COMPOSITE POSITIVE INTEGERS WITH AN AVERAGE PRIME FACTOR

FLORIAN LUCA AND FRANCESCO PAPPALARDI

Abstract. We study the distribution of the positive integers n which are composite and whose average prime divisor is an integer and a prime divisor of n.

Let $p(n)$ denote the average prime divisor of an integer n. That is,

$$p(n) = \frac{1}{\omega(n)} \sum_{\substack{p \text{ prime} \\ p|n}} p,$$

where $\omega(n)$ denotes the number of distinct prime divisors of n.

It is clear that if n is a prime power, then $p(n) \mid n$. In this paper we consider the set

$$\mathcal{A} = \{n \ : \ \omega(n) > 1, \ p(n) \in \mathbb{N}, p(n) \mid n \ \text{ and } p(n) \ \text{ is prime}\}.$$

It is obvious that $n \in \mathcal{A}$ if and only if the square-free part of n is in $\mathcal{A}$.

The first few square-free elements of $\mathcal{A}$ are: 105, 231, 627, 897, 935, 1365, 1581, 1729, 2465, 2967, 4123, 4301, 4715, 5313, 5487, 6045, 7293, 7685, 7881, 7917, 9717, 10707, 10965, 11339, 12597, 14637, 14993, 16377, 16445, 17353, 18753, 20213, 20757, 20915, 21045, 23779, 25327, 26331, 26765, 26961, 28101, 28497, 29341, 29607.

It is clear that $\mathcal{A}$ contains only odd numbers. Here, we prove the following result:

Theorem 1. *Let $\mathcal{A}(x) := \mathcal{A} \cap [1, x]$. The estimates*

$$\frac{x}{\exp\left((2 + o(1))\sqrt{\log x \log\log x}\right)} \le \#\mathcal{A}(x) \le \frac{x}{\exp\left((\frac{1}{\sqrt{2}} + o(1))\sqrt{\log x \log\log x}\right)}$$

hold as $x \to \infty$.

Since the counting function of the prime powers $n < x$ which are not primes is $O(\sqrt{x}/\log x)$, it follows that the same result is valid if we enlarge $\mathcal{A}$ to be the set of all composite integers n whose average prime factor is an integer and is a prime factor of n.

Our theorem complements the results from [1], where several results concerning the function $p(n)$ were obtained, such as the uniform distribution of the fractional parts $\{p(n)\}$ in the interval $[0, 1)$ when n ranges in the set of all positive integers, and the order of magnitude of the counting function of the set of positive integers n such that $p(n)$ is an integer.

Throughout, we use the Vinogradov symbols $\gg$ and $\ll$ and the Landau symbols O and o with their regular meanings. We use log for the natural logarithm and $\lfloor\ \rfloor$ for the 'integer part' function.

1991 *Mathematics Subject Classification.* 11N37 (primary), 11N60, 11P32 (secondary).

Proof of the upper-bound. Let us consider the following sets:

$$\mathcal{A}_1(x) = \{n \le x \mid P(n) < y\},$$

and

$$\mathcal{A}_2(x) = \left\{n \le x \mid n \notin \mathcal{A}_1(x), P(n)^2 \mid n\right\},$$

where y is a parameter which depends on x to be chosen later and which satisfies $\exp((\log\log x)^2) \le y \le x$, and $P(n)$ denotes the largest prime factor of n.

From standard estimates for smooth numbers [2], we know that if we set $u = \log x/\log y$, then

$$\#\mathcal{A}_1(x) \ll \frac{x}{\exp((1+o(1))u\log u)} \qquad (x \to \infty) \tag{1}$$

in our range for y versus x, while

$$\#\mathcal{A}_2(x) \le \sum_{\substack{p \text{ prime} \\ p \ge y}} \left\lfloor \frac{x}{p^2} \right\rfloor \le x \sum_{n \ge y} \frac{1}{n^2} \ll \frac{x}{y}. \tag{2}$$

