IJMMS 2003:19, 1185–1192
PII. S016117120320418X
http://ijmms.hindawi.com

ON A THIN SET OF INTEGERS INVOLVING THE LARGEST PRIME FACTOR FUNCTION

JEAN-MARIE DE KONINCK and NICOLAS DOYON

Received 22 April 2002

For each integer $n \geq 2$, let $P(n)$ denote its largest prime factor. Let $S := \{n \geq 2 : n \text{ does not divide } P(n)!\}$ and $S(x) := \#\{n \leq x : n \in S\}$. Erdős (1991) conjectured that S is a set of zero density. This was proved by Kastanas (1994) who established that $S(x) = O(x/\log x)$. Recently, Akbik (1999) proved that $S(x) = O(x\exp\{-(1/4)\sqrt{\log x}\})$. In this paper, we show that $S(x) = x\exp\{-(2+o(1))\times\sqrt{\log x\log\log x}\}$. We also investigate small and large gaps among the elements of S and state some conjectures.

2000 Mathematics Subject Classification: 11B05, 11N25.

1. Introduction. For each integer $n \geq 2$, let $P(n)$ denote its largest prime factor and let

$$S := \{n \geq 2 : n \text{ does not divide } P(n)!\}, \qquad S(x) := \#\{n \leq x : n \in S\}. \tag{1.1}$$

Thus, the first 25 elements of S are

$$\begin{gathered} 4,8,9,12,16,18,24,25,27,32,36,45,48,49, \\ 50,54,64,72,75,80,81,90,96,98,100, \end{gathered} \tag{1.2}$$

while using a computer, we easily obtain that $S(10) = 3$, $S(100) = 25$, $S(1000) = 127$, $S(10^4) = 593$, $S(10^5) = 2806$, $S(10^6) = 13567$, $S(10^7) = 67252$, and $S(10^8) = 342022$.

In 1991, Erdős [2] challenged his readers to prove that S is a set of zero density. In 1994, Kastanas [4] proved that result, while K. Ford (see [4]) observed that $S(x) = O(x/\log x)$. In 1999, Akbik [1] proved that $S(x) = O(x\exp\{-(1/4)\times\sqrt{\log x}\})$.

Our main goal here is to prove that

$$S(x) = x\exp\left\{-(2+o(1))\sqrt{\log x\log\log x}\right\}. \tag{1.3}$$

In order to prove (1.3), we establish the following two bounds valid for each fixed $\delta > 0$:

$$S(x) \gg x\exp\left\{-2(1+\delta)\sqrt{\log x \log\log x}\right\}, \tag{1.4}$$

$$S(x) \ll x\exp\left\{-2(1-\delta)\sqrt{\log x \log\log x}\right\}. \tag{1.5}$$

Finally, we investigate small and large gaps among the elements of S and state some conjectures.

2. The lower bound for $S(x)$. Let $\delta > 0$ be small and fixed. Since every integer $n \ge 2$ divisible by the square of its largest prime factor must belong to S, we have that

$$S(x) \ge \sum_{p\le\sqrt{x}} \sum_{\substack{mp^2\le x\\ P(m)\le p}} 1 = \sum_{p\le\sqrt{x}} \sum_{\substack{m\le x/p^2\\ P(m)\le p}} 1 = \sum_{p\le\sqrt{x}} \Psi\left(\frac{x}{p^2},p\right), \tag{2.1}$$

where $\Psi(x,y) := \#\{n \le x : P(n) \le y\}$.

Setting $u = \log x/\log y$, we recall Hildebrand's estimate [3]

$$\Psi(x,y) = x\rho(u)\left\{1+O\left(\frac{\log(u+1)}{\log y}\right)\right\} \tag{2.2}$$

which holds for

$$\exp\left\{(\log\log x)^{5/3+\varepsilon}\right\} \le y \le x, \tag{2.3}$$

where $\varepsilon > 0$ is any fixed real number, and where ρ stands for Dickman's function whose asymptotic behaviour is given by

$$\rho(u) = \exp\left\{-u\left(\log u + \log\log u - 1 + O\left(\frac{\log\log u}{\log u}\right)\right)\right\} \quad (u \longrightarrow \infty). \tag{2.4}$$

It follows from this last estimate that if u is sufficiently large, then

$$\log\rho(u) \ge -(1+\delta)u\log u. \tag{2.5}$$

Hence, if we choose r sufficiently large, say $r \ge r_0 \ge 2$, then for each $y \le x^{1/r}$, we have $u = \log x/\log y \ge r$, thereby guaranteeing the validity of (2.5).

