for such integers n we have $P(n+i) = p_i$ for $i = 0, 1, \dots, k-1$, although this is the case for the number $1\,294\,298$ when $\ell = 2$ and $p_0 = 103, p_1 = 29$ and $p_2 = 43$. As we will see, each set $E_{k,\ell}$ is very small. Nevertheless, it is natural to conjecture that, given any $k \geq 2$ and $\ell \geq 2$, the corresponding set $E_{k,\ell}$ is infinite. But how does one prove this? Here, we show that the set $E_{2,2}$ is infinite and study its size. We also examine other sets $E_{k,\ell}$ each time building strategies to find their elements. We also provide a heuristic argument whose goal is to determine the size of the (so far unknown) smallest elements of $E_{4,2}$ and $E_{5,2}$. In the process, given an arbitrary integer $\ell \geq 2$ and setting $G_\ell := \{n \in \mathbb{N} : P(n)^\ell \mid n\}$, we create two distinct algorithms which each provide the list of all elements of G_ℓ below a given bound, and both these algorithms run in polynomial time. To study the sets $E_{k,\ell}$, we rely on elementary, analytic and probabilistic number theory tools, as well as basic results from algebra.

We conclude our presentation by raising various open problems.

2. Various elements of $E_{k,\ell}$

Using a computer, one can check that the smallest element of $E_{2,3}$ is 6859, with

$$6859 = 19^3 \quad \text{and} \quad 6860 = 2^2 \cdot 5 \cdot 7^3.$$

Similarly, one can check that the smallest element of $E_{2,4}$ is 11 859 210, with

$$11\,859\,210 = 2 \cdot 3^4 \cdot 5 \cdot 11^4 \quad \text{and} \quad 11\,859\,211 = 7 \cdot 13 \cdot 19^4.$$

One may be surprised to learn that the next four elements of $E_{2,4}$ are much larger than its smallest element, as they are the four numbers

$$632\,127\,050\,601\,113\,666\,430, 1\,337\,735\,048\,956\,150\,266\,042\,387, \\ 2\,512\,088\,574\,784\,743\,818\,066\,896 \text{ and } 4\,974\,823\,008\,148\,736\,705\,412\,779.$$

It is as if the number 11859210 is completely out of tune with the other numbers with the same property; let's say it was lucky! This seemingly abnormal behaviour does happen at times, one reason being that the multiplicative structure of consecutive integers is very unpredictable.

In Section 7, we will show how one can attempt to find elements of the various sets $E_{2,\ell}$ for any given integer $\ell \geq 2$.

Before embarking on a broad study of the various sets $E_{k,\ell}$, we introduce some notation and state preliminary results.

3. Notation and preliminary results

From here on, the letters p and q will be used exclusively for primes. In particular, we write $\sum_p f(p)$ (resp. $\prod_p f(p)$) to denote sums (resp. products) which run over all primes p . We denote by $\pi(x)$ the number of primes $p \leq x$. The smallest prime factor of an integer $n \geq 2$ is denoted by $p(n)$.

Given two functions f and g defined on $[a, \infty)$, where $a \geq 0$ and $g(x) > 0$, we will write that $f(x) = O(g(x))$ if there exist two constants $M > 0$ and x_0 such that $|f(x)| < Mg(x)$ for all $x \geq x_0$. In particular, $f(x) = O(1)$ if $f(x)$ is a bounded function. Moreover, instead of writing $f(x) = O(g(x))$, we shall at times write $f(x) \ll g(x)$. Similarly, we shall write $f(x) = o(g(x))$ as $x \rightarrow \infty$ if, for any given real number $\varepsilon > 0$, there exists a constant $x_0 = x_0(\varepsilon)$ such that $|f(x)| < \varepsilon g(x)$ for all $x \geq x_0$. Also, we shall write $f(x) \gg g(x)$ if there exist two constants $M > 0$ and x_0 such that $|f(x)| > M|g(x)|$ for all $x \geq x_0$. Finally, we will write $f(x) \asymp g(x)$ if we simultaneously have $f(x) \ll g(x)$ and $g(x) \ll f(x)$, in which case we say that $f(x)$ and $g(x)$ are of the same order.

A continuous function $L : [a, \infty) \rightarrow \mathbb{R}^+$, where $a > 0$, is called a *slowly oscillating function* (or *slowly varying function*) if given any fixed number $c > 0$, we have $\lim_{x \rightarrow \infty} L(cx)/L(x) = 1$. We write \mathcal{L} for the set of slowly oscillating functions. Some common elements of \mathcal{L} are $\log x$, $\log \log x$ and $e^{(\log x)^\alpha}$ for any fixed number $\alpha \in (0, 1)$. It can be shown that any function $L : [a, \infty) \rightarrow \mathbb{R}^+$ with a continuous derivative (for all $x \geq a$) belongs to \mathcal{L} if and only if $\frac{xL'(x)}{L(x)} = o(1)$ as $x \rightarrow \infty$ (see Chapter 1 in the book of Seneta [17]). One of the nice properties of differentiable and increasing slowly oscillating functions L is that

$$(3.1) \quad \int_a^x \frac{dt}{L(t)} = (1 + o(1)) \frac{x}{L(x)} \quad (x \rightarrow \infty).$$

Indeed, using integration by parts and setting $\eta(t) := tL'(t)/L(t)$, we obtain

$$\begin{aligned} \int_a^x \frac{dt}{L(t)} &= \frac{t}{L(t)} \Big|_a^x + \int_a^x t \frac{L'(t)}{L(t)} \frac{1}{L(t)} dt = \\ &= \frac{x}{L(x)} + O(1) + \int_a^x \eta(t) \frac{dt}{L(t)} = \frac{x}{L(x)} + O(1) + o(1) \cdot \int_a^x \frac{dt}{L(t)}, \end{aligned}$$

from which (3.1) clearly follows.

We will be using the *prime number theorem* in one of its basic forms, namely

$$(3.2) \quad \pi(x) = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right),$$

and *Mertens' theorem* in the form

$$(3.3) \quad \prod_{p \leq x} \left(1 - \frac{1}{p}\right) = (1 + o(1)) \frac{e^{-\gamma}}{\log x} \quad (x \rightarrow \infty),$$

where $\gamma := \lim_{N \rightarrow \infty} \left(\sum_{n=1}^N \frac{1}{n} - \log N\right)$ is the Euler-Mascheroni constant.

4. The size of $E_{2,2}$

The 12 smallest elements of the set $E_{2,2}$ are

$$8, 49, 242, 288, 675, 1444, 1681, 2400, 2645, 6727, 6859, 9408.$$

The fact that this set is infinite follows from the Fermat-Pell equation $x^2 - 2y^2 = 1$, as it is well-known that this equation has infinitely many integer solutions. A nice proof of this can be found in the book of Joseph Silverman [18].

Once it is known that a particular set of positive integers E is infinite, a natural question arises: "How many elements of E are there below a given number x ?" So, let us introduce the counting function of the set $E_{2,2}$, that is,

$$E_{2,2}(x) := \#\{n \leq x : n \in E_{2,2}\}.$$

We will first search for a lower bound for $E_{2,2}(x)$. Given an integer $\ell \geq 2$, let G_ℓ be as defined in Section 1 and set $G_\ell(x) := \#\{n \leq x : n \in G_\ell\}$. For now, we are interested in G_2 , more precisely in finding integers n such that $n, n+1 \in G_2$. For this, we observe that

$$(4.1) \quad (2p^2 - 1)^2 - 1 = 4p^2(p-1)(p+1) \quad \text{for all primes } p.$$

It is clear that both $(2p^2 - 1)^2$ and $4p^2(p - 1)(p + 1)$ are in G_2 , implying that each corresponding number $n = 4p^2(p - 1)(p + 1) \in E_{2,2}$. How many of these n are there below x ? This amounts to asking how many primes p are there such that $(2p^2 - 1)^2 \leq x$. This last inequality is successively equivalent to

$$(4.2) \quad 2p^2 - 1 \leq \sqrt{x}, \quad p^2 \leq \frac{\sqrt{x} + 1}{2}, \quad p \leq \sqrt{\frac{\sqrt{x} + 1}{2}},$$

which in light of the prime number theorem in the form (3.2) and using the last inequality in (4.2) allows us to conclude that

$$(4.3) \quad E_{2,2}(x) \gg \frac{x^{1/4}}{\log x}.$$

What about a non trivial upper bound for $E_{2,2}(x)$? In 2013, De Koninck, Doyon and Luca [5] proved that

$$(4.4) \quad E_{2,2}(x) \ll x \exp\left\{-\frac{25}{24}\sqrt{2 \log x \log \log x}\right\}.$$

In 2020, de la Bretèche and Drappeau [8] showed that the constant $25/24$ can be replaced by 1.33. One may wonder which of the lower bound (4.3) and upper bound (4.4) is closer to the real value of $E_{2,2}(x)$? We will get back to this question in the last sections of this paper.