Let $\mathcal{A}_3(x) = \mathcal{A}(x) \setminus (\mathcal{A}_1(x) \cup \mathcal{A}_2(x))$. If $n \in \mathcal{A}_3(x)$, then we can write $n = P(n)m$, where $m > 1$ (because $\omega(n) > 1$). Furthermore, since $n \notin \mathcal{A}_2(x)$, $P(n) \nmid m$, and $p(n) < P(n)$ since the average of at least 2 distinct integers is less then the maximum of the integers. Thus, the condition that $p(n)$ is prime and divides n implies that $p(n) \mid m$, and so we can write

$$p(n) = \frac{P(n) + \sum_{q \mid m} q}{\omega(m) + 1},$$

which, solving for $P(n)$, gives

$$P(n) = p(n)(\omega(m)+1) - \sum_{q \mid m} q.$$

Hence, $P(n)$ is uniquely determined by $p(n)$ and by m. But since $p(n)$ is a prime divisor of m, it follows that for any fixed value of m, there are at most $\omega(m)$ possible values of $P(n)$. Furthermore, note that for the positive integers n under consideration, we have that $P(n) \ge y$, therefore $m \le x/y$, so

$$\#\mathcal{A}_3(x) \le \sum_{m \le x/y} \omega(m) \ll \frac{x \log\log x}{y}, \tag{3}$$

where we used the well known fact that

$$\sum_{t \le x} \omega(t) \ll \log\log x.$$

From estimates for (1), (2) and (3), we immediately deduce that

$$\begin{aligned} \#\mathcal{A}(x) &\le \#\mathcal{A}_1(x) + \#\mathcal{A}_2(x) + \#\mathcal{A}_3(x) \\ &\ll \frac{x\log\log x}{y} + \frac{x}{\exp((1+o(1))u\log u)}. \end{aligned}$$

To minimize the right hand side above we choose $y = \exp(u \log u)$, which amounts to

$$\log^2 y = \log x \log\left(\frac{\log x}{\log y}\right).$$

Thus, we get that $y = (1 + o(1))\sqrt{\log x \log\log x}$ as $x \to \infty$, and with this choice of y versus x we obtain

$$\#\mathcal{A}(x) \ll \frac{x}{\exp\left((\frac{1}{\sqrt{2}} + o(1))\sqrt{\log x \log\log x}\right)}$$

as $x \to \infty$. □

Proof of the lower-bound. Let y be a parameter depending on x (different from the one from the proof of the upper bound) and k an even positive integer depending also on x, both tending to infinity with x which we will choose later. For the moment we assume that $k > 5$ and $y > k^4$. Suppose that $P,\ Q,\ p_1,\ \ldots,\ p_k$ are prime numbers which lie in the respective intervals:

$$P \in (y/2, y], \qquad Q \in (y/4, y/2], \qquad \text{and}\ \ p_1,\ \ldots,\ p_k \in (y/2k^2, y/k^2].$$

It is clear that all the above primes are distinct and odd. Furthermore, the integer

$$N = (k+4)Q - P - (p_1 + \cdots + p_k)$$

is odd, positive, and lies in the interval $(ky/4, ky]$. By Vinogradov's Three Primes Theorem [3], we have that the equation

$$N = q_1 + q_2 + q_3$$

admits $\gg N^2/\log^3 N$ solution in primes $q_1 < q_2 < q_3$ as $N \to \infty$. It is also clear that, at the cost of reducing the constant implied by the above $\gg$, we can assume that $q_1 > c_1 N$, where c_1 is some absolute positive constant, and that the three primes above are distinct. Note that with these choices, $\min\{q_1, q_2, q_3\} > c_1 ky/4 > k^3 y/4 > y$, therefore the primes $q_1,\ q_2$ and q_3 are different from $P,\ Q,\ p_1,\ \ldots,\ p_k$.

Consider the integer

$$n = p_1 \cdots p_k \cdot q_1 \cdot q_2 \cdot q_3 \cdot P \cdot Q.$$

We claim the $n \in \mathcal{A}$. Indeed, $\omega(n) = k + 5$, and

$$\frac{1}{k+5}\left(p_1 + \cdots + p_k + q_1 + q_2 + q_3 + P + Q\right) = Q$$

is a prime factor of n. We are therefore only left with the task of counting the number integers up to a fixed upper bound x which can be constructed by the above method with suitable choices of y and k versus x.

