ACTA ARITHMETICA
XXVIII (1976)

An asymptotic formula in additive number theory

by

P. ERDÖS (Budapest), G. JOGESH BABU and K. RAMACHANDRA
(Bombay)

1. Introduction. In his paper [1], Erdös introduced the sequences of positive integers $b_1 < b_2 < \ldots$, with $(b_i, b_j) = 1$, for $i \neq j$, and $\sum b_i^{-1} < \infty$. With any such arbitrary sequence of integers, he associated the sequence $\{d_i\}$ of all positive integers not divisible by any b_j, and he showed that if $b_1 \geqslant 2$, there exists a $\theta < 1$ (independent of the sequence $\{b_i\}$) such that $d_{i+1} - d_i < d_i^{\theta}$, for $i \geqslant i_0$. Later, Szemerédi [4] made an important progress on the problem, showing that θ can be taken to be any number greater than $\frac{1}{2}$.

In this paper, we study this sequence from a different point of view. We study the number $N(n)$ of solutions of the equation $n = p + d$, where p is a prime and $d \not\equiv 0 \pmod{b_j}$ for any j. In fact we derive an asymptotic formula for $N(n)$, when $b_1 \geqslant 3$. We also study $N(n)$ when the condition $(b_i, b_j) = 1$ is dropped.

2. In what follows, we let $C_1, C_2, \ldots$ denote positive absolute constants and let C be a positive constant. p, q with or without subscript, always denote primes.

THEOREM 1. *Let $2 \leqslant b_1 < b_2 < \ldots$ be a sequence of natural numbers with the properties $(b_i, b_j) = 1$ whenever $i \neq j$ and*

$$\sum_{j=1}^{\infty} b_j^{-1} < \infty. \tag{2.1}$$

Then the number $N(n)$ of solutions of the equation $n = p + t$, where p is a prime and t is a natural number not divisible by any b_j, is given by

$$N(n) = n(\log n)^{-1} \prod_{(b_j, n)=1} \Big(1 - \big(\varphi(b_j)\big)^{-1}\Big) + o\big(n(\log n)^{-1}\big). \tag{2.2}$$

Remarks. If either $b_1 \geqslant 3$ or if n is even then $N(n)$ is asymptotic to the main term in (2.2). Similar remarks apply to Theorem 2 below, which can be proved along the same lines as Theorem 1. Also it easily follows from the prime number theorem for arithmetic progressions and the sieve of Eratosthenes that if $(b_i, b_j) = 1$ and $\sum_{j=1}^{\infty} \frac{1}{b_j} = \infty$ then $N(n) = o\left(\frac{n}{\log n}\right)$.

THEOREM 2. *Let l be any non-zero integer. Under the assumptions of Theorem 1, the number $N_l(x)$, of primes p not exceeding x such that $p+l$ is not divisible by any b_j, satisfies*

$$N_l(x) = x(\log x)^{-1} \prod_{(b_j, l)=1} \left(1-(\varphi(b_j))^{-1}\right) + o\left(x(\log x)^{-1}\right).$$

3. Proof of Theorem 1. We denote by ν, natural numbers not divisible by any b_j, and by d all finite power products $\prod b_j^{e_j}$ where $e_j = 0$ or 1, and we write $h(d) = (-1)^{\Sigma e_j}$. We begin with

LEMMA 1. *We have*

$$\sum \nu^{-s} = \zeta(s) \prod (1-b_j^{-s}) \quad \textit{and} \quad \prod (1-b_j^{-s}) = \sum h(d) d^{-s}.$$

Proof. The proof follows from the fact that every natural number m can be written uniquely in the form

$$m = \left(\prod b_j^{a_j}\right)\nu \quad (a_j \geqslant 0 \text{ are integers}).$$

This can be proved in the following way. Define a_j as the greatest integer such that $b_j^{a_j}$ divides m. This gives existence and the uniqueness is trivial.

LEMMA 2. *The two series*

$$\sum (\varphi(b_j))^{-1} \quad \textit{and} \quad \sum (\varphi(d))^{-1}$$

are convergent.