5. Searching for elements of $E_{3,2}$

Recall that $E_{3,2} = \{n \in \mathbb{N} : P(n)^2 \mid n, P(n+1)^2 \mid n+1, P(n+2)^2 \mid n+2\}$. Setting $E_{3,2}(x) := \#\{n \leq x : n \in E_{3,2}\}$ and using a computer, it is possible to show that $E_{3,2}(10^{15}) = 60$. The list of all elements of $E_{3,2}$ with 15 digits or less is given in De Koninck and Moineau [7].

How can one quickly generate elements of $E_{3,2}$? Instead of searching for three consecutive integers each belonging to G_2 , we will examine a bigger picture and search for consecutive polynomials $f(x)$, $f(x)+1$, $f(x)+2$, with integer coefficients, hoping that at integer arguments these three polynomials will reveal elements of $E_{3,2}$. Why this approach? Polynomials have many features that integers don't have, including having a degree, roots and a derivative. Let us see how this works.

As we did in [7], consider the consecutive polynomials

$$(5.1) \quad \begin{aligned} f(x) &= (2x^2 + 1)^2(x - 1)(x + 1), \\ f(x) + 1 &= x^2(4x^4 - 3), \\ f(x) + 2 &= (2x^2 - 1)^2(x^2 + 1). \end{aligned}$$

With $x = 5087$, we find the somewhat larger 23-digit numbers

$$\begin{aligned} 69\,315\,509\,064\,481\,032\,011\,329 &= 2^6 \cdot 3^7 \cdot 53 \cdot 53 \cdot 2543 \cdot 1916857^2, \\ 69\,315\,509\,064\,481\,032\,011\,330 &= 11 \cdot 71 \cdot 769 \cdot 1163 \cdot 1321 \cdot 2903 \cdot 5087^2, \\ 69\,315\,509\,064\,481\,032\,011\,331 &= 2 \cdot 5 \cdot 7^2 \cdot 17 \cdot 29^2 \cdot 167^2 \cdot 181 \cdot 44273^2. \end{aligned}$$

The bad news with the system of polynomials (5.1) is that the polynomials involved are of degree 6 and are therefore bound (unfortunately) to generate very large elements of $E_{3,2}$, thus skipping many of the small elements of $E_{3,2}$ which we would also like to identify. This motivates replacing x^2 by x in (5.1), yielding a new system of degree 3 polynomials

$$(5.2) \quad \begin{aligned} g(x) - 1 &= (2x + 1)^2(x - 1), \\ g(x) &= x(4x^2 - 3), \\ g(x) + 1 &= (2x - 1)^2(x + 1). \end{aligned}$$

The good news with this new system is that the polynomials involved are of degree 3. The bad news is that we lost the square factor in the polynomial $g(x)$. To overcome this hurdle, we choose $x = mP(m)$, a number clearly belonging to G_2 . Letting m run up to 100 000 000, we find 13 042 elements of $E_{3,2}$.

Another way to take advantage of system (5.2) is to allow for “half-integers”, that is by setting $x = y/2$ in the above system, which yields the new system

$$(5.3) \quad \begin{aligned} f(y) - 1 &= \frac{1}{2}(y - 2)(y + 1)^2, \\ f(y) &= \frac{1}{2}y(y^2 - 3), \\ f(y) + 1 &= \frac{1}{2}(y + 2)(y - 1)^2, \end{aligned}$$

which nevertheless always yield three consecutive integers. This approach allows one to find more elements of $E_{3,2}$. Indeed, system (5.3) is more “prolific” than system (5.2); for instance, the number $122\,169\,948\,877\,430 \in E_{3,2}$ is found through system (5.3), but not through system (5.2).

Using a variety of approaches, including the systems of polynomials mentioned above and others, we found 12 027 elements of $E_{3,2}$ with 24 digits or less. The following is a table showing the number of elements of $E_{3,2}$ we found so far.

Number of digits	Number of elements	Number of digits	Number of elements
7	2	16	53
8	0	17	101
9	1	18	187
10	2	19	442
11	4	20	941
12	4	21	1 927
13	5	22	2 974
14	14	23	3 202
15	28	24	2 140

Table 1: Number of elements of $E_{3,2}$ found so far

In Table 1, the number of elements of $E_{3,2}$ in each k -digits category with $k \leq 18$ is accurate. However, for each $k \in \{19, \dots, 24\}$, the real figures for the number of k -digits numbers belonging to $E_{3,2}$ are most likely much larger than the ones listed above. A list of all the known members of $E_{3,2}$ with no more than 21 digits is available on the first author's web site [3].

6. Searching for elements of $E_{3,3}$

Recall that

$$E_{3,3} := \{n \in \mathbb{N} : P(n+i)^3 \mid n+i, i=0,1,2\}.$$

Can one find three consecutive polynomials of the form

$$(6.1) \quad \begin{aligned} g(x) - 1 &= (2x - 1)^3 \times g_1(x), \\ g(x) &= x^3 \times g_2(x), \\ g(x) + 1 &= (2x + 1)^3 \times g_3(x), \end{aligned}$$

where the $g_i(x)$'s are of minimal degree d ? To find this "small" number d , we will rely on Mason's theorem.

Mason's theorem (1984). *Let a, b, c be nonzero co-prime polynomials, not all constant. If $a + b = c$, then*

$$\max\{\deg(a), \deg(b), \deg(c)\} < N_0(abc),$$

where $N_0(abc)$ denotes the number of distinct roots of the polynomial abc .

An elegant proof of Mason's theorem has been given by Noah Snyder [19]. This result is a mathematical theorem about polynomials, analogous to the well known abc conjecture for integers.

Applying Mason's theorem to the polynomials in (6.1), we seek to find the minimum value of d . To do so, write

$$\underbrace{g(x)^2}_{c(x)} = \underbrace{(g(x) - 1)(g(x) + 1)}_{a(x)} + \underbrace{1}_{b(x)}.$$

Mason's theorem implies that

$$\begin{aligned} \max\{\deg a(x), \deg b(x), \deg c(x)\} &< N_0((g(x) - 1)g(x)(g(x) + 1)), \\ 2(3 + d) &< 3(1 + d), \end{aligned}$$

implying that $d > 3$, so that $d \geq 4$. So, the polynomials g_i will each be of degree 4. Now, how can we obtain each $g_i(x)$ explicitly? To do so, we first observe that since the three polynomials appearing in system (6.1) only differ by a constant, they must have the same derivative. Hence we have that

$$(g(x) - 1)' = (g(x))' = (g(x) + 1)'.$$

It is clear that this common derivative $f_0(x)$ is of degree 6 and moreover of the form

$$f_0(x) = ax^2(2x - 1)^2(2x + 1)^2 = ax^2(4x^2 - 1)^2$$

for some positive constant a . Hence, we have that

$$g(x) = \int f_0(x) dx,$$

implying that

$$g(x) = \int ax^2(4x^2 - 1)^2 dx.$$

This gives

$$g(x) = a \left(\frac{16x^7}{7} - \frac{8x^5}{5} + \frac{x^3}{3} \right).$$

Choosing $a = \text{LCM}[7, 5, 3] = 105$, we find that

$$g(x) = x^3(240x^4 - 168x^2 + 35).$$

From this explicit form of $g(x)$, we deduce that system (6.1) can be written as

$$\begin{aligned} g(x) - 1 &= (2x - 1)^3(30x^4 + 45x^3 + 24x^2 + 6x + 1), \\ g(x) &= x^3(240x^4 - 168x^2 + 35), \\ g(x) + 1 &= (2x + 1)^3(30x^4 - 45x^3 + 24x^2 - 6x + 1). \end{aligned}$$

