SUMMA
BRASILIENSIS
MATHEMATICÆ

ON INTEGERS OF THE FORM $2^k + p$ AND SOME RELATED PROBLEMS

BY
P. ERDÖS

ON INTEGERS OF THE FORM 2^k+p AND SOME RELATED PROBLEMS *

By *P. Erdös*

Romanoff [1] proved that the integers of the form 2^k+p have positive density. In other words, the number of integers $\leq x$ of the form 2^k+p is greater than $c_1 x$. Throughout this paper the c's denote positive absolute constants. This result is surprising since it follows from the prime number theorem (or a more elementary result of Tchebicheff) that the number of solutions in k and p of $2^k+p \leq x$ is less than $c_2 x$. Romanoff [1] proved in fact the following result: Denote by $f(n)$ the number of solutions of $2^k+p=n$. Then

$$(1) \qquad \lim . \sup. \frac{1}{x} \sum_{n=1}^{x} f^2(n) < \infty .$$

The fact that the numbers of the form 2^k+p have positive density follows immediately from (1), Schwartz's inequality and the fact that the number of solutions in k and p of $2^k+p \leq x$ is $> c_3 x$.

Following a question of Turán [2] we first prove the following

THEOREM 1. $\lim.\sup f(n) = \infty$. *In fact for infinitely many n*

$$(2) \qquad f(n) > c \; \log \log n.$$

THEOREM 2. *For every k*

$$(3) \qquad \lim .\sup . \frac{1}{x} \sum_{n=1}^{x} f^k(n) < \infty .$$

Also following a question of Romanoff [3] we prove

THEOREM 3. *There exists an arithmetic progression consisting only of odd numbers, no term of which is of the form 2^k+p.*

Finally we prove

THEOREM 4. *Let $a_1 < a_2 < \ldots$ be an infinite sequence of integers satisfying $a_k \mid a_{k+1}$. Then the necessary and sufficient condition that the sequence $p + a_k$ should have positive density is that*

$$(4) \qquad \qquad \lim.\sup. \frac{\log a_k}{k} < \infty \; ,$$

$$(5) \qquad \qquad \sum_{d \mid a_i} \frac{1}{d} < c_5 \; .$$

Finally we discuss a few related unsolved problems.

Proof of Theorem 1. Define A as the product of the odd consecutive primes $< (\log x)^{1/2}$. It follows from the results of Tchebicheff[4] on the function $\theta(x)$ that

$$A < \exp \theta \left[(\log x)^{1/2} \right] < \exp 2 \, (\log x)^{1/2} \; ,$$

where $\exp z = e^z$.

Denote by S the number of solutions of

$$2^k + p \equiv 0 \pmod{A}, \quad 2^k + p \leq x \; .$$

Let $k \leq \log x$, then if $p < x/2$, $2^k + p < x$. The number of primes satisfying

$$p < x/2, \quad p \equiv -2^k \pmod{A}$$

is, by the result of Rodosskii [5], greater than

$$[1 + o(1)] \frac{x}{2\varphi(A) \log x} = [1 + o(1)] \frac{x}{2 \, A A^* \log x} \cdot c_0 \frac{x \log \log x}{A \log x} \; ,$$

where

$$A^* = \prod_{p \mid A} \left(1 - \frac{1}{p} \right) = 0 \left(\frac{1}{\log \log x} \right) \quad (\text{see } ^6) \; .$$

Thus by summing on k we obtain

$$(6) \qquad \qquad S > c_6 \, x \log \log x / A \; .$$

[4] See e.g. the first few pages of Ingham's Cambridge tract on the distribution of primes or Hardy and Wright's Number Theory.

[5] On the distribution of prime numbers in short arithmetic progressions. Izvestiya Akad. Nauk. S.S.S.R. Ser. Mat. 12, 123-128 (1948).

[6] 4 ibid.

But

$$(7) \qquad S = \sum_{k=1}^{x/A} f(kA).$$

Hence from (6) and (7) we obtain that for some multiple lA of A not exceeding x we have $f(lA) > c_6 \log \log x$, which proves Theorem 1. If we only want to prove that $\lim \sup f(n) = \infty$, the application of the prime number theorem for arithmetic progressions would have been sufficient.