For given y and k, the number of choices for P, Q and $(p_1, \ldots, p_k)$ are respectively:

$$\pi(y) - \pi(y/2), \quad \pi(y/2) - \pi(y/4) \quad \text{and} \quad \binom{\pi(y/k^2) - \pi(y/2k^2)}{k}.$$

Therefore the number of possible n's, when $k^4 < y$ and k is large, is

$$\gg \frac{y}{2\log y} \cdot \frac{y}{4\log y} \cdot \left(\frac{y}{6k^3 \log(y/k^2)}\right)^k \cdot \frac{c_1(ky/4)^2}{(\log ky)^3}, \tag{4}$$

where in the above estimates we used the Prime Number Theorem and the fact that if $a > 2b$, then

$$\binom{a}{b} \gg \left(\frac{a-b}{b}\right)^b > \left(\frac{a}{2b}\right)^b$$

with the choices $a = \pi(y/k^2) - \pi(y/2k^2) > y/(3k^2 \log(y/k^2)) > 2k$ and $b = k$ (the first estimate above holds for large k by the Prime Number Theorem, while the second holds for large k by the fact that $y > k^4$).

A further calculation shows that the expression appearing at (4) above is

$$\gg \frac{y^{k+4}}{4^k k^{3k-3} (\log y)^{k+5}}. \tag{5}$$

We now need to find a lower bound on the above expression under the constraint that

$$n = p_1 \cdots p_k \cdot q_1 \cdot q_2 \cdot q_3 \cdot P \cdot Q \le \left(\frac{y}{k^2}\right)^k (ky)^3 y^2 := x. \tag{6}$$

We will do this by choosing $k = \left\lfloor c\sqrt{\log x / \log\log x}\right\rfloor + \nu$, where $\nu \in \{0, 1\}$ is such that k even and c is a constant to be determined later. Then, by estimate (5), we get

$$\begin{aligned} \#\mathcal{A}(x) &\ge \frac{x}{\exp\left(k \log 4k + \log y + (k+5)\log\log y\right)} \\ &= x \exp\Big(-c/2\sqrt{\log x \log\log x} - \log y \; c\sqrt{\log x/\log\log x}\log\log y \\ &\quad - O(k + \log\log y)\Big). \end{aligned}$$

Estimate (6) together with the choice of k leads to the conclusion that $\log y = c^{-1}(1 + o(1))\sqrt{\log x \log\log x}$ as $x \to \infty$, which, in turn, leads to the lower-bound

$$\#\mathcal{A}(x) \gg \frac{x}{\exp\left((c + c^{-1} + o(1))\sqrt{\log x \log\log x)}\right)}.$$

The minimum of the function $c \mapsto c + c^{-1}$ is attained at $c = 1$. Hence, choosing $c = 1$, we get the lower bound of the statement. $\square$

ACKNOWLEDGEMENTS This work was done in February of 2006, while both authors were in residence at the Centre de Recherche Mathemátique of the Université de Montréal for the thematic year *Analysis and Number Theory.* The authors thank the organizers for the opportunity of participating in this program.

References

[1] W. D. Banks, M. Z. Garaev, F. Luca and I. E. Shparlinski, *Uniform distribution of the fractional part of the average prime factor*, Forum Mathematicum. **17** (2005), 885–903.

[2] A. Hildebrand, *On the number of positive integers $\le x$ and free of prime factors $> y$*, J. Number Theory. **22** (1986), 289–307.

[3] I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, Comptes Rendues (Doklady) de l'Academie des Sciences de l'USSR. **15** (1937), 191–294.

(F. Luca) Instituto de Matemáticas, Universidad Nacional, Autónoma de México, C.P. 58089, Morelia, Michoacán, México

E-mail address: `fluca@matmor.unam.mx`

(F. Pappalardi) Dipartimento di Matematica, Università Roma Tre, Largo S. L. Murialdo, 1, Roma, 00146, Italy

E-mail address: `pappa@mat.uniroma3.it`