Using a computer, one will find that at $x = 39\,682\,272\,446$ the above system reveals the 77-digit number n_1 which is such that

$$\begin{aligned} n_1 - 1 &= 3^3 \cdot 137 \cdot 251 \cdot 49253 \cdot 6892241 \cdot 1400173417 \\ &\quad \cdot 1749071927 \cdot 2602138829 \cdot 26454848297^3, \\ n_1 &= 2^3 \cdot 1162 \cdot 31 \cdot 3301 \cdot 92639 \cdot 376627 \cdot 474994139 \\ &\quad \cdot 573384841 \cdot 5057839271 \cdot 19841136223^3, \\ n_1 + 1 &= 13 \cdot 23 \cdot 41 \cdot 61^3 \cdot 233 \cdot 3767 \cdot 9551 \cdot 977719 \\ &\quad \cdot 5076637 \cdot 367839041 \cdot 396464197 \cdot 1301058113^3. \end{aligned}$$

In 2014, Péter Burcsi and Gabor Gévay (private communication) found the 77-digit number n_2 which satisfies

$$\begin{aligned} n_2 - 1 &= 2^7 \cdot 53 \cdot 4253 \cdot 27631 \cdot 27953 \cdot 1546327 \cdot 2535271 \\ &\quad \cdot 17603683 \cdot 1472289739 \cdot 16476952799^3, \\ n_2 &= 3^6 \cdot 19 \cdot 37 \cdot 787 \cdot 711163 \cdot 2181919 \cdot 137861107 \\ &\quad \cdot 318818473 \cdot 937617607 \cdot 7323090133^3, \\ n_2 + 1 &= 2 \cdot 12899 \cdot 133451 \cdot 421607 \cdot 2198029 \cdot 8046041 \\ &\quad \cdot 19854409 \cdot 555329197 \cdot 32953905599^3. \end{aligned}$$

Most likely, the above 77-digit numbers are not the two smallest elements of $E_{3,3}$.

7. Searching for elements of $E_{2,\ell}$

How can one find integers n such that $P(n)^6 \mid n$ and $P(n+1)^6 \mid n+1$? One way is to look for a polynomial $g(x)$ such that

$$(7.1) \quad \begin{cases} g(x) &= x^6 &\times g_1(x), \\ g(x) + 1 &= (x-1)^6 &\times g_2(x), \end{cases}$$

where the polynomials $g_i(x)$ are of degree 5, as a consequence of Mason's theorem. Indeed, letting d be the minimal degree of the polynomials g_i 's and rewriting the system (7.1) as

$$\begin{cases} h(x) - 1 &= x^6 &\times g_1(x), \\ h(x) &= (x-1)^6 &\times g_2(x), \end{cases}$$

we have the trivial representation

$$\underbrace{h(x)}_{c(x)} = \underbrace{h(x) - 1}_{a(x)} + \underbrace{1}_{b(x)}.$$

Mason's theorem implies that

$$\max\{\deg h(x), \deg(h(x) - 1), \deg 1\} < N_0(h(x)(h(x) - 1))$$

and therefore that $d + 6 < 2(d + 1)$, that is, $d > 4$, which means that $d \geq 5$. Hence, we can assume that the polynomials $g_1(x)$ and $g_2(x)$ are of degree 5. As before, this means that $g(x)$ and $g(x) + 1$ have a common derivative of degree $5 + 6 = 11$, with factors x^5 and $(x - 1)^5$. Therefore, $g(x) = \int a x^5 (x - 1)^5 dx$ for some positive constant a . Choosing $a = 2772$, we finally obtain that

$$\begin{aligned} g(x) &= x^6(252x^5 - 1386x^4 + 3080x^3 - 3465x^2 + 1980x - 462), \\ g(x) + 1 &= (x - 1)^6(252x^5 + 126x^4 + 56x^3 + 21x^2 + 6x + 1). \end{aligned}$$

Choosing $x = 20\,905\,825\,364$, we find the 116-digit integer n_3 which satisfies

$$\begin{aligned} n_3 &= 2^{13} \cdot 5 \cdot 83 \cdot 22157 \cdot 127139 \cdot 12177577 \cdot 17565259 \\ &\quad \cdot 372289003 \cdot 1659308773 \cdot 3257215037 \cdot 5226456341^6, \\ n_3 + 1 &= 569 \cdot 32939 \cdot 122489 \cdot 146359 \cdot 50300881 \cdot 919974911 \\ &\quad \cdot 4166729363 \cdot 15532846993 \cdot 20905825363^6. \end{aligned}$$

Given an arbitrary integer $\ell \geq 2$, the following addresses the general case of $E_{2,\ell}$.

Theorem 1. *Let $\ell \geq 2$ be a fixed integer. Then, there exist $g_1(x), g_2(x) \in \mathbb{Z}[x]$ each of degree $\ell - 1$ such that*

$$(7.2) \quad x^\ell \cdot g_1(x) + (-1)^\ell = (x - 1)^\ell \cdot g_2(x).$$

Explicitly, for ℓ even and $C = \frac{(\ell - 1)! \cdot (\ell - 1)!}{(2\ell - 1)!}$, we obtain

$$g_1(x) = \frac{1}{C} \left(\frac{x^{\ell-1}}{2\ell-1} - \binom{\ell-1}{1} \frac{x^{\ell-2}}{2\ell-2} + \cdots + \binom{\ell-1}{\ell-2} \frac{x}{\ell+1} - \frac{1}{\ell} \right).$$

Applying Theorem 1 to each $\ell = 2, 3, 4, 5$, we obtain respectively the following pairs of polynomials:

$$\begin{aligned} g(x) &= (2x - 1)(x + 1)^2, \\ g(x) + 1 &= (2x + 3)x^2. \end{aligned}$$

$$\begin{aligned} g(x) &= (6x^2 + 15x + 10)x^3, \\ g(x) + 1 &= (6x^2 - 3x + 1)(x + 1)^3. \end{aligned}$$

$$\begin{aligned} g(x) &= (20x^3 - 10x^2 + 4x - 1)(x + 1)^4, \\ g(x) + 1 &= (20x^3 + 70x^2 + 84x + 35)x^4. \end{aligned}$$

$$\begin{aligned}g(x) &= (70x^4 + 315x^3 + 540x^2 + 420x + 126)x^5, \\g(x) + 1 &= (70x^4 - 35x^3 + 15x^2 - 5x + 1)(x + 1)^5.\end{aligned}$$

A detailed proof of Theorem 1 is given in [7].

Finally, regarding the sets $E_{2,\ell}$, let us add that recently, Karl Dilcher and Maciej Ulas [9], mentioning the above Theorem 1 as part of their motivation, revealed interesting properties of the polynomials appearing in (7.2).

8. Searching for elements of $E_{4,2}$ using polynomials

As of now no members of $E_{4,2}$ are known. A naive approach is to examine numbers $n \in E_{3,2}$ and check if either $n - 1 \in G_2$ or $n + 3 \in G_2$, since any of these cases would reveal a member of $E_{4,2}$. After all, as we showed in Section 5, we have at hand 13 042 members of $E_{3,2}$.

Using an approach through consecutive polynomials similar to the ones used above, but this time with the objective of finding elements of $E_{4,2}$, one could consider the system of polynomials

$$(8.1) \quad \begin{cases} f(x) &= \frac{1}{18}x^2(10x^5 - 105x^4 + 427x^3 - 840x^2 + 805x - 315), \\ f(x) + 1 &= \frac{1}{18}(x - 1)^2(10x^5 - 85x^4 + 247x^3 - 261x^2 + 36x + 18), \\ f(x) + 2 &= \frac{1}{18}(x - 2)^2(10x^5 - 65x^4 + 127x^3 - 72x^2 + 9x + 9), \\ f(x) + 3 &= \frac{1}{18}(x - 3)^2(10x^5 - 45x^4 + 67x^3 - 33x^2 + 4x + 6) \end{cases}$$

and try replacing x by 1, 2, 3, ... expecting to eventually find an element of $E_{4,2}$. This approach, although theoretically interesting, is practically hopeless, as we will now see. For this we need some background results on “friable numbers”, also called “smooth numbers”.