It can be conjectured that $f(n) = o(\log n)$. This if true is probably rather deep. I cannot even prove that for all sufficiently large n not all the integers

$$(8) \qquad n - 2^k, \qquad 1 \le k < \frac{\log n}{\log 2}$$

can be prime. For $n = 105$ all the integers (8) are prime, but it is easy to see from a study of the prime tables that in the interval $105 < n \le 3 . 5^2 . 11 . 13 . 19 = 203775$ there is no other such integer. It seems likely that 105 is the largest exceptional integer. I believe that the following result holds: Let c be any constant and n sufficiently large, $a_1 < a_2 < \ldots < a_x \le n$, $x > \log n$. Then there exists an m so that the number of solutions of $m = p + a_i$ is greater than c. This would be a generalization of Theorem 1.

Proof of Theorem 2. Denote by $\varphi(x; i_1, \ldots, i_k)$ the number of solutions of the equations

$$(9) \qquad p_{i_1} + 2^{i_1} = p_{i_2} + 2^{i_2} = \ldots = p_{i_k} + 2^{i_k}$$

in primes $p_{i_r} < x$. A simple argument shows that

$$(10) \qquad \sum_{n=1}^{x} f^k(n) < k^k \left[\Sigma \, \varphi(x; i_1, \ldots, i_k) + x \right]$$

where the summation is extended over all the distinct i's satisfying $2^i \le x$. (The first term of the right side of (10) comes from the n with $f(n) \ge k$, the second term from the n with $f(n) < k$.)

Let $g < x^{c_7}$, g prime, c_7 sufficiently small, and g does not divide $D_{i_1} \ldots_{i_k} = \Pi \, (2^i U - 2^i V)$, $1 \le V < U \le k$. Then if $p_{i_1} > x^{c_7}$ satisfies (9), we clearly must have

$$p_{i_1} \equiv \cdot - (2^{i_1} - 2^{i_r}) \quad (\text{mod } g) \quad \text{false for} \quad r = 1, 2, \ldots, k.$$

(since $p_{i_1} + 2^{i_r} - 2^{i_1} = p_{i_r}$ is a prime). Since $g/D_{i_1} \ldots_{i_k}$ is false, these k residues are all different. Thus by Brun's method [7]

$$(11) \qquad \varphi(x; i_1, \ldots, i_k) < x^{c_7} + c_8 \, x \, \Pi_1 \left(1 - \frac{k}{g} \right) <$$

$$< c_9 \frac{x}{(\log x)^k} \Pi_2 \left(1 + \frac{k}{g} \right),$$

where Π_1 denotes the product for all $g < x^{c_7}$, $g \,/\, D_{i_1} \ldots_{i_k}$ false and Π_2 the product for all $g \,/\, D_{i_1} \ldots_{i_k}$.

Now by the inequality of the geometric and arithmetic means

$$\Pi_2 \left(1 + \frac{k}{g} \right) \le \frac{2}{k \, (k-1)} \, \Sigma_1 \Pi_3 \left(1 + \frac{k}{g} \right)^{k \, (k-1)/2},$$

where Σ_1 denotes the sum subject to the conditions $1 \le V < U \le k$ and Π_3 the product for all $g \,/\, 2^i U - 2^i V$.

Thus by a simple argument

$$(12) \quad \Sigma_2 \Pi_2 \left(1 + \frac{k}{g} \right) < c_{10} \, (\log x)^{k-2} \, \Sigma_3 \, \Pi_4 \left(1 + \frac{k}{g} \right)^{k \, (k-1)/2},$$

on the left side of (12) the summation Σ_2 being extended over the distinct sets of i's satisfying $2^i \le x$ and on the right side Σ_3 denoting the sum subject to

$$1 \le v < u \le \frac{\log x}{\log 2}$$

and Π_4 the product over all $g \,/\, 2^u - 2^v$. Now trivially for a fixed $B = B \, (k)$

$$(13) \qquad \left(1 + \frac{k}{g} \right)^{k \, (k-1)/2} < 1 + \frac{B}{g}.$$

(7) See e.g. P. Erdös, Proc. Cambridge Phil. Soc. 33, 6-12 (1937).