Therefore, it follows from (2.4) and (2.5) that, with $u = \log(x/p^2)/\log p = \log x/\log p - 2$,

$$\log\rho(u) \geq -(1+\delta)\frac{\log x}{\log p}\log\log x \quad (u \geq r_0) \tag{2.6}$$

and hence (2.1) and (2.2) yield

$$\begin{aligned} S(x) &\gg x \sum_{e^{(\log\log x)^{5/3+\varepsilon}} \leq p \leq x^{1/r}} \frac{1}{p^2 e^{(1+\delta)(\log x/\log p)\log\log x}} \\ &= x\int_{e^{(\log\log x)^{5/3+\varepsilon}}}^{x^{1/r}} \frac{d\pi(t)}{t^2 \cdot e^{(1+\delta)(\log x/\log t)\log\log x}}, \end{aligned} \tag{2.7}$$

where $\pi(t)$ stands for the number of primes not exceeding t. Now, set

$$L_\delta(x) := \sqrt{(1+\delta)\log x\log\log x} \quad (x \geq 3) \tag{2.8}$$

so that, for any $\delta_1 > 0$, we have, for x sufficiently large,

$$[L_\delta(x), (1+\delta_1)L_\delta(x)] \subset \Big[(\log\log x)^{5/3+\varepsilon}, \frac{1}{r}\log x\Big]. \tag{2.9}$$

Using this, it follows from (2.7) that setting $J(x) := [e^{L_\delta(x)}, e^{(1+\delta_1)L_\delta(x)}]$,

$$\begin{aligned} S(x) &\gg x\int_{t\in J(x)} \frac{d\pi(t)}{t^2 \cdot e^{(1+\delta)(\log x/\log t)\log\log x}} \\ &> x \min_{t\in J(x)}\left(\frac{1}{t^2 \cdot e^{(1+\delta)(\log x/\log t)\log\log x}}\right)\int_{t\in J(x)} d\pi(t). \end{aligned} \tag{2.10}$$

Now, observe that since $t/\log t < \pi(t) < 2(t/\log t)$ for $t \geq 11$, we have that

$$\begin{aligned} \int_{t\in J(x)} d\pi(t) &= \pi(e^{(1+\delta_1)L_\delta(x)}) - \pi(e^{L_\delta(x)}) \\ &> \frac{e^{(1+\delta_1)L_\delta(x)}}{(1+\delta_1)L_\delta(x)} - \frac{e^{L_\delta(x)}}{L_\delta(x)} \\ &\gg \frac{e^{(1+\delta_1)L_\delta(x)}}{(1+\delta_1)L_\delta(x)}. \end{aligned} \tag{2.11}$$

On the other hand, setting $v = \log t$ and afterwards $w = v/L_\delta(x)$, we have

$$\begin{aligned}
\min_{t\in J(x)} &\left(\frac{1}{t^2\cdot e^{(1+\delta)(\log x/\log t)\log\log x}}\right)\\
&= \min_{L_\delta(x)\le v\le(1+\delta_1)L_\delta(x)}\left(\frac{1}{e^{2v+(1+\delta)(\log x/v)\log\log x}}\right)\\
&= \min_{1\le w\le 1+\delta_1}\left(\frac{1}{e^{2wL_\delta(x)+(1+\delta)(\log x/wL_\delta(x))\log\log x}}\right)\\
&= \min_{1\le w\le 1+\delta_1}\left(\frac{1}{e^{(2w+1/w)L_\delta(x)}}\right)\\
&\gg \frac{1}{e^{(3+2\delta_1)L_\delta(x)}}
\end{aligned}\tag{2.12}$$

since $2w+1/w \le 2+2\delta_1+1 = 3+2\delta_1$ for each $w\in[1,1+\delta_1]$.

Hence, using (2.11) and (2.12), it follows from (2.10) that

$$\begin{aligned}
S(x) &\gg x\frac{e^{(1+\delta_1)L_\delta(x)}}{(1+\delta_1)L_\delta(x)}\cdot\frac{1}{e^{(3+2\delta_1)L_\delta(x)}}\\
&= x\frac{e^{-(2+\delta_1)L_\delta(x)}}{(1+\delta_1)L_\delta(x)}\\
&\gg xe^{-2(1+\delta_1)L_\delta(x)},
\end{aligned}\tag{2.13}$$

which establishes (1.4) by taking δ_1 sufficiently small.