An integer $n \geq 2$ is said to be y -friable (or y -smooth) if $P(n) \leq y$. To count how many y -friable numbers there are up to a given number x , it will be convenient to work with the function

$$\Psi(x, y) := \#\{n \leq x : P(n) \leq y\} \quad (2 \leq y \leq x).$$

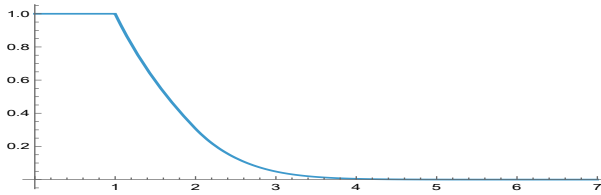
Setting $u = \log x / \log y$ and keeping u fixed, one can show (see Theorem 9.3 in the book of De Koninck and Luca [5]) that

$$(8.2) \quad \Psi(x, y) = (1 + o(1))x\rho(u) \quad (x \rightarrow \infty),$$

where $\rho(u)$ is the *Dickman function*, which is defined as the unique continuous function $\rho : [0, \infty) \rightarrow (0, 1]$ which is differentiable on $[1, \infty)$ and satisfies

$$\begin{aligned}\rho(u) &= 1 && \text{for } 0 \leq u \leq 1, \\ u\rho'(u) + \rho(u - 1) &= 0 && \text{for } u \geq 1.\end{aligned}$$

The function $\rho(u)$ decreases very rapidly as u increases, as suggested by the following graph and table of values.



k	1	2	3	4
$\rho(k)$	1.0	0.3068	0.0486	0.00491
k	5	6	7	
$\rho(k)$	0.000354	0.0000196	0.00000087	

Table 2: Values of $\rho(k)$ for $k \in \{1, 2, \dots, 7\}$

It follows from (8.2) that for each positive integer k , as $x \rightarrow \infty$,

$$\begin{aligned} \Psi(x, x^{1/k}) &= \#\{n \leq x : P(n) \leq x^{1/k}\} = (1 + o(1))x\rho\left(\frac{\log x}{\log x^{1/k}}\right) \\ &= (1 + o(1))x\rho(k). \end{aligned}$$

Given an irreducible polynomial $f \in \mathbb{Z}[x]$, what can we say about the friable values of $f(n)$? In 2002, Greg Martin [15] conjectured that for such a function f , we should have

$$(8.3) \quad \#\{n \leq x : P(f(n)) \leq y\} = (1 + o(1))x\rho\left(\frac{\log f(x)}{\log y}\right) \quad (x \rightarrow \infty).$$

How is that of any help in gauging the size of the smallest elements of $E_{4,2}$? We discuss this in the next section.

9. Returning to the search for elements in $E_{4,2}$

If Martin's conjecture (8.3) is true, then, for large n ,

$$\text{Prob}(P(f(n)) \leq y) = \rho\left(\frac{\log f(n)}{\log y}\right).$$

Writing the system of polynomials (8.1) as

$$\begin{aligned} f(x) &= \frac{1}{18}x^2 \cdot f_0(x), \\ f(x) + 1 &= \frac{1}{18}(x-1)^2 \cdot f_1(x), \\ f(x) + 2 &= \frac{1}{18}(x-2)^2 \cdot f_2(x), \\ f(x) + 3 &= \frac{1}{18}(x-3)^2 \cdot f_3(x), \end{aligned}$$

in order for the above to eventually reveal a number r such that $f(r) \in E_{4,2}$, this particular number r will have to satisfy the four conditions

$$P(f_i(r)) \leq P(r-i) \quad \text{for } i = 0, 1, 2, 3.$$

According to (8.3), this will occur (relying on the values given in Table 2) with a probability smaller than

$$\rho(5)^4 \approx 0.000354^4 \approx 1.57 \cdot 10^{-16},$$

implying that the smallest such r would be larger than 6×10^{15} , with a corresponding $n \in E_{4,2}$ whose size would be larger than $(6 \times 10^{15})^7 \approx 10^{109}$, meaning that this number n should have at least 109 digits. Not good! All this to say that we need another approach, one that will allow us to find a much smaller element of $E_{4,2}$. In fact, we will see in Section 13 that the smallest element of $E_{4,2}$ is most likely a number with fewer than 40 digits.

10. Two algorithms for generating all elements of G_ℓ below a given bound

We start by taking a closer look at the set $G_2 := \{n \in \mathbb{N} : P(n)^2 \mid n\}$ and to its counting function $G_2(x) := \#\{n \leq x : n \in G_2\}$. How can one manage to quickly obtain the list of all elements of G_2 up to a given number x , say $x = 2^{30}$? Perhaps naively, one could compute the set

$$\{m \cdot P(m) : 1 \leq m \leq 2^{15}\},$$

which is clearly a subset of the desired set $\{n \leq 2^{30} : n \in G_2\}$. Unfortunately, this approach will be missing too many elements of G_2 .

Another way would be to try to check one by one each positive integer $n \leq 10^9$ to see if it belongs to G_2 . This approach would be very time consuming, the main reason being that one would need to factor each such n , and it is well known that ‘‘factoring’’ cannot (so far) be done in polynomial time.

As we will now see, there are other ways.

10.1. A first algorithm

We create a first algorithm, one that can quickly generate all the elements of G_2 below any given power of 2. In fact, as we will see, this algorithm runs in polynomial time (that is, very fast), by opposition to most algorithms which involve factoring, which can only be done in exponential time or at best in sub-exponential time (that is, very slowly). For a rigorous explanation of the distinction between a “polynomial time” algorithm and an “exponential time” or “subexponential time” algorithm, see Episode 35 in the book of De Koninck and Doyon [7].

The idea behind the construction of the algorithm is the following. Fix an integer N and observe that if $n \in (N/2, N] \cap G_2$, then $2n \in (N, 2N] \cap G_2$, and similarly for any prime p , that if $n \in (N/p, 2N/p] \cap G_2$ and $p(n) \geq p$, then $pn \in (N, 2N] \cap G_2$. This simple observation means that knowing all $n \in [1, N] \cap G_2$ allows one to obtain all $n \in (N, 2N] \cap G_2$, since

$$\begin{array}{ll}
 n \in G_2 \text{ and } \frac{N}{2} < n \leq N & \text{if and only if } 2n \in G_2 \\
 & \text{and } N < 2n \leq 2N \\
 n \in G_2, \frac{N}{3} < n \leq \frac{2N}{3} \text{ and } p(n) \geq 3 & \text{if and only if } 3n \in G_2, \\
 & N < 3n \leq 2N \\
 & \text{and } p(3n) = 3 \\
 n \in G_2, \frac{N}{5} < n \leq \frac{2N}{5} \text{ and } p(n) \geq 5 & \text{if and only if } 5n \in G_2, \\
 & N < 5n \leq 2N \\
 & \text{and } p(5n) = 5 \\
 & \vdots \\
 n \in G_2, \frac{N}{p_r} < n \leq \frac{2N}{p_r} \text{ and } p(n) \geq p_r & \text{if and only if } p_r n \in G_2, \\
 & N < p_r n \leq 2N \\
 & \text{and } p(p_r n) = p_r
 \end{array}$$

for n not prime, and where $r = \pi((2N)^{1/3})$. This choice of r comes from the string of inequalities

$$2N \geq p_r n \geq p_r p(n)^2 \geq p_r^3.$$

The above allows one to identify most $n \in G_2$ located in the interval $(N, 2N]$, but not all, since some elements of $(N, 2N] \cap G_2$ do not appear in the above enumeration. Indeed, there remains the squares of primes located between N and $2N$, which are obviously also members of G_2 . We must therefore add the $\pi(\sqrt{2N}) - \pi(\sqrt{N})$ numbers $p^2 \in (N, 2N]$.

This is why we have

$$\begin{aligned} \{n \in (N, 2N] : n \in G_2\} = & \\ = & 2 \cdot \{n \in (N/2, N] \cap G_2\} \cup \\ & \cup 3 \cdot \{n \in (N/3, 2N/3] \cap G_2 \text{ and } p(n) \geq 3\} \cup \\ & \cup 5 \cdot \{n \in (N/5, 2N/5] \cap G_2 \text{ and } p(n) \geq 5\} \cup \\ & \cup \dots \cup p_r \cdot \{n \in (N/p_r, 2N/p_r] \cap G_2 \text{ and } p(n) \geq p_r\} \cup \\ & \cup \{p^2 : \sqrt{N} < p \leq \sqrt{2N}\}. \end{aligned}$$

Because the sets $p_i \cdot \{n \in (N/p_i, 2N/p_i] \cap G_2 \text{ and } p(n) \geq p_i\}$ coincide with $\{p_i n \in (N, 2N] \cap G_2 \text{ and } p(p_i n) = p_i\}$, these sets are disjoint. In practice, the condition $p(n) \geq p_i$ is equivalent to requiring that n is not divisible by p_{i-1} , as all numbers with smaller prime factors have been accounted for in previous terms of the union. This algorithm can be programmed using MATHEMATICA: see the first author’s web site [3].

