SOME PROBLEMS ON CONSECUTIVE PRIME NUMBERS

P. ERDŐS

Let $2 = p_1 < p_2 < \ldots$ be the sequence of consecutive prime numbers. Put $d_n = p_{n+1} - p_n$. Turán and I proved [**1**] that the inequalities $d_{n+1} > d_n$ and $d_{n+1} < d_n$ both have infinitely many solutions. It is not known if $d_n = d_{n+1}$ has infinitely many solutions. The answer is undoubtedly affirmative but the proof will probably be very difficult [**2**]. It was a great surprise and disappointment to us that we could not prove that $d_{n+2} > d_{n+1} > d_n$ has infinitely many solutions. We could not even prove that $(-1)^n (d_{n+1} - d_n)$ changes sign infinitely often. It seems certain that the answer to both of these questions is affirmative and perhaps a simple proof can be found.

I proved that for a certain $c > 0$ ($c, c_1, c_2 \ldots$ will denote positive absolute constants) $d_{n+1} > (1+c)d_n$ and $d_{n+1} < (1-c)d_n$ both have infinitely many solutions [**3**]. An obvious guess is that d_{n+1}/d_n is everywhere dense in $(0, \infty)$, a proof of Ricci and myself only gives that the set of limit points has positive measure, but we can not find a single element of this set of positive measure.

Turán and I also asked: Let $a_1, \ldots, a_k$ be real numbers. What is the necessary and sufficient condition that

$$\sum_{i=1}^{k} a_i p_{n+i}, n = 1, 2, \ldots \tag{1}$$

should have infinitely many changes of sign? We observed that $\sum_{i=1}^{k} a_i = 0$ is clearly a necessary condition and Pólya observed that if (1) has infinitely many changes of sign then the k numbers $\alpha_j = \sum_{i=1}^{j} a_i$ cannot all have the same sign. Put $\left(\sum_{i=1}^{k} a_i = 0\right)$

$$\sum_{i=1}^{k} a_i p_{n+i} = -\sum_{i=1}^{k-1} \alpha_i d_{n+i}. \tag{2}$$

It would be reasonable to conjecture that Pólya's condition is necessary and sufficient for (2) to change sign infinitely often. Unfortunately the proof of this is not likely to succeed at the present stage of science. It is easy to see that the conjecture is equivalent with the following conjecture: for every i,

$$\overline{\lim_{n\to\infty}} \frac{d_n}{\sum_{\substack{j=-i \\ j\neq 0}}^{i} d_{n+j}} = \infty. \tag{3}$$

(3) seems quite hopeless to me at present. As stated previously it is not even known that $\overline{\lim}_{n=\infty} d_n/d_{n+1} = \infty$. As far as I know it has not yet been proved that

$$d_n > d_{n+1} + d_{n+2}$$

[MATHEMATIKA **19** (1972), 91–95]

has infinitely many solutions. Perhaps by sieve method techniques this could be done, but I have not succeeded in proving it.

In this note we prove the following much more modest (we can clearly assume $k \geqslant 3$)

THEOREM 1. *Assume* $\sum_{i=1}^{k-1} \alpha_i = 0$ *and* $\alpha_{k-1} \neq 0$. *Then* (2) *changes sign infinitely often.*

Theorem 1 clearly implies our old result with Turán that $d_{n+1} - d_n$ changes sign infinitely often.

Before we prove our Theorem we state one more question: Turán and I also proved [1] that the determinant

$$\begin{vmatrix} p_n, & p_{n+1} \\ p_{n+1}, & p_{n+2} \end{vmatrix}$$

changes sign infinitely often. Presumably this could be generalized to show that for every k the determinant

$$\begin{vmatrix} p_n, & \ldots, p_{n+k-1} \\ \vdots & \vdots \quad \vdots \\ p_{n+k-1}, & \ldots, p_{n+2k-2} \end{vmatrix}$$

changes sign infinitely often.