3. The upper bound for $S(x)$. First, we establish that

$$S(x) < \sum_{2\le r<\log x/\log 2}\ \sum_{p<x^{1/r}}\Psi\left(\frac{x}{p^r}, pr\right). \tag{3.1}$$

Actually, this inequality is based on a very simple observation; namely, the fact that if $n\in S$, then there exist a prime p and an integer $r\ge 2$ such that p^r divides n but does not divide $P(n)!$, in which case $P(n) < pr$. Hence, writing $n = p^r m$, we have that $P(m)\le P(n) < pr$. These conditions imply that if $n\in S$ and $n\le x$, then we have $r<\log x/\log 2$, $p<x^{1/r}$, $m<x/p^r$, and $P(m)<pr$, thus proving (3.1).

We now move to find an upper bound for the inner sum on the right-hand side of (3.1); namely, $\sum_{p<x^{1/r}}\Psi(x/p^r, pr)$, uniformly for all $r\ge 2$. For this purpose, we fix $r\ge 2$ and separate this sum on p into three distinct sums as follows:

$$\sum_{p<x^{1/r}}\Psi\left(\frac{x}{p^r}, pr\right) = S_1(x)+S_2(x)+S_3(x), \tag{3.2}$$

where the sums $S_1(x)$, $S_2(x)$, and $S_3(x)$ run, respectively, in the following ranges:

$$\begin{gathered} p \le \exp\{(\log\log x)^2\}, \\ \exp\{(\log\log x)^2\} < p \le \exp\left\{2\sqrt{\log x \log\log x}\right\}, \\ \exp\left\{2\sqrt{\log x \log\log x}\right\} < p < x^{1/r}. \end{gathered} \tag{3.3}$$

The first sum is negligible since it is clear that, using the well-known estimate,

$$\Psi(X,Y) \ll Xe^{-(1/2)\log X/\log Y} \quad (X \ge Y \ge 2) \tag{3.4}$$

(see, e.g., Tenenbaum [5, Chapter III.5, Theorem 1]), we get that

$$\begin{aligned} S_1(x) &< \exp\{(\log\log x)^2\}\Psi\left(x, \frac{\log x}{\log 2}\exp\{(\log\log x)^2\}\right) \\ &\ll xe^{(-1/2+o(1))(\log x/(\log\log x)^2)}. \end{aligned} \tag{3.5}$$

The third one is also easily bounded since

$$\begin{aligned} S_3(x) &< \sum_{\exp\{2\sqrt{\log x \log\log x}\}<p<x^{1/r}} \frac{x}{p^r} \\ &\ll x \sum_{p>\exp\{2\sqrt{\log x\log\log x}\}} \frac{1}{p^2} \\ &\ll x\exp\left\{-2\sqrt{\log x\log\log x}\right\}. \end{aligned} \tag{3.6}$$

To estimate $S_2(x)$, we use essentially the same technique as in the proof of (1.4).

First, it follows from (2.4) that

$$\log\rho(u) \le -u\log(u) \tag{3.7}$$

provided u is sufficiently large. Then, with the same approach as in the proof of (1.4), we get that, for each fixed integer $r \ge 2$,

$$S_2(x) \ll x\int_1^{2\sqrt{\log x\log\log x}} \frac{dv}{v^{r-1}e^{v+\log x\log\log x/v}}. \tag{3.8}$$

Now, set $f(v) = v + \log x \log\log x / v$. Since $f'(v) = 1 - \log x \log\log x / v^2$ and $f'(v) = 0$ when $v = v_0 = \sqrt{\log x \log\log x}$, it is easy to see that v_0 is indeed a minimum for f. From this, it follows that

$$\begin{aligned} v + \frac{\log x \log\log x}{v} & \\ & \geq f(v_0) = 2\sqrt{\log x \log\log x} \quad \text{for each } v \in \left[1, 2\sqrt{\log x \log\log x}\right]. \end{aligned} \tag{3.9}$$

Using this in (3.8), we conclude that

$$\begin{aligned} S_2(x) & \ll x \exp\left\{-2\sqrt{\log x \log\log x}\right\} \int_1^{2\sqrt{\log x \log\log x}} \frac{dv}{v^{r-1}} \\ & \ll x \log\left(2\sqrt{\log x \log\log x}\right) \exp\left\{-2\sqrt{\log x \log\log x}\right\}. \end{aligned} \tag{3.10}$$

Combining (3.1), (3.2), (3.5), (3.6), and (3.10), we get (1.5).