Using the algorithm, one can obtain the following table of the exact values of $G_2(N)$ for $N = 2^k$, $k = 5, \dots, 43$.

$N = 2^k$	$G_2(N)$	$N = 2^k$	$G_2(N)$	$N = 2^k$	$G_2(N)$
2^5	8	2^{18}	3 642	2^{31}	2 059 504
2^6	13	2^{19}	5 838	2^{32}	3 404 363
2^7	22	2^{20}	9 397	2^{33}	5 637 783
2^8	36	2^{21}	15 154	2^{34}	9 353 062
2^9	57	2^{22}	24 517	2^{35}	15 543 299
2^{10}	89	2^{23}	39 757	2^{36}	25 873 449
2^{11}	142	2^{24}	64 606	2^{37}	43 138 297
2^{12}	224	2^{25}	105 249	2^{38}	72 036 070
2^{13}	354	2^{26}	171 873	2^{39}	120 475 143
2^{14}	562	2^{27}	281 236	2^{40}	201 782 341
2^{15}	893	2^{28}	461 254	2^{41}	338 446 875
2^{16}	1 425	2^{29}	758 039	2^{42}	568 461 667
2^{17}	2 276	2^{30}	1 248 269	2^{43}	956 088 615

Table 3: Values of $G_2(2^k)$ for $k = 5, 6, \dots, 43$

By slightly adapting the above algorithm, one will obtain all the elements of G_ℓ for any given integer $\ell \geq 3$. For instance, doing so, one will obtain that $G_3(2^{52}) = 733\,981\,869$, $G_4(2^{57}) = 190\,689\,911$, $G_5(2^{66}) = 337\,714\,064$, $G_6(2^{74}) = 449\,806\,051$ and $G_7(2^{79}) = 248\,763\,764$.

Remark 10.1. The only limit to the expansion of Table 3 is storage capacity. Indeed, building the list of all those $n \in (N, 2N] \cap G_2$ requires knowledge of all those $n \in (N/2, N] \cap G_2$, which therefore must have been previously stored.

Remark 10.2. As mentioned at the beginning of this section, it would be laborious to establish by “brute force” that $G_2(2^{30}) = 1\,248\,269$. Using the above algorithm on a computer, one can obtain the same outcome in seconds.

Remark 10.3. The algorithm can be further improved by first adapting it to find all odd positive integers belonging to $[9, 2^k) \cap G_2$. Let A be the resulting set. Then, the set $[4, 2^k] \cap G_2$ can be obtained by simply gathering the disjoint sets

$$A, \quad 2 \cdot A \cap [4, 2^{k-1}), \quad 2^2 \cdot A \cap [4, 2^{k-2}), \quad \dots, \quad 2^{k-2} \cdot A \cap [4, 2^2).$$

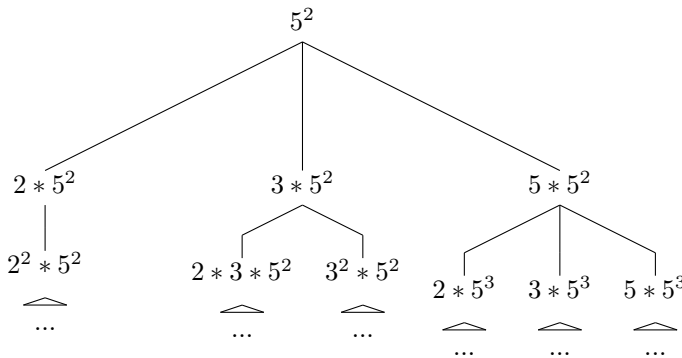
Doing so allows one to obtain the value of $G_2(2^{30})$ at least twice as fast.

In the above algorithm, we showed how one could use the set of elements of G_2 contained in the interval $[1, N)$ in order to obtain those $n \in G_2$ belonging to $[N, 2N)$. We also showed how this algorithm can be generalized to yield all the elements of G_ℓ for an arbitrary integer $\ell \geq 2$. We will now exhibit another algorithm, one that does not require to store those elements of G_ℓ already found.

10.2. A second algorithm

Let N be a positive integer and assume that our first goal is to compute $G_2(N)$. Let $M = \pi(\sqrt{N})$, so that p_M is the largest prime number not exceeding \sqrt{N} . For each positive integer $m \leq M$, we will count the number of m -tuples of non negative integers (a_1, a_2, \dots, a_m) such that $\prod_{i=1}^m p_i^{a_i} \leq N$ with $a_m \geq 2$. Here p_1, p_2, \dots stands for the list of all primes in increasing order.

To do so, we recursively construct a tree of all valid combinations and count its nodes. Here is an example of what the tree looks like in the particular case $p_m = 5$:



In the above tree, we counted so far 10 nodes. The way we did it is that, at each step of the recursive function, we simply considered those primes q such that $2 \leq q \leq p_m$ and $q \times \text{node} \leq N$, so that each of the nodes is counted exactly once. The above approach can be generalized in order to evaluate $G_\ell(N)$ for an arbitrary integer $\ell \geq 2$.

In fact, the algorithm to compute $G_\ell(N)$ (for a given integer $\ell \geq 2$) – which we chose to call Algorithm B – can easily be programmed using python: see the first author’s web site [3].

The main advantage of Algorithm B over Algorithm A is that one only needs to store the primes $\leq N^{1/\ell}$ instead of all the actual elements of G_ℓ not exceeding $N/2$. Another non negligible advantage of Algorithm B is that its running (polynomial) time is about ten times faster than that of Algorithm A.

11. The distance between successive elements of G_2

With no surprise, the sum of the reciprocals of the elements of G_2 converges. This follows from the fact that, using Mertens' theorem (see (3.3)), we have

$$\begin{aligned} \sum_{n \in G_2} \frac{1}{n} &= \sum_p \frac{1}{p^2} \prod_{q \leq p} \left(1 + \frac{1}{q} + \frac{1}{q^2} + \dots\right) = \sum_p \frac{1}{p^2} \prod_{q \leq p} \left(1 - \frac{1}{q}\right)^{-1} \ll \\ &\ll \sum_p \frac{\log p}{p^2} < \infty. \end{aligned}$$

As a consequence, if we write $g_1 < g_2 < \dots$ for the list of all elements of G_2 , then,

$$\limsup_{i \rightarrow \infty} (g_{i+1} - g_i) = \infty.$$

Note that in Section 4, we proved that $\liminf_{i \rightarrow \infty} (g_{i+1} - g_i) = 1$.

Now that we know that the distance between consecutive elements of G_2 can be arbitrarily large, one might inquire about the average distance between consecutive elements of G_2 . To do so, we first prove the following.

Theorem 2. *Letting λ stand for the Golomb–Dickman constant defined by*

$$\lambda = \int_0^\infty \frac{\rho(u)}{u+2} du \approx 0.62433,$$

we have

$$(11.1) \quad \sum_{\substack{n \leq x \\ n \in G_2}} P(n) = \lambda x \left(1 + O\left(\frac{\log \log x}{\sqrt{\log x}}\right)\right).$$

Proof. First write

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \in G_2}} P(n) &= \sum_{p \leq \sqrt{x}} p \sum_{\substack{mp^2 \leq x \\ P(m) \leq p}} 1 = \sum_{p \leq \sqrt{x}} p \Psi\left(\frac{x}{p^2}, p\right) = \\ &= \sum_{p \leq \exp\{\sqrt{\log x}\}} p \Psi\left(\frac{x}{p^2}, p\right) + \sum_{\exp\{\sqrt{\log x}\} < p \leq \sqrt{x}} p \Psi\left(\frac{x}{p^2}, p\right) = \\ (11.2) \quad &= S_1(x) + S_2(x). \end{aligned}$$

To bound the sum $S_1(x)$, first recall the inequality

$$(11.3) \quad \Psi(x, y) \ll x \exp \left\{ -\frac{1}{2} \frac{\log x}{\log y} \right\} \quad (2 \leq y \leq x).$$

(For a proof of this, see Théorème 5.1 in the book of Tenenbaum [20]). Using this inequality, we may write that

$$(11.4) \quad \begin{aligned} S_1(x) &\ll \sum_{p \leq \exp\{\sqrt{\log x}\}} p \frac{x}{p^2} \exp \left\{ -\frac{1}{2} \left(\frac{\log x}{\log p} - 2 \right) \right\} \ll \\ &\ll x \exp \left\{ -\frac{1}{2} \sqrt{\log x} \right\}. \end{aligned}$$

For the evaluation of $S_2(x)$, we will be using a classical result due to Hildebrand [11] which we state as follows.

