ERDÖS, P.
1959
CAN. J. MATH.
Vol. 11, pp 161–167

SOME REMARKS ON PRIME FACTORS OF INTEGERS

PAUL ERDÖS

Reprinted from
Canadian Journal of Mathematics

SOME REMARKS ON PRIME FACTORS OF INTEGERS

P. ERDÖS

1. Let $1 < a_1 < a_2 < \ldots$ be a sequence of integers and let $N(x)$ denote the number of a's not exceeding x. If $N(x)/x$ tends to a limit as x tends to infinity we say that the a's have a density. Often one calls it the asymptotic density to distinguish it from the Schnirelmann or arithmetical density. The statement that almost all integers have a certain property will mean that the integers which do not have this property have density 0. Throughout this paper p, q, r will denote primes.

I conjectured for a long time that, if $\epsilon > 0$ is any given number, then almost all integers n have two divisors d_1 and d_2 satisfying

$$d_1 < d_2 < (1 + \epsilon)\, d_1. \tag{1}$$

I proved **(1**, p. 691**)** that the integers with two divisors satisfying (1) have a density, but I cannot prove that this density has the value 1. However, analogous questions can be asked about the prime divisors of integers and a more complete result is contained in the following theorem.

THEOREM 1. *Let $\epsilon_p > 0$, $\delta_p = \epsilon_p$ if $\epsilon_p \leqslant 1$ and $\delta_p = 1$ if $\epsilon_p \geqslant 1$. The divergence of $\sum_p \delta_p/p$ is a necessary and sufficient condition that almost all integers should have two prime factors p and q satisfying*

$$p < q < p^{1+\epsilon_p}. \tag{2}$$

From the prime number theorem we have

$$p_n = (1 + o(1))n \log n;$$

thus $\sum_p \epsilon_p p^{-1}$ will diverge if $\epsilon_p = (\log \log p)^{-1}$, but will converge if $\epsilon_p = (\log \log p)^{-1-c}$, for any $c > 0$.

Further, we shall outline a proof of

THEOREM 2. *The density of integers n which have two prime factors p and q satisfying*

$$p < q < p^{1+c/\log\log n}$$

equals $1 - e^{-c}$.

Let $p_1 < p_2 < \ldots < p_k$ be the distinct prime factors of n. Define the real number η_i by $p_i^{\eta_i} = p_{i+1}$. A famous result of Hardy and Ramanujan **(4)** asserts that $k = (1 + o(1)) \log \log n$ for almost all n. I proved **(2**, p.

Received April 23, 1958.

533, Theorem 9) that, for almost all n, the number of η's not exceeding t ($t > 1$) is

$$(1 + o(1))\left(1 - \frac{1}{t}\right) \log \log n.$$

Theorem 2 can be stated as follows: the density of integers with

$$\min_{1 \leqslant i \leqslant k} \eta_i < 1 + \frac{c}{\log \log n}$$

is $1 - e^{-c}$. By similar methods, we can prove that the density of integers n satisfying

$$\max_{1 \leqslant i \leqslant k} \eta_i > c \log \log n$$

is $1 - \exp[-1/c]$. Further, we can prove that the divergence of $\sum_p \delta_p/p$ ($\delta_p < 1$) is the necessary and sufficient condition that almost all integers n should have a prime factor p such that $n \equiv 0 \pmod{p}$, and $n \not\equiv 0 \pmod{q}$ for all primes with

$$p \leqslant q < p^{\delta_p^{-1}}.$$

We shall not give the proof of these results, since they are similar to those of Theorems 1 and 2.

2. First, we show that the condition of Theorem 1 is necessary. In fact, we show that if $\sum_p \delta_p/p < \infty$, then the upper density of integers having two prime divisors satisfying (2) is less than one. Since $\sum_p \delta_p p^{-1} < \infty$, it is clear that

$$\sum_{\epsilon_p > 1} p^{-1} < \infty.$$

Denote by $b_1 < b_2 < \ldots$ the integers consisting of the primes p satisfying $\epsilon_p \geqslant 1$ and the integers of the form pq, where $\epsilon_p < 1$ and $p < q < p^{1+\epsilon_p}$. Clearly the integers not divisible by any b have no divisor of the form pq satisfying (2). But $\sum b_i^{-1} < \infty$; thus by a well-known and simple argument (**3**, p. 279) one can show that the density of integers divisible by a b is less than one. We really only proved that if $\sum_p \delta_p/p < 1$ then the upper density of integers having a divisor of the form pq satisfying (2) is less than one. In fact it would be quite easy to show that the density in question exists.