Thus to prove Theorem 2 it will suffice to show that (by (10), (11), (12) and (13))

$$\Sigma_4 \Pi_4 \left(1 + \frac{B}{g} \right) < c_{11} \log x .$$

where Σ_4 denotes the sum subject to $2^u \leq x$ and Π_4 refers to the product extended over all $g \,/\, 2^u - 1$.

Now if $l_2(d)$ denotes the smallest exponent z for which $2^z \equiv 1 \pmod d$ and $v(d)$ denotes the number of distinct prime factors of d, then by a simple argument (interchanging the order of summation)

$$\Sigma_4 \Pi_4 \left(1 + \frac{B}{g} \right) < \Sigma_4 \, \Sigma_5 \, \frac{B^{v(d)}}{d} < c_{12} \log x \sum_{d=1}^{\infty} \frac{B^{v(d)}}{d \, l_2(d)}$$

where Σ_5 refers to the sum subject to $d \,/\, 2^u - 1$.

Thus to complete the proof of Theorem 2 we have to prove that

$$(14) \qquad\qquad \sum_{d=1}^{\infty} \frac{B^{v(d)}}{d \, l_2(d)} < \infty .$$

The proof of (14) will be very similar to the proof given by Turán and myself [8] for the theorem of Romanoff $\sum_{d=1}^{\infty} \frac{1}{d \, l_2(d)} < \infty$.

To prove (14) we split the integers into two classes. In the first class are the integers with $l_2(d) < (\log d)^{c_{13}}$, $c_{13} > B$ and in the second class are the other integers. Clearly the integers of the first class are all composed of the prime factors of the set

$$(15) \qquad\qquad 2^k - 1 , \qquad 1 < k \leq (\log x)^{c_{13}} .$$

Since the number of prime factors of $2^k - 1$ is clearly $< k$, the number of the prime factors of the sequence (15) is clearly $< (\log x)^{2c_{13}}$. Denote the prime factors of the sequence (15) by $p_1, p_2, \ldots, p_r$, $r < (\log x)^{2c_{13}}$. Clearly the number of integers of the first class is less than the number of integers $\leq x$ composed of $p_1, p_2, \ldots, p_r$. We split the integers $\leq x$ composed entirely of the p's into two groups. In the first group are the integers having less than

$$D = \log x \,/\, 4 \cdot c_{13} \cdot \log \log x$$

different prime factors and in the second group are the other integers composed of the p's. The number of integers of the first group is clearly less than

$$(16) \quad \left[\binom{r}{D} + \binom{r}{D-1} + \ldots + \binom{r}{0} \right] \times$$
$$\times \left(\frac{\log x}{\log 2} \right)^D < r^D \left(\frac{\log x}{\log 2} \right)^D < x^{3/4} \text{ for } x > x_0 .$$

The factor $(\log x / \log 2)^D$ comes from the fact that the number of powers $p^s \leq x$ equals $\left[\dfrac{\log x}{\log 2} \right] \leq \dfrac{\log x}{\log 2}$.

The number of integers of the second group is clearly less than

$$(17) \quad \frac{x}{D!} \left(\sum_{i=1}^{r} \sum_{s} \frac{1}{p_i^s} \right)^D < \frac{x}{D!} (c_{11} \log \log \log x)^D < x^{1-\epsilon}$$

since $D! > x^\delta$ and

$$\sum_{i=1}^{r} \sum_{s} \frac{1}{p_i^s} < c_{15} \log \log r < c_{14} \log \log \log x .$$

Thus from (16) and (17) we obtain that the number of integers $\leq x$ of the first class is less than $2x^{1-\epsilon}$ for sufficiently large x. Denote by $d_1 < d_2 < \ldots$ the integers of the first class. We evidently have for sufficiently large k

$$(18) \quad d_k > x^{1+\epsilon/2} .$$

Since $v(d) < c_{16} \log d / \log \log d$, we obtain by (18)

$$(19) \quad \Sigma \frac{B^{v(d)}}{d \, l_2(d)} < c_{17} \sum_{k=1}^{\infty} \frac{1}{k^{1+\epsilon/4}} < \infty$$

where in the left side of (19) the summation is extended over the integers of the first class, i.e. the integers satisfying $l_2(d) \leq (\log d)^{c_{13}}$.