Proposition 1. (Hildebrand) *Let $\varepsilon > 0$. Then, setting $u = \log x / \log y$, we have*

$$\Psi(x, y) = x \rho(u) \left(1 + O_\varepsilon \left(\frac{\log(u+1)}{\log y} \right) \right)$$

uniformly for $x \geq 3$ and $\exp\{(\log \log x)^{\frac{5}{3}+\varepsilon}\} \leq y \leq x$.

Choosing $y = \exp\{\sqrt{\log x}\}$ in Proposition 1 and reckoning that

$$\frac{\log(u+1)}{\log y} = O \left(\frac{\log \log x}{\sqrt{\log x}} \right),$$

it follows, using the prime number theorem in the form (3.2), that

$$(11.5) \quad \begin{aligned} S_2(x) &= \left(1 + O \left(\frac{\log \log x}{\sqrt{\log x}} \right) \right) \sum_{\exp\{\sqrt{\log x}\} < p \leq \sqrt{x}} p \frac{x}{p^2} \rho \left(\frac{\log x}{\log p} - 2 \right) = \\ &= x \left(1 + O \left(\frac{\log \log x}{\sqrt{\log x}} \right) \right) \int_{e^{\sqrt{\log x}}}^{\sqrt{x}} \frac{1}{t} \rho \left(\frac{\log x}{\log t} - 2 \right) d\pi(t) = \\ &= x \left(1 + O \left(\frac{\log \log x}{\sqrt{\log x}} \right) \right) \left(1 + O \left(\frac{1}{\log x} \right) \right) \times \\ &\quad \times \int_{\exp\{\sqrt{\log x}\}}^{\sqrt{x}} \rho \left(\frac{\log x}{\log t} - 2 \right) \frac{dt}{t \log t} = \\ &= x \left(1 + O \left(\frac{\log \log x}{\sqrt{\log x}} \right) \right) \int_2^{\sqrt{\log x}} \rho(u-2) \frac{du}{u} = \\ &= x \left(1 + O \left(\frac{\log \log x}{\sqrt{\log x}} \right) \right) \int_0^\infty \frac{\rho(w)}{w+2} dw, \end{aligned}$$

where we first used the change of variable $u = \log x / \log t$ and thereafter $w = u - 2$.

Finally, using (11.4) and (11.5) in (11.2), the proof of (11.1) is complete. ■

As an immediate consequence of Theorem 2, we have the following.

Corollary 1. The distance d between consecutive elements $n, n + d$ of G_2 is on average

$$\frac{P(n)}{\lambda} = P(n) \times 1.60172 \dots$$

12. Focusing on the size of $G_2(x)$

In 1986, Erdős, Ivić and Pomerance [10] proved that, as x becomes large,

$$(12.1) \quad \sum_{n \leq x} \frac{1}{P(n)} = x \delta(x) \left(1 + O \left(\sqrt{\frac{\log \log x}{\log x}} \right) \right),$$

where

$$\delta(x) := \int_2^x \rho \left(\frac{\log x}{\log t} \right) \frac{dt}{t^2}$$

can be approximated as

$$(12.2) \quad \delta(x) = \exp \left\{ -\sqrt{2 \log x \log \log x} (1 + g_0(x) + o(g_0(x))) \right\},$$

where

$$g_0(x) := \frac{\log_3 x - 2 - \log 2}{2 \log_2 x} \left(1 + \frac{2}{\log_2 x} \right) - \frac{(\log_3 x - \log 2)^2}{8 \log_2^2 x}$$

(here $\log_k x$ stands for the logarithm of x iterated k times).

In 2002, A. Ivić [12] proved that

$$(12.3) \quad G_2(x) = \sqrt{\frac{\log x}{2} (\log_2 x + \log_3 x - \log 2 + o(1))} \times \sum_{n \leq x} \frac{1}{P(n)}.$$

Combining (12.1), (12.2) and (12.3), we get

$$(12.4) \quad \begin{aligned} G_2(x) &= x \sqrt{\frac{\log x}{2} (\log_2 x + \log_3 x - \log 2 + o(1))} \times \\ &\times \exp \left\{ -\sqrt{2 \log x \log \log x} (1 + g_0(x) + o(g_0(x))) \right\}. \end{aligned}$$

The unfortunate situation with the approximation of $\delta(x)$ given in (12.2) is that the function $g_0(x)$ is negative for all $x < 10^{2834447}$ and positive for

all $x > 10^{2834448}$. Practically, this means that the approximation of $G_2(x)$ provided by estimate (12.4) only makes sense for extremely large values of x (namely for numbers with more than two million digits), which is not adequate since we are seeking accurate values of $G_2(x)$ for those numbers x with one hundred digits or less. This is why we need to find another approach. A first naive observation is “why make use of an approximation for $\delta(x)$ when we could work with the exact value of $\delta(x)$?”, a value which in turn will yield a more accurate value of $\sum_{n \leq x} 1/P(n)$ through (12.1). Indeed, by a simple change of variables, we get that

$$(12.5) \quad \delta(x) = \int_2^x \rho \left(\frac{\log x}{\log t} \right) \frac{dt}{t^2} = \log x \int_1^{\log x / \log 2} \frac{\rho(u)}{u^2} x^{-1/u} du = \log x \cdot I(x),$$

say, which inserted in (12.1) gives

$$(12.6) \quad \sum_{n \leq x} \frac{1}{P(n)} = (1 + o(1))x \log x \cdot I(x) \quad (x \rightarrow \infty).$$

Combining (12.3) and (12.6) gives

$$(12.7) \quad G_2(x) = x \log x \sqrt{\frac{\log x}{2} (\log \log x + \log \log \log x - \log 2 + o(1))} \cdot I(x).$$

As we will see, this last approximation of $G_2(x)$ is much better than the one given by (12.4).

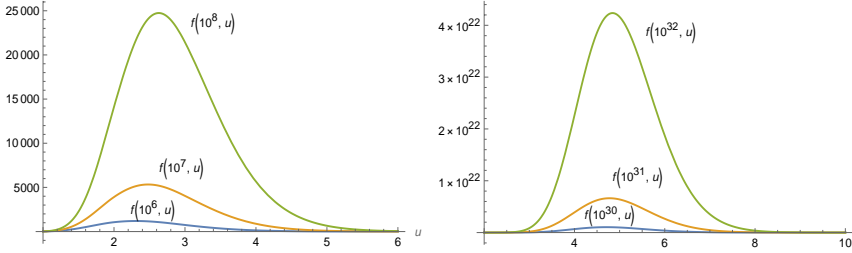
To see this, first set

$$f(x, u) := x^{1-1/u} \log x \frac{\rho(u)}{u^2},$$

so that, in light of (12.5) and (12.6) we may write that

$$(12.8) \quad \begin{aligned} \sum_{n \leq x} \frac{1}{P(n)} &= (1 + o(1))x\delta(x) = \\ &= (1 + o(1)) \int_1^{\log x / \log 2} f(x, u) du \quad (x \rightarrow \infty). \end{aligned}$$

Now, the first graph below shows the curves $f(x, u)$ corresponding to $x = 10^6, 10^7$ and 10^8 . The same for the second graph, this time for $x = 10^{30}, 10^{31}$ and 10^{32} . For each of these curves, the corresponding number $x\delta(x)$ is equal to the area under the curve $f(x, u)$ from $u = 1$ to $u = \log x / \log 2$.



One will notice that most of the points under the two families of curves have their abscissa located within a very narrow interval (in fact, the intervals $[1, 6]$ and $[1, 8]$, respectively), which explains why the corresponding integral in (12.8) converges very rapidly.