Now we prove the sufficiency of Theorem 1. We first show that it will suffice to prove the following

THEOREM 1′. *Let* $\epsilon_p < \frac{1}{4}$, $\epsilon_p \to 0$, $\sum_p \epsilon_p/p = \infty$. *Then the density of integers* n *having two prime divisors* p *and* q *satisfying*

$$p < q < p^{1+\epsilon_p}$$

is 1.

To deduce the sufficiency of the condition of Theorem 1 from Theorem 1′ it will suffice to show that if $\sum_p \delta_p/p = \infty$ there always exists an $\epsilon_p' \leqslant \epsilon_p$,

$\epsilon_p' < \frac{1}{4}$, $\sum_p \epsilon_p'/p = \infty$. To show this we observe that if $\sum_p \delta_p/p = \infty$ then either there exists a subsequence p_i with

$$\sum_i \epsilon_{p_i} p_i^{-1} = \infty, \qquad \epsilon_{p_i} < \tfrac{1}{4}$$

and then we put

$$\epsilon'_{p_i} = \epsilon_{p_i}, \qquad 1 \leqslant i < \infty,$$

$\epsilon_p' = 0$ if $p \neq p_i$, or for a certain

$$c \geqslant \tfrac{1}{8}, \qquad \sum_{\epsilon_p > c} p^{-1} = \infty.$$

But in this case there clearly exists an $\epsilon_p' < \epsilon_p$ such that

$$\epsilon'_p \to 0, \qquad \epsilon'_p < \tfrac{1}{4}, \qquad \sum \frac{\epsilon'_p}{p} = \infty,$$

which completes our proof.

Now we prove Theorem 1′. Put

$$\sum_{p<x} \frac{1}{p} \sum_{p<q<p^{1+\epsilon_p}} \frac{1}{q} = A(x);$$

then, since $\sum_p \epsilon_p/p = \infty$,

$$A(x) \to \infty \quad \text{as} \quad x \to \infty.$$

We have to show that almost all integers have at least one divisor of the form pq, where $p < q < p^{1+\epsilon_p}$. Instead of this we shall prove the stronger result that if $f(n)$ denotes the number of divisors of n of the above form then, for almost all n,

$$f(n) = (1 + o(1))A(n). \tag{3}$$

Or, because of the slow growth of $A(n)$, we shall in fact prove that

$$f(n) = (1 + o(1))A(x), \tag{4}$$

except for $o(x)$ values of $n \leqslant x$. It is easy to see that (3) and (4) are equivalent since

$$A(x) - A(x^{\frac{1}{2}}) = \sum_{x^{\frac{1}{2}}<p<x} \frac{1}{p} \sum_{p<q<p^{1+\epsilon_p}} \frac{1}{q} = \sum_{x^{\frac{1}{2}}<p<x} \frac{\epsilon_p + o(1)}{p} = o(1) \tag{5}$$

by the well-known estimate

$$\sum_{p \leqslant x} p^{-1} = \log \log x + c_1 + O\left(\frac{1}{\log x}\right).$$

To prove (4) we shall use Turan's method (**6**, pp. 274–6). We have

$$\sum_{n=1}^{x} (f(n) - A(x))^2 = x(A(x))^2 - 2A(x) \sum_{n=1}^{x} f(n) + \sum_{n=1}^{x} f^2(n). \tag{6}$$

Since

$$f(n) = \sum_{\substack{pq \mid n \\ p<q<p^{1+\epsilon_p}}} 1,$$

we may write

(7) $$\sum_{n=1}^{x} f(n) = \sum_{p\leq x} \sum_{p<q<p^{1+\epsilon_p}} \left[\frac{x}{pq}\right] = x \sum_{p}{}' \sum_{p<q<p^{1+\epsilon_p}}{}' \frac{1}{pq} + O(x),$$

where the dash indicates that $pq \leqslant x$.* Now $\epsilon_p < \frac{1}{4}$ implies that for $p < x^{\frac{1}{2}}$, $pq < x$

$$A(x^{\frac{1}{2}}) \leqslant \sum_{p}{}' \sum_{p<q<p^{1+\epsilon_p}}{}' \frac{1}{pq} \leqslant A(x).$$