Now we have to show that

$$(20) \quad \Sigma \frac{B^{v(d)}}{d \, l_2(d)} < \infty$$

where in (20) the summation is extended over the integers satisfying $l_2(d) > (\log d)^{c_{13}}$. In fact we shall prove that

$$(21) \qquad \sum_{d=1}^{\infty} \frac{B^{v(d)}}{d \, (\log d)^{c_{13}}} < \infty$$

and this clearly implies (20). By partial summation we obtain that to prove (21) it will suffice to show that

$$(22) \qquad \sum_{x=1}^{\infty} \frac{\sigma_x}{x^2 \, (\log x)^{c_{13}}} < \infty \qquad \text{where} \qquad \sigma_x = \sum_{d=1}^{x} B^{v(d)} \, .$$

We have

$$\sum_{d=1}^{x} 2^{v(d)} = \sum_{\substack{l=1 \\ \mu(l) \neq 0}}^{x} \left[\frac{x}{l} \right] = 0 \, (x \log x)$$

Assume now that

$$\sum_{d=1}^{x} (B-1)^{v(d)} = 0 \left[x \, (\log x)^{B-2} \right] \, .$$

We then have by induction, interchanging the order of summation and partial summation

$$\sigma_x = \sum_{d=1}^{x} \left[(B-1) + 1 \right]^{v(d)} = \sum_{\substack{l=1 \\ \mu(l) \neq 0}}^{x} (B-1)^{v(l)} \left[\frac{x}{l} \right] <$$

$$< x \sum_{l=1}^{x} \frac{(B-1)^{v(l)}}{l} < c \, x \sum_{l=1}^{x} \frac{l \, (\log l)^{B-2}}{l^2} = 0 \left[x \, (\log x)^{B-1} \right] \, .$$

Thus (22) converges for $c_{13} > B$. Hence from (19) and (20) it follows that (14) converges, which completes the proof of Theorem 2.

Proof of Theorem 3. Clearly every integer satisfies at least one of the following congruences: 0 (mod 2), 0 (mod 3), 1 (mod 4), 3 (mod 8), 7 (mod 12), 23 (mod 24). Therefore, if x is congruent to: 1 (mod 2), 1 (mod 7), 2 (mod 5), 2^3 (mod 17), 2^7 (mod 13), 2^{23} (mod 241), then for any $k, x-2^k$ is a multiple of one of the primes 3, 5, 7, 13, 17, 241. This proves Theorem 3.

The reason for the success of this proof is that, by a well known theorem ([9]), for any $n \neq 6$ there exists a prime p satisfying $p \mid 2^n - 1$ and $p \nmid 2^m - 1$ false for all $m < n$.

The simplest system of congruences $a_i \pmod{n}$, $n_1 < n_2 < \ldots < n_k$ so that every integer satisfies at least one of them is: $0 \pmod 2$, $0 \pmod 3$, $1 \pmod 4$, $5 \pmod 6$, $7 \pmod{12}$. But because of the modulus 6 this system cannot be used to prove Theorem 3.

The following such system does not contain the modulus 2: $0 \pmod 3$, $0 \pmod 4$, $0 \pmod 5$, $1 \pmod 6$, $6 \pmod 8$, $3 \pmod{10}$, $5 \pmod{12}$, $11 \pmod{15}$, $7 \pmod{20}$, $10 \pmod{24}$, $2 \pmod{30}$, $34 \pmod{40}$, $59 \pmod{60}$, $98 \pmod{120}$. Davenport found a slightly more complicated system somewhat earlier. It seems likely that for every c there exists such a system all the moduli of which are $> c$. This would imply, by the same argument which proved Theorem 3, that for every c' there exists an arithmetic progression no term of which is of the form $2^k + u$, where $v(u) < c'$.