Proof. Let B_1 be the set of those b's which are primes and let B_2 be the set of the remaining b's. Clearly, the number of b's in B_2 not exceeding x is less than $\sqrt{x}$. Thus (2.1) implies convergence of the first series. Convergence of the second series follows from convergence of the first series and the identity

$$\sum (\varphi(d))^{-1} = \prod \left(1-(\varphi(b_i))^{-1}\right).$$

LEMMA 3. *Let $N'(n)$ be the number of solutions of*

$$n = p + t', \quad t' > 0, \quad t' \not\equiv 0 \pmod{b_i} \quad \textit{for every } b_i \leqslant \log\log n.$$

Then

$$N'(n) = n(\log n)^{-1} \prod_{(b_i, n)=1} \left(1-(\varphi(b_i))^{-1}\right) + o\left(n(\log n)^{-1}\right).$$

Proof. Let d' denote a product of the form $\prod b_i^{e_i}$, where $e_i = 0$ or 1 and $b_i \leqslant \log\log n$. By Siegel–Walfisz theorem (see [3], Satz 8.3, p. 144) and by Lemmas 1 and 2, we have

$$N'(n) = \sum_{n=p+t'} 1 = \sum_{p+md'=n} h(d') = \sum_{\substack{p+md'=n\\(d',n)=1}} h(d') + \sum_{\substack{p+md'=n\\(d',n)>1}} h(d') = \Sigma_1+\Sigma_2.$$

Note that, if $d(n)$ denotes the number of divisors of n, then

$$\Sigma_2 = \Big|\sum_{\substack{p+md'=n\\(d',n)=p}} h(d')\Big| \leqslant \sum_{p|n}\sum_{\substack{d'|n-p\\(d',n)=p}} h(p) \leqslant \sum_{p|n} d(n-p) \ll n^{1/2}\log n,$$

since $|h(d')| \leqslant 1$ and $d(n) \ll n^{\varepsilon}$ for any $\varepsilon > 0$.

$$\Sigma_1 = \sum_{(d',n)=1}\left(\frac{h(d')}{\varphi(d')}\frac{n}{\log n}\big(1+O\big((\log n)^{-1}\big)\big)\right)$$
$$= \frac{n}{\log n}\left(\sum_{(d,n)=1}\frac{h(d)}{\varphi(d)}\right)+o\left(\frac{n}{\log n}\right).$$

Thus

$$N'(n) = \Sigma_1+\Sigma_2 = n(\log n)^{-1}\prod_{(b_i,n)=1}\big(1-\big(\varphi(b_i)\big)^{-1}\big)+o\big(n(\log n)^{-1}\big).$$

This completes the proof of the lemma.

LEMMA 4. *There exists a function $\eta(\varepsilon)\to 0$ as $\varepsilon\to 0$, such that the number of primes $p\leqslant n$ satisfying*

$$n-p \equiv 0 (\operatorname{mod} b_i), \quad \textit{for some } b_i\in(n^{1-\varepsilon}, n]$$

is less than

$$\big(\eta(\varepsilon)+o(1)\big)n(\log n)^{-1}, \quad \textit{for every } \varepsilon\in(0, \tfrac{1}{4}).$$

Proof. First note that the number of composite b_i's not exceeding n is at most $n^{1/2}$. For a fixed $b_i\in(n^{1-\varepsilon}, n]$, $n-p\equiv 0(\operatorname{mod} b_i)$ has at most $(n/b_i) < n^{\varepsilon}$ solutions. Thus the contribution of the composite b_i's is less than $n^{1/2+\varepsilon}$. To complete the proof it, thus, suffices to show that the number of solutions of

$$n \equiv p(\operatorname{mod} q), \quad n^{1-\varepsilon} < q < n, \ q \text{ prime},$$

is less than

$$\big(\eta(\varepsilon)+o(1)\big)n(\log n)^{-1}.$$

In other words we have to prove that the number of solutions of

$$n = p+aq, \quad p, q \text{ primes not exceeding } n \text{ and } a < n^{\varepsilon}$$

is less than

$$\big(\eta(\varepsilon)+o(1)\big)n(\log n)^{-1}.$$

First note that the number of solutions of

$$n = p+aq, \quad a < n^{\varepsilon}, \quad (a, n) > 1 \text{ and } p, q \text{ primes not exceeding } n$$

is less than

$$\sum_{a<n^{\varepsilon}} \sum_{p|a} 1 \ll n^{2\varepsilon} = o\big(n(\log n)^{-1}\big),$$

since $\varepsilon < 1/4$.