4. Small and large gaps among elements of S. We can easily show that there are infinitely many $n \in S$ such that $n+1 \in S$. This follows from the fact that the Pell equation

$$x^2 - 2y^2 = 1 \tag{4.1}$$

has infinitely many solutions. Indeed, if (x,y) is a solution of (4.1), then by setting $n = 2y^2$ and $n+1 = x^2$, we have that $P(n)^2 \mid n$ and $P(n+1)^2 \mid (n+1)$, in which case n does not divide $P(n)!$ and $n+1$ does not divide $P(n+1)!$, which guarantees that $n, n+1 \in S$. In fact, if T_2 stands for the set of those $n \in S$ such that $n+1 \in S$ and if $T_2(x) = \#\{n \leq x : n \in T_2\}$, then it follows easily from the above that $T_2(x) \gg \log x$. In fact, most certainly, the true order of $T_2(x)$ is much larger than $\log x$, but we could not prove it.

It seems strange that such *twin elements* of S, that is, pairs of numbers n and $n+1$ both in S, are more difficult to count than pairs of numbers n and $n+4$ both in S. Indeed, if F_4 stands for the set of those $n \in S$ such that $n+4 \in S$ and if $F_4(x) = \#\{n \leq x : n \in F_4\}$, then we can show that

$$F_4(x) \gg \frac{x^{1/4}}{\log x}. \tag{4.2}$$

Indeed, observe that given any prime p, then both numbers $n = p^4 - 4p^2 = p^2(p^2-4) = p^2(p-2)(p+2)$ and $n+4 = p^4 - 4p^2 + 4 = (p^2-2)^2$ belong to S. Since there are at least $\pi(x^{1/4})$ such pairs up to x, estimate (4.2) follows from

Chebychev's inequality $\pi(y) \gg y/\log y$. Finally, note that $T_2(10^8) = 1175$, while $F_4(10^8) = 1261$.

More generally, we conjecture that given any positive $k \geq 3$, the set $T_k := \{n \in S : n+1,\ n+2,\dots,n+k-1 \in S\}$ is also an infinite set. We could not prove this to be true, even in the case where $k = 3$. Note that the only numbers less than 10^8 belonging to T_3 are 48, 118579, 629693, 1294298, 9841094, and 40692424.

As for large gaps among consecutive elements of S, it follows from the fact that S is a set of zero density that given any positive integer k, there are infinitely many integers n such that the intervals $[n, n+k]$ contain no element of S. Table 4.1 gives, for each positive integer k, the smallest integer $n = n(k) \in S$ such that both n and $n+100k$ belong to S, while the open interval $(n, n+100k)$ contains no element of S.

TABLE 4.1

$100k$	$n = n(k)$	$100k$	$n = n(k)$
100	21025	600	738606
200	78408	700	946832
300	369303	800	8000325
400	1250256	900	5382888
500	1639078	1000	5775000

It is quite easy to show that

$$n(k) \geq 2500k^2 - 100k + 1. \tag{4.3}$$

Indeed, since all perfect squares belong to S and since $(m+1)^2 - m^2 = 2m + 1$, it follows that the interval $(n, n+2m+1)$ contains no element of S and, therefore, that $n \geq m^2$. Hence, given a positive integer k, choose m so that $100k = 2m+2$, that is, $m = 50k - 1$. Then, clearly, we have that $n(k) \geq m^2 = (50k-1)^2$, which proves (4.3).

It would also be interesting to obtain a decent upper bound for $n(k)$.

ACKNOWLEDGMENT. This research was supported in part by a grant from the Natural Sciences and Engineering Research Council (NSERC).

REFERENCES

[1] S. Akbik, *On a density problem of Erdős*, Int. J. Math. Math. Sci. **22** (1999), no. 3, 655–658.
[2] P. Erdős, *Problem 6674*, Amer. Math. Monthly **98** (1991), no. 10, 965.
[3] A. Hildebrand, *On the number of positive integers* $\leq x$ *and free of prime factors* $> y$, J. Number Theory **22** (1986), no. 3, 289–307.
[4] I. Kastanas, *Solution to Problem 6674*, Amer. Math. Monthly **101** (1994), no. 2, 179.

[5] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Studies in Advanced Mathematics, vol. 46, Cambridge University Press, Cambridge, 1995.

Jean-Marie De Koninck: Département de Mathématiques et de Statistique, Université Laval, Québec, Québec, Canada G1K 7P4
E-mail address: jmdk@mat.ulaval.ca

Nicolas Doyon: Département de Mathématiques et de Statistique, Université de Montréal, Montréal, Québec, Canada H3C 3J7
E-mail address: doyon@dms.umontreal.ca