Proof of Theorem 4. The proof of the necessity is easy. If (4) is not satisfied then for a suitable sequence n_i the number of the $a_k \leq n_i$ is $o(\log n_i)$ and since $\pi(n_i) < 2n_i / \log n_i$ we obtain that the number of the integers of the form $p + a_k \leq n_i$ is $o(n_i)$, or (4) is necessary.

To show the necessity of (5), assume that (5) does not hold. Let A be large and j chosen so large that $\sum_{d/a_j} 1/d > A$. Let n be sufficiently large. We split the integers of the form $p + a_k \leq n$ into three classes. In the first class are the integers $\leq n$ of the form

$$p + a_k, \qquad k \leq j \, ;$$

the number of these integers is less than $j \, \pi(n) = o(n)$.

The integers of the second class satisfy $p \mid a_j$. The number of these integers is clearly less than $v(a_j)$ times the number of the a's not exceeding n, thus it is $O(\log n) = o(n)$.

In the third class are all the other integers of the form $p + a_k \leq n$. Clearly from $a_j \mid a_k$ for $j < k$ and $(p, a_j) = 1$, we obtain that these

---

integers are all relatively prime to a_j. Thus the number of integers of the third class is less than

$$n \prod_{r/a_j} \left(1 - \frac{1}{p} \right) + 0(1) < n \left(\sum_{d/a_j} \frac{1}{d} \right)^{-1} < \frac{n}{A},$$

which completes the proof of the necessity of (4) and (5).

Now we prove the sufficiency of (4) and (5). We shall prove that the number of distinct integers of the form

$$p + a_k \leq n, \qquad k \leq c_{13} \log n$$

is greater than $c_{19} n$. First of all it follows from (4) that if c_{18} is sufficiently small then $a_k < n/2$ for $1 \leq k < c_{18} \log n$. Now we estimate from below the number of integers of the form $p + a_k \leq n$ which are not of the form $p + a_j$, $1 \leq j < k$. If $p_1 + a_k = p_2 + a_j$, then

$$p_2 - p_1 = a_k - a_j.$$

By a result of Schnirelmann ([9]) the number of solutions of this equation in primes $p_1 \leq n$, $p_2 \leq n$ is less than (by(5))

$$(23) \qquad c_{20} \frac{n}{(\log n)^2} \prod_5 \left(1 + \frac{1}{p} \right) < c_{20} \frac{n}{(\log n)^2} \sum_6 \frac{1}{d} <$$
$$< c_{21} \frac{n}{(\log n)^2} \sum_7 \frac{1}{d}$$

where $\prod_5$ is extended over all $p/a_k - a_l$, $\sum_6$ refers to the sum extended over all $d \mid a_k - a_l$ and $\sum_7$ is extended over all $d \mid a_k - a_l$, $(d, u_k) = 1$ Now we make use of the following

LEMMA. *Let* $a_1 < a_2 < \ldots$, a_k/a_{k+1}, *and assume that* (4) *and* (5) *are satisfied. Then there exists an absolute constant* c_{22} *so that*

$$\sum_{l < k} \sum_7 \frac{1}{d} < c_{22} k.$$

Let us assume that the lemma is already proved. Since $a_k < n/2$ the number of integers of the form $p + a_k \leq n$ is by the results of

Tchebicheff greater than $n/4 \log n$. Thus from (23) and our lemma we obtain that the number of integers $\leq n$ of the form $p + a_k$ which are not of the form $p + a_j$, $1 \leq j < k$ is greater than

$$\frac{n}{4 \log n} - c_{21} c_{22} c_{18} \frac{n}{\log n} > \frac{n}{5 \log n}$$

for sufficiently small c_{18}. Thus the number of integers $\leq n$ of the form $p + a_k$, $k \leq c_{18} \log n$ is greater than $c_{18} n/5 > c_{19} n$ which proves Theorem 4.