Comparing the data from the two approaches, namely the one using the approximate value of $\delta(x)$ given by (12.2) and the other one using the exact value of $\delta(x)$ given by

$$(12.9) \quad \delta(x) = \log x \int_1^{\log x / \log 2} \frac{\rho(u)}{u^2} x^{-1/u} du,$$

we obtain the following table.

x	$G_2(x)$ real value	$G_2(x)$ predicted value using (12.2)	$G_2(x)$ predicted value using (12.9)
10^7	44 947	524 371	44 162
10^{10}	6 298 636	51 731 829	6 534 731
10^{13}	1 052 806 859	6 852 450 401	1 120 107 017

Table 4: The predicted values of $G_2(x)$ using two different approaches

As one will notice, using the exact value of $\delta(x)$ given by (12.9) (rather than the one given by (12.2)) provides a better approximation of $G_2(x)$. Setting

$$L_1(x) := \left(\log x \sqrt{\frac{\log x}{2} (\log \log x + \log \log \log x - \log 2)} \cdot I(x) \right)^{-1},$$

we obtain

$$(12.10) \quad G_2(x) = (1 + o(1)) \frac{x}{L_1(x)} \quad (x \rightarrow \infty),$$

an expression which will allow for appropriate approximations for the counting functions for each of the sets $E_{k,2}$ for $k = 2, 3, 4, 5$.

13. The counting functions of $E_{k,2}$

In Section 4, we provided lower and upper bounds for the counting function of the set $E_{2,2}$. We could not obtain the exact order of $E_{2,2}(x)$ as we could only prove that, for some positive constant c ,

$$\frac{x^{1/4}}{\log x} \ll E_{2,2}(x) \ll x \exp\{-c\sqrt{2 \log x \log \log x}\},$$

which seems to indicate that finding an asymptotic expression for $E_{2,2}(x)$ may be a very difficult problem. Nevertheless, what can one expect regarding the size of $E_{k,2}(x)$ for a fixed integer $k \geq 2$? In light of the asymptotic value of $G_2(x)$ revealed by relation (12.10), one can argue heuristically that the probability that an integer n belongs to G_2 is $1/L(n)$ and moreover since we can assume that the two events “ n belongs to G_2 ” and “ $n+1$ belongs to G_2 ” are independent events, then it is reasonable to pretend that the probability that n and $n+1$ both belong to G_2 is $1/L(n)^2$. From this reasoning, we deduce that

one can expect that the true order of $E_{2,2}(N)$ is $\sum_{n=2}^N \frac{1}{L(n)^2}$. Approximating this sum by an integral and since $L(x)$ is a slowly increasing function, we can use (3.1) to conclude that

$$\sum_{n=2}^N \frac{1}{L(n)^2} = (1 + o(1)) \int_2^N \frac{dt}{L(t)^2} = (1 + o(1)) \frac{N}{L(N)^2} \quad (N \rightarrow \infty).$$

thus implying that, as x becomes large,

$$(13.1) \quad E_{2,2}(x) \approx \frac{x}{L(x)^2}.$$

The above argument is in tune with the following more general heuristic result.

Proposition 2. (Heuristic) *Let A be an infinite set of positive integers, where the events “ $n \in A$ ” and “ $n+1 \in A$ ” are independent events. Assume that its counting function $A(x) := \#\{n \leq x : n \in A\}$ satisfies the estimate*

$$A(x) = (1 + o(1)) \frac{x}{L(x)} \quad (x \rightarrow \infty),$$

where $L(x)$ is a slowly increasing function. Given a fixed integer $k \geq 2$, let

$$A_k(x) := \#\{n \leq x : n, n+1, \dots, n+k-1 \in A\}.$$

Then,

$$(13.2) \quad A_k(x) \asymp \frac{x}{L(x)^k} \quad (x \rightarrow \infty).$$

Choosing $k = 2$ in Proposition 2 and setting $F_2(x) := x/L_1(x)^2$, we can expect to have

$$(13.3) \quad E_{2,2}(x) \asymp F_2(x),$$

which reveals the following table of values.

r	6	8	10	12	14
$a = E_{2,2}(10^r)$	79	549	4 231	36 426	358 582
$b = \lfloor F_2(10^r) \rfloor$	75	526	4 270	39 509	408 900
b/a	0.9494	0.9581	1.0092	1.0846	1.1403

In the above table, by $\lfloor c \rfloor$, we mean $\lfloor c + 1/2 \rfloor$, that is, the closest integer to c .

In the case $k = 2$, one can go a step further with a heuristic argument. Indeed, as we choose at random two integers and require that they each belong to G_2 , we must also require that they be consecutive. For this last condition to hold, the two integers must be coprime. Now it is well-known that the probability that two positive integers chosen at random are coprime is $6/\pi^2$. In other words, it seems reasonable to conjecture that relation (13.3) can be replaced by

$$E_{2,2}(x) = (1 + o(1)) \frac{6}{\pi^2} F_2(x) \quad (x \rightarrow \infty),$$

in which case we obtain the comparison table

k	6	8	10	12	14
$a = E_{2,2}(10^k)$	79	549	4 231	36 426	358 582
$b = \lfloor \frac{6}{\pi^2} F_2(10^k) \rfloor$	46	320	2 596	24 018	248 581
b/a	0.582278	0.582878	0.613567	0.659364	0.693233

which one may argue is more realistic and accurate.

Choosing $k = 3$ and $k = 4$ in Proposition 2, one can expect to have

$$E_{3,2}(x) \asymp \frac{x}{L_1(x)^3}$$

and

$$E_{4,2}(x) \asymp \frac{x}{L_1(x)^4}.$$

Using the (incomplete) data from Table 1, we find that $E_{3,2}(10^{21}) \geq 3711$, which is “getting closer” to the predicted value

$$\left\lfloor \frac{10^{21}}{L_1(10^{21})^3} \right\rfloor = 4928.$$

On the other hand, setting $F_4(x) := \frac{x}{L_1(x)^4}$, we obtain the table

r	30	31	32	33
$F_4(10^r)$	0.616303	1.11522	2.05816	3.87101

r	34	35	36
$F_4(10^r)$	7.41463	14.4541	28.6588

which indicates that one can expect the smallest element of $E_{4,2}$ to have fewer than 40 digits.

Finally, setting $F_5(x) := \frac{x}{L_1(x)^5}$, we obtain

r	67	68	69	70
$F_5(10^r)$	0.875797	1.79317	3.70491	7.72323

r	71	72	73
$F_5(10^r)$	16.241	34.4466	73.6777

which indicates that one can expect the smallest element of $E_{5,2}$ to have around 70 digits.

14. Open problems

Open problems regarding consecutive integers divisible by a power of their largest prime factor are numerous. We list a few.

14.1. The case of squares amongst $n, n+1, n+2$ for $n \in E_{3,2}$

In the search for numbers belonging to $E_{3,2}$, one may be tempted to start by examining the perfect squares in the hope that the numbers near them would also belong to G_2 . Indeed, that does occur. For instance, we found six such numbers:

$n \in E_{3,2}$	$\rightarrow n + j = m^2$
20 944 167 840	$n + 1 = (17 \cdot 8\,513)^2$
23 153 476 981 634	$n + 2 = (2 \cdot 67 \cdot 149 \cdot 241)^2$
20 509 990 465 165 715 808	$n + 1 = (10\,399 \cdot 435\,503)^2$
57 615 604 841 725 651 874	$n + 2 = (2 \cdot 61 \cdot 1\,823 \cdot 34\,129)^2$
1 940 866 749 835 432 563 168	$n + 1 = (593 \cdot 983 \cdot 75\,577)^2$
1 202 552 397 955 428 157 263 374	$n + 2 = (2 \cdot 274\,152\,375\,281)^2$

Table 5: Integers $n \in E_{3,2}$ with $n + j = m^2$ for some $j \in \{0, 1, 2\}$

Are there any other members of $E_{3,2}$ with that same property? Most likely! Infinitely many? Difficult to say!

What if we were to somewhat relax the condition and only require that for some $n \in E_{3,2}$, one of the integers $n, n+1, n+2$ is a powerful number (also called “squarefull number”)? Recall that a positive integer n is called *powerful* if $p \mid n$

implies that $p^2 \mid n$. Since it is known that $\#\{n \leq x : n \text{ is powerful}\} = c_2 \sqrt{x} + O(x^{1/3})$, where $c_2 = \zeta(3/2)/\zeta(3) \approx 2.173$ (see for instance Ivić and Shiu [14]), one can expect the occurrence of numbers $n \in E_{3,2}$ with $n+j$ powerful ($n+j$ not a square) for some $j \in \{0, 1, 2\}$. Indeed, we did find three more numbers with that property, namely the number $n = 11\,587\,121\,710 \in E_{3,2}$ for which we have $n+2 = 2^4 \cdot 3^3 \cdot 5179^2$, the number $n = 1\,589\,922\,788\,612\,140\,124 \in E_{3,2}$ for which we have

$$n+1 = 3^2 \cdot 5^3 \cdot 11^2 \cdot 13^2 \cdot 151^2 \cdot 1741^2$$

and finally the number

$$n = 282\,671\,477\,408\,473\,458\,447\,091\,665\,772\,991\,620\,124 \in E_{3,2}$$

for which we have

$$n+1 = 3^{10} \cdot 5^3 \cdot 11^2 \cdot 17^2 \cdot 401^2 \cdot 82\,526\,558\,591^2.$$

Finding other numbers with that property will certainly be a challenge.