Now for a fixed $a < n^{\varepsilon}$ and $(n, a) = 1$, the number of primes $q < n$, for which $n - aq$ is a prime, by Lemma 1.4 of [2], if C_2 is a sufficiently small constant, is less than

$$C_1 \frac{n}{a} \prod_{2<p<n^{C_2}} \left(1-\frac{2}{p}\right) \prod_{p|n} \left(1-\frac{1}{p}\right)^{-1} < C_3 \frac{n}{a} \prod_{2<p<n^{C_2}} \left(1-\frac{2}{p}\right) \prod_{p|n} \left(1+\frac{1}{p}\right)$$
$$< C_4 \frac{n}{a} (\log n)^{-2} \prod_{p|n} \left(1+\frac{1}{p}\right).$$

Thus summing for all $a < n^{\varepsilon}$, $(a, n) = 1$, we immediately obtain that the number of solutions of

$$n - aq = p, \quad a < n^{\varepsilon},\ (a, n) = 1 \text{ and } p, q \text{ primes } (\leqslant n)$$

is less than

$$\eta(\varepsilon) n (\log n)^{-1}.$$

Now the lemma follows easily.

To complete the proof of Theorem 1, it is enough to show, in view of Lemma 3, that

$$N(n) - N'(n) = o\big(n(\log n)^{-1}\big).$$

To show this it will clearly be sufficient to show that the number of solutions of

$$n = p + R, \quad R > 0,\ R \equiv 0 (\operatorname{mod} b_j) \text{ for some } b_j > \log\log n$$

is

$$o\big(n(\log n)^{-1}\big).$$

First observe that if $b_i \leqslant n^{1-\varepsilon}$ ($\varepsilon > 0$, small), then the number of primes $p \leqslant n$ with $n \equiv p (\operatorname{mod} b_j)$ is, by Brun–Titchmarsh Theorem (see [3], Satz 4.1, p. 44), less than $\big(C_5 n/\varepsilon\varphi(b_i)\log n\big)$. Thus the number of primes $p \leqslant n$ for which $n \equiv p (\operatorname{mod} b_i)$ for some $b_i \in (\log\log n, n^{1-\varepsilon}]$ is less than

$$(C_5 n/\varepsilon \log n) \sum_{b_i > \log\log n} \big(\varphi(b_i)\big)^{-1} = o(n/\varepsilon \log n).$$

Now the theorem follows from Lemma 4.

4. If $(b_i, b_j) = 1$, for $i \neq j$, is not assumed, it is easy to give a sequence $2 < b_1 < b_2 < \ldots$ for which

$$\sum_{i=1}^{\infty} \big(\varphi(b_i)\big)^{-1} < \infty,$$

but there is an infinite sequence $0 < n_1 < n_2 < \ldots$ so that the number of solutions of

$$n_i = p+t, \quad p \text{ prime}, \ t > 0 \text{ and } t \not\equiv 0 (\bmod b_j), \text{ for all } j,$$

is

$$o(n_i/\log n_i) \quad \text{as} \quad i \to \infty.$$

We define $b_1 < b_2 < \ldots$ as follows. Suppose $\{n_i\}$ be an increasing sequence of natural numbers tending to infinity sufficiently fast and $\varepsilon_i = (\log\log n_i)^{-1}$. Now take the b's to be the integers of the form

$$n_i - p, \quad p < (1-\varepsilon_i) n_i, \quad i = 1, 2, \ldots$$

Clearly the number of

$$n_i = p+t, \quad t > 0, \ t \not\equiv 0 (\bmod b_j), \text{ for all } j,$$

is less than

$$\big(\varepsilon_i + o(1)\big)(n_i/\log n_i) = o(n_i/\log n_i).$$

Since

$$\varphi(m) \geqslant C_6 m (\log\log m)^{-1}, \tag{4.1}$$

we have

$$\sum_{p<(1-\varepsilon_i)n_i} \frac{1}{\varphi(n_i-p)} < \frac{C_6 n_i}{\log n_i} \frac{\log\log n_i}{\varepsilon_i n_i} = \frac{C_6 (\log\log n_i)}{\log n_i}.$$

Thus

$$\sum_{i=1}^{\infty} \big(\varphi(b_i)\big)^{-1} \leqslant \sum_{i=1}^{\infty} \sum_{p<(1-\varepsilon_i)n_i} \big(\varphi(n_i-p)\big)^{-1} \leqslant C_6 \sum_{i=1}^{\infty} \frac{(\log\log n_i)^2}{\log n_i} < \infty,$$

if $n_i \to \infty$ sufficiently fast.