Thus to complete the proof of Theorem 4 we only have to prove our lemma. We split the integers d, $(d, a_k) = 1$, into two classes. In the first class are the d's for which there are less than $k/(\log d)^2$ values of l with $a_k - a_l \equiv 0 \pmod d$, and in the second class are the other d's. We evidently have

$$(24) \qquad \sum_{l < k} \Sigma_8 \frac{1}{d} < k \, \Sigma_8 \frac{1}{d \, (\log d)^2} < c_{23} \, k \; ,$$

where in Σ_8 the summation is extended over the d's of the first class. Next we prove that for sufficiently large y the number of d's of the second class not exceeding y is less than $y/(\log y)^2$. Let $d \leq y$ be any integer of the second class. Denote by $1 \leq l_1 < l_2 < \ldots < l_r \leq k$ the l's satisfying

$$(25) \qquad a_k - a_{l_i} \equiv 0 \pmod d, \quad r > \frac{k}{(\log d)^2} \geq \frac{k}{(\log y)^2} \; .$$

Clearly for at least $\dfrac{1}{2} \dfrac{k}{(\log y)^2}$ of the l_i's we have

$$(26) \qquad l_i - l_{i-1} \leq 2 \, (\log y)^2 \; .$$

We consider only these l_i's. Further we consider only the l_i's satisfying

$$(27) \qquad a_{l_i} / a_{l_{i-1}} < \exp (\log y)^3 \; .$$

We evidently have by (4)

$$\prod_{i=1}^{r} a_{l_i} / a_{l_{i-1}} \leq a_k < c_{24}{}^k \; .$$

Thus if U denotes the number of l's which do not satisfy (27), we have

(28) $\exp U (\log y)^3 < c_{24}{}^k$ or $U = o (k / (\log y)^2)$

Denote now by $l_{j_1}, l_{j_2}, ..., l_{j_t}$ the l's which satisfy (26) and (27). We evidently have $t > k / 4 (\log y)^2$. Therefore if d runs through all the integers of the second class (and if we assume that the number of integers of the second class is greater than $y/(\log y)^2$) we obtain at least $ky/4 (\log y)^4$ differences $l - l_{i-1}$ for which l satisfies (26) and (27). But these differences are all $< 2 (\log y)^2$. Thus there must exist a fixed l_1 and l_2 satisfying (26) and (27) which occurs for at least $y/8 (\log y)^6$ d's. In other words the integer T

$$T = a_{l_2} / a_{l_1} - 1 < \exp (\log y)^3$$

has at least $y/8 (\log y)^6$ divisors not exceeding y. But this is easily seen to be impossible. Denote by $p_1, p_2, ..., p_s$ the prime factors of T. Clearly $s < (\log y)^3$. Clearly all the d's dividing T must be $\leq y$ and composed of the p's. We split these integers into two classes. In the first class are the integers having less than $100 \log \log y$ different prime factors and in the second class are the other integers $\leq y$ composed of the p's. The number of integers of the first class is less than (by the same argument as used in proving (16))

$$(\log y)^{3.100 \log \log y} \left(\frac{\log y}{\log 2} \right)^{100 \log \log y} < y^{1/2} .$$

If m is an integer of the second class then $d(m) \geq 2^{100 \log \log y}$. Thus from $\sum_{m=1}^{y} d(m) = 0 (y \log y)$ we obtain that the number of integers of the second class is $y/(\log y)^{10}$. Thus T has less than $y/(\log y)^{10} + y^{1/2} < y/8 (\log y)^6$ divisors for $y > y_0$. This contradiction proves the lemma and therefore Theorem 4 is proved.

Theorem 4 clearly generalizes Romanoff's ([1]) result according to which the density of the integers of the form 2^k+p is positive.

UNIVERSITY OF ABERDEEN SCOTLAND

SUMMA BRASILIENSIS MATHEMATICAE

Vol. I (1945-1946)

Vol. II (1947-....)

ANAIS DA ACADEMIA BRASILEIRA DE CIENCIAS