Now we prove Theorem 1. We have by a simple argument ($k \geqslant 3$)

$$\left| \sum_{n=1}^{x} \sum_{i=1}^{k-1} \alpha_i d_{n+i} \right| < \sum_{i=1}^{k-1} |\alpha_i| \, |d_1 + \ldots + d_{k-1} + d_{x+1} + \ldots + d_{x+k-1}|.$$

From the prime number theorem (or a more elementary theorem) it follows that for infinitely many x, $d_{x+k-1} - d_x < c_1 \log x$. Thus for infinitely many x

$$\left| \sum_{n=1}^{x} \sum_{i=1}^{k-1} \alpha_i d_{n+i} \right| < c_2 \log x. \tag{4}$$

Thus if (2) would change sign only finitely often we would obtain that for arbitrarily large values of x

$$\left| \sum_{i=1}^{k-1} \alpha_i d_{n+i} \right| < |\alpha_{k-1}| \tag{5}$$

for all but $2c_2 \log x$ values of $n \leqslant x$. But this clearly implies that, for all but $2c_2 \log x$ values of n, $\{d_{n+1}, \ldots, d_{n+k-2}\}$ uniquely determines d_{n+k-1}. Hence (5) implies that there are at most

$$(10k \log x)^{k-2} + 2c_2 \log x < c_3 (10k)^{k-2} (\log x)^{k-2} \tag{6}$$

$(k-1)$-tuples

$$\{d_{n+1}, \ldots, d_{n+k-1}\}, \qquad \max_{1 \leqslant i \leqslant k-1} d_{n+i} < 10k \log x, \; 1 \leqslant n \leqslant x.$$

Now we show

THEOREM 2. *The number of* $(k-1)$*-tuples*

$$\{d_{n+1}, \ldots, d_{n+k-1}\}, \qquad \max_{1 \leqslant i \leqslant k-1} d_{n+i} < 10k \log x,\ 1 \leqslant n \leqslant x$$

is greater than $c_4(\log x)^{k-1}$.

Since $k \geqslant 3$ Theorem 2 contradicts (6) and hence (5); thus Theorem 2 implies Theorem 1.

In Theorem 2, $10k$ can certainly be replaced by a smaller constant but it will be difficult to replace $10k$ by a constant independent of k.

It seems certain that for every k

$$\liminf_{n=\infty} \frac{\max(d_{n+1}, \ldots, d_{n+k})}{\log n} = 0 \tag{7}$$

and

$$\limsup_{n=\infty} \frac{\min(d_{n+1}, \ldots, d_{n+k})}{\log n} = \infty; \tag{8}$$

(7) and (8) seem to be very deep and I could only prove (8) for $k = 2$ [**4**].

It might be possible to give an elementary proof of Theorem 1 (perhaps following our proof with Turán for the case $k = 3$), but it seems to me that Theorem 2 has independent interest. In proving Theorem 2 we use Brun's method; better values of the constants could of course be obtained by using Selberg's sieve. We need the following

LEMMA 1. *The number of primes* $p < x$ *for which all the* $k-1$ *integers*

$$p + \sum_{j=1}^{i} d_j,\ 1 \leqslant i \leqslant k-1$$

are primes is less than

$$\frac{c_5 x}{(\log x)^k} \Pi' \left(1 + \frac{k}{p}\right), \tag{9}$$

where the primes in Π' *run through all the prime factors of*

$$A(d_1, \ldots, d_{k-1}) = \prod_{1 \leqslant i_1 \leqslant i_2 \leqslant k-1} \left(\sum_{j=i_1}^{i_2} d_j\right). \tag{10}$$

The Lemma immediately follows from Brun's method. The sieving primes are the primes $p < x^\delta$, $p \nmid A(d_1, \ldots, d_{k-1})$. For these primes the k residues

$$0, d_1, d_1 + d_2, \ldots, d_1 + \ldots + d_{k-1}$$

are all distinct. Thus Brun's method and the well known result

$$\prod_{p < x^\delta} \left(1 - \frac{k}{p}\right) < c_\delta \frac{x}{(\log x)^k}$$

immediately give Lemma 1.

LEMMA 2. *To every $\varepsilon > 0$ there is a δ so that if S is any set of fewer than δy^{k-1} $(k-1)$-tuples of integers $1 \leqslant d_1, \ldots, d_{k-1} < y$ we have*

$$\sum{}' \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{k}{p}\right) < \varepsilon y^{k-1}, \tag{11}$$

where the dash indicates that $d_1, \ldots, d_{k-1}$ runs through the $(k-1)$-tuples of S and $A(d_1, \ldots, d_{k-1})$ is given by (10).