14.2. The occurrence of cubes

Given k integers $\ell_0, \ell_1, \dots, \ell_{k-1}$, each ≥ 2 , consider the set

$$F(\ell_0, \ell_1, \dots, \ell_{k-1}) := \{n \in \mathbb{N} : P(n+i)^{\ell_i} \mid n+i \text{ for } i = 0, 1, \dots, k-1\},$$

so that in particular $E_{k,\ell} = F(\underbrace{\ell, \dots, \ell}_k)$. Most likely, each set $F(\ell_0, \dots, \ell_{k-1})$ is infinite, but besides the set $F(2, 2) = E_{2,2}$, no such statement has been proved.

Now, consider the sets $F(3, 2, 2)$, $F(2, 3, 2)$ and $F(2, 2, 3)$. We found 38 numbers with 30 digits or less belonging to either one of these three sets. The first 24 of these (with 22 digits or less) are listed in the first author's web site [3].

What is the smallest number n_0 belonging to exactly two of the sets $F(3, 2, 2)$, $F(2, 3, 2)$ and $F(2, 2, 3)$? This number n_0 is clearly smaller than the numbers n_1 and n_2 mentioned in Section 6, but how much smaller?

14.3. Elements of G_ℓ in arithmetic progressions

Another natural extension to the study of the sets $E_{k,\ell}$ is the following. Instead of searching for consecutive integers n such that $P(n+j)^\ell \mid n+j$ for $j = 0, 1, \dots, k-1$, one could search for integers in arithmetic progression belonging to G_ℓ . For instance, integers $n, n+d, \dots, n+(k-1)d$ such that $P(n+jd)^\ell \mid n+jd$ for $j = 0, 1, \dots, k-1$ for some $d \geq 2$. Of course, given any fixed integer $d \geq 2$, one can associate to any $n \in E_{k,\ell}$ the k -tuples $(n, n+d, \dots, n+(k-1)d)$ for which we have that $P(n+jd)^\ell \mid n+jd$ for $j = 0, 1, \dots, k-1$. Moreover, it is easy to prove the following.

Proposition 3. *Given arbitrary integers $k \geq 2$ and $\ell \geq 2$, there exists an arithmetic progression $n, n + d, \dots, n + (k - 1)d$, where $P(n + jd)^\ell \mid n + jd$, $j = 0, 1, \dots, k - 1$, for some positive integers n and d .*

Proof. Let p be the smallest prime such that $p \geq k - 1$ and consider the numbers

$$1 \cdot p^\ell, 2 \cdot p^\ell, \dots, (p + 1) \cdot p^\ell.$$

Each of the above numbers belongs to G_ℓ and they are in an arithmetic progression with common difference p^ℓ . ■

Given a fixed integer $\ell \geq 2$, the real challenge is to find a fixed integer $d \geq 2$ for which there exist an infinite sequence of integers n_1, n_2, \dots such that $n_i, n_i + d, n_i + 2d, \dots, n_i + (k - 1)d$ all belong to G_ℓ .

Acknowledgments. The authors are grateful to the referee for the careful reading of the paper and for helpful suggestions. The work of the first author was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

- [1] **Bunyakovsky, V.**, Nouveaux théorèmes relatifs à la distinction des nombres premiers et à la décomposition des entiers en facteurs, *Mém. Acad. Sc. St. Pétersbourg*, **6** (1857), 305–329.
- [2] **De Koninck, J.-M.**, *Those Fascinating Numbers*, American Mathematical Society, Providence, RI, 2009, xviii+426 pp.
<https://doi.org/10.1090/mbk/064>
- [3] **De Koninck, J.-M.**,
www.jeanmariedekoninck.mat.ulaval.ca/documents.
- [4] **De Koninck, J.-M. and N. Doyon**, *The Life of Primes in 37 Episodes*, American Mathematical Society, Providence, RI, 2021] xiv+329 pp.
<https://doi.org/10.1090/mbk/139>
- [5] **De Koninck, J.-M., N. Doyon and F. Luca**, Consecutive integers divisible by the square of their largest prime factors, *Journal of Combinatorics and Number Theory*, **5** (2013), No.2 (June), 81–93.
- [6] **De Koninck, J.-M. and F. Luca**, *Analytic Number Theory: Exploring the Anatomy of Integers*, Graduate Studies in Mathematics, Vol. 134, American Mathematical Society, Providence, RI, 2012, xviii+414 pp. 2012.
<https://doi.org/10.1090/gsm/134>
- [7] **De Koninck, J.-M. and M. Moineau**, Consecutive integers divisible by a power of their largest prime factor, *Journal of Integer Sequences*, **21** (2018), no. 9, Article 18.9.3, 16 pp.

- [8] **de la Bretèche, R. et S. Drappeau**, Niveau de répartition des polynômes quadratiques et crible majorant pour les entiers friables [Level of distribution of quadratic polynomials and upper-bound sieve for friable integers], *J. Eur. Math. Soc. (JEMS)*, **22** (2020), no. 5, 1577–1624.
<https://doi.org/10.4171/jems/951>
- [9] **Dilcher, K. and M. Ulas**, Arithmetic properties of polynomial solutions of the Diophantine equation $P(x)x^{n+1}+Q(x)(x+1)^{n+1} = 1$, *Period. Math. Hungar.*, **84(1)** (2022), 89–118.
<https://doi.org/10.1007/s10998-020-00376-5>
- [10] **Erdős, P., A. Ivić and C. Pomerance**, On sums involving reciprocals of the largest prime factor of an integer, *Glas. Mat. Ser. III*, **21(2)**(41) (1986), 283–300.
- [11] **Hildebrand, A.**, On the number of positive integers $\leq x$ and free of prime factors $> y$, *J. Number Theory*, **22(3)** (1986), 289–307.
[https://doi.org/10.1016/0022-314X\(86\)90013-2](https://doi.org/10.1016/0022-314X(86)90013-2)
- [12] **Ivić, A.**, On certain large additive functions, in: *Paul Erdős and his mathematics*, I (Budapest, 1999), 319–331, Bolyai Soc. Math. Stud., 11, János Bolyai Math. Soc., Budapest, 2002.
- [13] **Ivić, A. and C. Pomerance**, Estimates for certain sums involving the largest prime factor of an integer, in: *Topics in classical number theory*, Vol. I, II (Budapest, 1981), 769–789, Colloq. Math. Soc. Janos Bolyai, 34, North-Holland, Amsterdam, 1984.
- [14] **Ivić, A. and P. Shiu**, The distribution of powerful numbers, *Illinois J. Math.*, **26** (1982), 576–590.
<https://doi.org/10.1215/ijm/1256046597>
- [15] **Martin, G.**, An asymptotic formula for the number of smooth values of a polynomial, *J. Number Theory*, **93** (2002), 108–182.
<https://doi.org/10.1006/jnth.2001.2722>
- [16] **Mason, R.C.**, *Diophantine Equations over Functions Fields*, Cambridge, England: Cambridge University Press, 1984.
<https://doi.org/10.1017/CB09780511752490>
- [17] **Seneta, E.**, *Regularly Varying Functions*, Springer-Verlag, Berlin, 1976.
<https://doi.org/10.1007/BFb0079658>
- [18] **Silverman, J.**, *A Friendly Introduction to Number Theory*, 4th Edition, Pearson, 2012.
- [19] **Snyder, N.**, An alternate proof of Mason’s theorem, *Elem. Math.*, **55(3)** (2000), 93–94.
<https://doi.org/10.1007/s000170050074>
- [20] **Tenenbaum, G.**, *Introduction à la théorie analytique et probabiliste des nombres*, Dunod, 2022.

Jean-Marie De Koninck

<https://orcid.org/0000-0003-4506-959X>

Département de mathématiques

Université Laval

Québec G1V 0A6

Canada

`jmdk@mat.ulaval.ca`

Matthieu Moineau

Interstellar Lab

10 Rue de Penthièvre

Paris 75008

France

`moineau53.moineau@gmail.com`