It might be possible to construct a sequence $2 < b_1 < b_2 < \ldots$ of integers such that $\sum b_i^{-1}$ is convergent and for which

$$n = p+t, \quad p \text{prime}, \ t > 0, \ t \not\equiv 0 (\bmod b_i), \text{ for all } i,$$

has no solution for infinitely many n. But we are unable to find such a sequence.

On the other hand, if $B(x)$, defined b,

$$B(x) = \sum_{b_i \leqslant x} 1,$$

is not too large, then the condition $(b_i, b_j) = 1$, for $i \neq j$, is quite unnecessary. In this direction, we have the following

THEOREM 3. *Let $3 \leqslant b_1 < b_2 < \ldots$ be a sequence of integers such that*

$$B(x) = o\Big(x/\big((\log x)^2 \log\log x\big)\Big). \tag{4.2}$$

Then

$$N(n) > Cn(\log n)^{-1}.$$

Proof of Theorem 3. Let, for any $k \geqslant 1$, $N(n, k)$ be the number of solutions of $n = p+t$, p prime, $t > 0$ and $t \not\equiv 0 \pmod{b_j}$, for all $j \leqslant k$, and let $A(n, k)$ be the number of solutions of $n = p+t$, $t > 0$, $t \equiv 0 \pmod{b_j}$ for some $j > k$. We need the following lemmas.

LEMMA 5. *For every $k \geqslant 1$, there exists $n(k)$ such that*

$$N(n, k) \geqslant C_7\big(n/(\log n)(\log k)\big), \quad \text{for all } n \geqslant n(k).$$

Proof. Since each $b_i \geqslant 3$, either $b_i \equiv 0 \pmod{2^2}$, or there exists a prime $q_i' \geqslant 3$ such that $b_i \equiv 0 \pmod{q_i'}$. Let $l(k)$ be the number of distinct primes in the set $\{q_i'\}$. Let these be denoted by q_i, $i = 1, \ldots, l(k)$.

Note that, $N(n, k)$ is not less than the number of solutions of

$$n = p+t, \quad t > 0,\ t \equiv 0 \pmod{2^2} \text{ and } t \equiv 0 \pmod{q_i} \text{ for all } i \leqslant l(k).$$

This latter number solutions, by Theorem 1, is not less than

$$\begin{aligned}\left(1-\frac{1}{\varphi(4)}\right)\prod_{i\leqslant l(k)}\left(1-\frac{1}{\varphi(q_i)}\right)\frac{n}{\log n}+o\left(\frac{n}{\log n}\right)\\ \geqslant \frac{1}{2}\prod_{i\leqslant k}\left(1-\frac{1}{p_i-1}\right)\frac{n}{\log n}+o\left(\frac{n}{\log n}\right)\\ \geqslant \frac{C_8}{\log k}\,\frac{n}{\log n} \quad \text{for all } n \geqslant n(k),\end{aligned}$$

where p_i is the ith odd prime number and $n(k)$ is a sufficiently large integer. This completes the proof of Lemma 5.

LEMMA 6. *We have*

$$\sum_{i\geqslant k}\big(\varphi(b_i)\big)^{-1} = o\big((\log k)^{-1}\big). \tag{4.3}$$

Proof. By (4.1), (4.2) and by partial integration, we have

$$\begin{aligned}\sum_{i\geqslant k}\big(\varphi(b_i)\big)^{-1} &\leqslant \sum_{i\geqslant k}\frac{\log\log b_i}{b_i} = \int_{b_k}^{\infty}\frac{\log\log t}{t}\,dB(t)\\ &= \frac{1}{t}B(t)\log\log t\Big]_{b_k}^{\infty} + \int_{b_k}^{\infty}\frac{B(t)}{t^2}\left(\log\log t-\frac{1}{\log t}\right)dt\\ &= o\big((\log b_k)^{-2}\big)+o\left(\int_{b_k}^{\infty}\frac{dt}{t(\log t)^2}\right) = o\big((\log b_k)^{-1}\big)\\ &= o\big((\log k)^{-1}\big).\end{aligned}$$