The proof of Lemma 2 follows standard techniques but we give all the details. First we show

$$\sum_{\max d_i \leqslant y} \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{k}{p}\right)^2 < c_k y^{k-1}. \tag{12}$$

Since $(1+k/p)^2 < (1+3k^2/p)$, to prove (12) it suffices to show that

$$\sum_{\max d_i \leqslant y} \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{l}{p}\right) < c_l y^{k-1}, \tag{13}$$

with $l = 3k^2$.

By interchanging the order of summation we evidently have

$$\sum_{\max d_i \leqslant y} \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{l}{p}\right) = \sum \frac{l^{v(t)} F(t)}{t}, \tag{14}$$

where $v(t)$ denotes the number of prime factors of the squarefree integer t and $F(t)$ denotes the number of solutions of

$$A(d_1, \ldots, d_{k-1}) \equiv 0 \pmod t, \ \max d_i \leqslant y.$$

Denote by p_t the greatest prime factor of t. It is easy to see that

$$F(t) \leqslant F(p_t) < 2^{k-1} \frac{(2y)^{k-1}}{p_t} < 4^k \frac{y^{k-1}}{p_t}. \tag{15}$$

From (14) and (15) we have (p and q are primes)

$$\sum_{\max d_i \leqslant y} \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{l}{p}\right) < c_k y^{k-1} \sum_{t=1}^{\infty} \frac{l^{v(t)}}{t p_t}$$
$$< c_k y^{k-1} \sum_p \frac{l}{p^2} \prod_{q<p} \left(1 + \frac{l}{q}\right) < c_k y^{k-1} \sum_p \frac{c_6{}^l (\log p)^l}{p^2} < c_l y^{k-1}$$

which completes the proof of (13) and (12).

Lemma 2 easily follows from (12). If (11) did not hold for a certain set having fewer than δy^{k-1} k-tuples we would have from Cauchy's inequality

$$\sum_{\max d_i \leqslant y} \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{k}{p}\right)^2 > \sum{}' \prod_{p|A(d_1, \ldots, d_{k-1})} \left(1 + \frac{k}{p}\right)^2$$
$$\geqslant \frac{\left(\sum{}' \prod_{p|A(d_1, \ldots, d_k)} \left(1 + \frac{k}{p}\right)\right)^2}{\sum 1} > \frac{(\varepsilon y^{k-1})^2}{\delta y^{k-1}} = \frac{\varepsilon^2}{\delta} y^{k-1},$$

which contradicts (12) for $\delta = \varepsilon^2/c_k$ and hence Lemma 2 is proved.

Now we easily obtain Theorem 2. It follows from the prime number theorem (or a more elementary theorem) that

$$p_{n+k} - p_{n+1} < 10k \log x$$

has more than $(x/2 \log x)$ solutions in primes $p_n < x$.

Thus the primes $p_n < x$ give at least $(x/2 \log x)$ $(k-1)$-tuples

$$\{d_{n+1}, \ldots, d_{n+k-1}\}, \max_{1 \leqslant i \leqslant k-1} d_{n+1} < 10k \log x, \tag{16}$$

where multiple occurrences are counted multiply. By Lemma 1 a $(k-1)$-tuple can occur with a multiplicity at most

$$\frac{c_5 x}{(\log x)^k} \Pi' \left(1 + \frac{k}{p}\right)$$

and by Lemma 2 ($y = 10k \log x$) there are at least $c_7 (\log x)^{k-1}$ distinct $(k-1)$-tuples satisfying (16) if c_7 is sufficiently small. This completes the proof of Theorems 2 and 1.

References

1. P. Erdős and P. Turán, " On some new questions on the distribution of primes ", *Bull. Amer. Math. Soc.*, 44 (1948), 271–278.
2. ———, and A. Rényi, " Some problems and results on consecutive primes ", *Simon Stevin* (1949), 115–125.
3. ———, " On the difference of consecutive primes ", *Bull. Amer. Math. Soc.*, 44 (1948), 885–889.
4. ———, " Problems and results on the difference of consecutive primes ", *Publicationes Delruen*, 1 (1949), 37–39.

Department of Combinatories and Optimisation,
University of Waterloo,
Canada.

10H20: *Number theory; Multiplicative theory; Distribution of primes.*

(*Received on the* 8*th of November*, **1971.**)

Printed by C. F. Hodgson & Son, Ltd., 50 Holloway Road, London N7 8JL.