Thus from (5),

(8) $$\sum_{n=1}^{x} f(n) = xA(x) + O(x).$$

Similarly,

(9) $$\sum_{n=1}^{x} f^2(n) = \sum_{p\leq x} \sum_{p<q<p^{1+\epsilon_p}} \left[\frac{x}{pq}\right] + \sum \sum \left[\frac{x}{\{p_1q_1, p_2q_2\}}\right],$$

where in the second sum

$$p_1 < q_1 < p_1^{1+\epsilon_{p_1}}, \qquad p_2 < q_2 < p_2^{1+\epsilon_{p_2}},$$

$p_1q_1 \neq p_2q_2$, and ($\{p_1q_1, p_2q_2\}$ denotes the least common multiple of p_1q_1 and p_2q_2).

The first sum on the right of (9) is $(1 + o(1))xA(x)$. For the second sum we have

(10) $$\sum \sum \left[\frac{x}{\{p_1q_1, p_2q_2\}}\right] = x \sum{}' \sum{}' \frac{1}{\{p_1q_1, p_2q_2\}} + O(x),$$

where the dash indicates that $p_1q_1 \neq p_2q_2$ and $\{p_1q_1, p_2q_2\} \leqslant x$. Clearly, from (5), if $p_1 < x^{\frac{1}{4}}$, $p_2 < x^{\frac{1}{4}}$, $\{p_1q_1, p_2q_2\} < x$

(11) $$\sum{}' \sum{}' \frac{1}{\{p_1q_1, p_2q_2\}} > (A(x^{\frac{1}{4}}))^2 + O(1) = (1 + o(1))(A(x))^2.$$

On the other hand, by a simple argument,

(12) $$\sum{}' \sum{}' \frac{1}{\{p_1q_1, p_2q_2\}} < A^2(x) + 4 \sum{}'' \frac{1}{r_1r_2r_3},$$

where in $\sum''$

$$r_1 < r_2 < r_1^{1+\epsilon_{r_1}}, \qquad r_3 < \max(r_1^{1+\epsilon_{r_1}}, \qquad r_2^{1+\epsilon_{r_2}}),$$

or $r_3 < r_1^2$, and $r_1 \leqslant x$. (12) follows from the fact that $r_1r_2r_3 = \{p_1q_1, p_2q_2\}$ has four solutions. Now

*Since $p < q$, the equation $pq = \lambda$ has at most one solution p, q and so there are at most x terms in the double sum. Hence the error in omitting the square brackets is at most x.

$$\sum{}'' \frac{1}{r_1 r_2 r_3} \leqslant \sum_{p<x} \frac{1}{p} \sum_{p<q<p^{1+\epsilon_p}} \frac{1}{q} \sum_{p<r<p^2} \frac{1}{r} < c\, A(x),$$

hence

$$\sum{}' \sum{}' \frac{1}{\{p_1 q_1,\, p_2 q_2\}} = (1 + o(1)) A^2(x). \tag{13}$$

Thus, by (9), (10), and (13),

$$\sum_{n=1}^{x} f^2(n) = (1 + o(1)) x (A(x))^2. \tag{14}$$

Hence from (6), (8), and (14)

$$\sum_{n=1}^{x} (f(n) - A(x))^2 = o(xA^2(x)),$$

which proves that $f(n) = (1 + o(1))A(x)$, except for $o(x)$ values of $n \leqslant x$. Thus Theorem 1 is proved.

3. Now we outline the proof of Theorem 2. Denote by $a_1 < a_2 < \ldots < a_k \leqslant x$, the integers not exceeding x of the form pq, where

$$p < q < p^{1+c/\log \log x}.$$

Clearly the a's depend on x and $a_1 \to \infty$ as x tends to infinity. Denote by $N_c(a_1, \ldots, a_k; x)$ the number of integers not exceeding x which are not divisible by any of the a_i's. Further, denote by $M_c(x)$ the number of integers $n \leqslant x$ which do not have two prime factors p and q satisfying