LEMMA 7. *There exists a k_0 such that, for every $k \geqslant k_0$, there exists $n_0(k)$ satisfying*

$$A(n,k) \leqslant \frac{C_7}{2\log k}\frac{n}{\log n} \quad \textit{for all } n \geqslant n_0(k).$$

Proof. Since the number of solutions of $n \equiv p(\mathrm{mod}\, b_i)$ is, by Brun–Titchmarsh theorem for $b_i \leqslant \sqrt{n}$, less than $\big(C_8 n/\varphi(b_i)\log n\big)$, thus, for any $k \geqslant 1$, the number of solutions of

$$n = p+t, \quad p \leqslant n,\, t \equiv 0(\mathrm{mod}\, b_j), \text{ for } b_j \leqslant \sqrt{n} \text{ and } j > k$$

is less than

$$C_8 n(\log n)^{-1}\sum_{i>k}\big(\varphi(b_i)\big)^{-1}. \tag{4.4}$$

By Lemma 6, there exists a k_0 such that for $k \geqslant k_0$, (4.4) is less than

$$\frac{C_7}{10\log k}\frac{n}{\log n}. \tag{4.5}$$

Let, next, $b_j > \sqrt{n}$. By Brun–Titchmarsh Theorem the number of solutions of

$$n \equiv p(\mathrm{mod}\, b_j), \quad p \leqslant n,$$

is less than

$$\left(C_9 n/\varphi(b_j)\log\frac{n}{b_j}\right).$$

So, if $s \geqslant 1$ and $2^s < \sqrt{n}$, then the number of solutions of

$$n \equiv p(\mathrm{mod}\, b_j), \quad \frac{n}{2^{s+1}} < b_j \leqslant \frac{n}{2^s}, \quad p \leqslant n,$$

is less than

$$B(n/2^s)C_{10}\frac{2^s}{s}\log\log n = o\big(s^{-1}n(\log n)^{-2}\big) \quad \text{as} \quad n \to \infty. \tag{4.6}$$

Here we used (4.2). Since, for each $b_j \in (n/2, n]$, there exists at most one prime $p \leqslant n$ such that $n \equiv p(\mathrm{mod}\, b_j)$, the number of solutions of

$$n \equiv p(\mathrm{mod}\, b_j), \quad p \leqslant n,\ b_j \in (n/2, n]$$

is less than

$$B(n) = o\Big(n/\big((\log n)^2\log\log n\big)\Big). \tag{4.7}$$

By summing (4.6) over s and adding (4.7) to the result, we get that the number of solutions of

$$n \equiv p(\mathrm{mod}\, b_j), \quad \text{for some } b_j \geqslant \sqrt{n},\, p < n$$

is

$$o\left(n(\log n)^{-1}\right).$$

Now the lemma follows from (4.5).

To complete the proof of Theorem 3, first note that for any $k \geqslant 1$

$$N(n) \geqslant N(n, k) - A(n, k). \tag{4.8}$$

Now the theorem follows immediately from (4.8) and Lemmas 5 and 7.

Without much difficulty we could obtain an asymptotic formula for $N(n)$ even if we only assume

$$B(x) = o\left(\frac{x}{\log x \log\log x}\right).$$

We hope to return to this problem on another occasion.

References

[1] P. Erdös, *On the difference of consecutive terms of sequences defined by divisibility properties*, Acta Arith. 12 (1966), pp. 175–182.

[2] J. Kubilius, *Probabilistic methods in the theory of numbers*, Transl. of Math. Monographs, Amer. Math. Soc. 11 (1964).

[3] K. Prachar, *Primzahlverteilung*, Berlin 1957.

[4] E. Szemerédi, *On the difference of consecutive terms of sequences defined by divisibility properties II*, Acta Arith. 23 (1973), pp. 359–361.

MATHEMATICAL INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
Budapest, Hungary
SCHOOL OF MATHEMATICS
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
Colaba, Bombay 5, India

Received on 20. 4. 1974 (562)