$$p < q < p^{1+c/\log \log n}.$$

We have to prove that

$$M_c(x) = (1 + o(1)) e^{-c} x. \tag{15}$$

Clearly

$$M_c(x) \leqslant N_c(a_1, a_2, \ldots, a_k; x),$$

but because of the slow increase of $\log \log n$ it is easy to see that

$$M_c(x) = N_c(a_1, a_2, \ldots, a_k; x) + o(x).$$

Thus to prove Theorem 2 it will suffice to show that

$$N_c(a_1, a_2, \ldots, a_k; x) = xe^{-c} + o(x). \tag{16}$$

We obtain by a simple sieve process the well-known formula

$$N_c(a_1, a_2, \ldots, a_k; x) = x \sum_{l=0}^{k} (-1)^l \sum{}_l,$$

where

$$\sum{}_0 = 1, \qquad \text{and} \qquad \sum{}_l = \sum \frac{1}{\{a_{i_1}, \ldots, a_{i_l}\}},$$

where $i_1, i_2, \ldots, i_l$ runs through all distinct l-tuples from 1 to k. (The curly bracket in the denominator denotes least common multiple.)

By a well-known combinatorial argument*

$$(17) \quad x \sum_{l=0}^{2t-1} \left((-1)^l \sum\nolimits_l\right) \leqslant N(a_1, \ldots, a_k; x) \leqslant x \sum_{l=0}^{2t} \left((-1)^l \sum\nolimits_l\right),$$

for every $t > 0$. We evidently have, by a simple computation (the dashes indicate that $p < q < p^{1+c/\log\log x}$ and $pq < x$)

$$(18) \quad \sum_{i=1}^{k} \frac{1}{a_i} = \sum{}' \frac{1}{p} \sum{}' \frac{1}{q} = \frac{(1+o(1))c}{\log\log x} \sum_{p<x} \frac{1}{p} + o(1) = c + o(1),$$

by the estimate for $\sum_{p<x} p^{-1}$. Further, for every fixed l (the two dashes indicate that

$$\{a_{i_1}, a_{i_2}, \ldots, a_{i_l}\} \leqslant x)$$

$$(19) \qquad \sum\nolimits_l = \sum \left[\frac{x}{\{a_{i_1}, \ldots, a_{i_l}\}}\right] = x \sum\nolimits_l'' \frac{1}{\{a_{i_1}, \ldots, a_{i_l}\}} + o(x)$$
$$= x \sum\nolimits_l'' + o(x),$$

since there are only $o(x)$ l-tuples satisfying

$$\{a_{i_1}, \ldots, a_{i_l}\} < x.$$

This last statement follows from the fact that the integers

$$\{a_{i_1}, \ldots, a_{i_l}\}$$

have at most $2l$ prime factors and, by a well-known theorem of Landau (**5**, Vol. I, pp. 208–11), the number of integers not exceeding x having $2l$ prime factors equals

$$(1 + o(1)) \frac{x}{\log x} \frac{(\log\log x)^{2l-1}}{(2l-1)!} = o(x),$$

and finally a simple argument shows that the number of solutions of

$$y = \{a_{i_1}, \ldots, a_{i_l}\}$$

is less than a constant depending only on l.

Now we outline the proof of

$$(20) \qquad \sum\nolimits_l'' = \frac{c^l}{l!} + o(1).$$

For $l = 1$, (20) follows from (18). For $l > 1$ we can prove (20) by a simple induction process, similar but a bit more complicated than that used in the estimations in Theorem 1. We do not give the details since they are somewhat cumbersome.

*This is one of the basic ideas of Brun's method, see for example, Landau *Zahlentheorie*, Vol. 1, Kap. 2.

From (17) and (20) we have

$$(21) \qquad N_c(a_1, \ldots, a_k; x) = x \sum_{l=0}^{\infty} \frac{(-1)^l c^l}{l!} + o(x) = x\, e^{-c} + o(x),$$

which is (16).

References

1. P. Erdös, *On the density of some sequences of integers*, Bull. Amer. Math. Soc., *54* (1948), 691.
2. ———, *Some remarks about additive and multiplicative functions*, Bull. Amer. Math. Soc., *52* (1946), 533, theorem 9.
3. ———, *On the density of the abundant numbers*, J. Lond. Math. Soc., *9* (1934), 279.
4. G. H. Hardy and S. Ramanujan, *The normal number of prime factors of n*, Quart. J. Math., *48* (1917), 76–92.
5. E. Landau, *Verteilung der Primzahlen*,
6. P. Turán, *On a theorem of Hardy and Ramanujan*, J. Lond. Math. Soc., *9* (1934), 274–6.

University of Alberta