ON THE DIFFERENCE OF CONSECUTIVE PRIMES

P. ERDÖS

The present paper contains some elementary results on the difference of consecutive primes. Theorem 2 has been announced in a previous paper.[1] Also some unsolved problems are stated.

Let $p_1=2$, $p_2=3, \cdots, p_k, \cdots$ be the sequence of consecutive primes. Put $d_k=p_{k+1}-p_k$. We have:

THEOREM 1. *There exist positive real numbers c_1 and c_2, $c_1<1$, $c_2<1$, such that for every n the number of k's satisfying both*

$$(1)\qquad d_{k+1} > (1+c_1)d_k, \qquad k \leqq n,$$

and the number of l's satisfying both

$$(2)\qquad d_{l+1} < (1-c_1)d_l, \qquad l \leqq n,$$

are each greater than $c_2 n$.

We shall prove Theorem 1 later. From Theorem 1 we easily deduce:

THEOREM 2. *For every t and all sufficiently large n the number of solutions in k and l of each of the two sets of inequalities*

$$(3)\qquad \left(\frac{p_{k+1}^t + p_{k-1}^t}{2}\right)^{1/t} > p_k,\ k \leqq n; \qquad \left(\frac{p_{l+1}^l + p_{l-1}^t}{2}\right)^{1/t} < p_l,\ l \leqq n,$$

is greater than $(c_2/2)n$.

Let ϵ be sufficiently small but fixed. It is well known that $p_n < 2\cdot n \log n$. Thus the number of $k \leqq n$, with $p_{k+1}>(1+\epsilon)p_k$, is less than $c \log n$. Hence it follows from Theorem 1 that the number of k's satisfying

$$(4)\qquad p_{k+1} < (1+\epsilon)p_k, \qquad d_k > (1+c_1)d_{k-1}, \qquad k \leqq n,$$

is greater than $(c_2/2)n$. A simple calculation now shows that the primes satisfying (4) also satisfy the first inequality of (3) if $\epsilon=\epsilon(c_1)$ is chosen small enough. The second inequality of (3) is proved in the same way, which proves Theorem 2.

Further, we obtain, as an immediate corollary of Theorem 1, that[2]

Received by the editors October 17, 1947.

[1] P. Erdös and P. Turán, *Some new questions on the distribution of primes*, Bull. Amer. Math. Soc. vol. 54 (1948) pp. 371–378.

[2] This result was also stated in the above paper.

$$\limsup d_{k+1}/d_k > 1, \qquad \liminf d_{k+1}/d_k < 1.$$

At present I can not decide whether $d_{k+2} > d_{k+1} > d_k$ has infinitely many solutions. The following question might be of some interest: Let $\epsilon_n = 1$ if $d_{n+1} > d_n$, otherwise $\epsilon_n = 0$. It may be conjectured that $\sum_{n=1}^{\infty} \epsilon_n/2^n$ is irrational. I can not even prove that from a certain point on ϵ_n is not alternatively 1 and 0.

In order to prove Theorem 1 we need two lemmas.

LEMMA 1. *For sufficiently small $c_1 > 0$ the number of solutions in k of the inequalities*

$$1 + c_1 > d_{k+1}/d_k > 1 - c_1, \qquad k \leqq n, \tag{5}$$

is less than $n/4$.

Denote by $g(n; a, b)$ the number of solutions of the simultaneous equations

$$d_{k+1} = a, \qquad d_k = b, \qquad k \leqq n.$$

Denote by V the number of primes $r < 2 \cdot n \cdot \log n$ for which $r+a$ and $r+a+b$ are also primes. Since $p_n < 2 \cdot n \cdot \log n$, we evidently have

$$g(n; a, b) \leqq V. \tag{6}$$

Now let $c_1 > 0$ be sufficiently small and $q_1, q_2, \cdots$ run through the primes less than n^{c_3}. Then V is not greater than n^{c_3} plus the number U of integers $m < 2 \cdot n \cdot \log n$, which satisfy, for all i,

$$m \not\equiv 0 \pmod{q_i}, \qquad m \not\equiv -a \pmod{q_i}, \qquad m \not\equiv -(a+b) \pmod{q_i}.$$

If $q \nmid a \cdot b \cdot (a+b)$ then these three residues are all different. In a previous paper[3] I stated the following theorem: Let $q_1, q_2, \cdots$ be primes all less than n^{c_3}. Associate with each q_i t distinct residues $r_1^{(i)}, \cdots, r_t^{(i)}$. Then the number of integers $m \leqq n$ for which

$$m \not\equiv r_j^{(i)} \pmod{q_i}, \qquad j = 1, 2, \cdots, t; \; i = 1, 2, \cdots,$$

is less than

$$cn \prod_i \left(1 - \frac{t}{q_i}\right).$$

The proof of this theorem follows easily from Brun's method.[3] Thus

[3] P. Erdös, Proc. Cambridge Philos. Soc. vol. 33 (1937) p. 8, Lemma 2. A book of Rosser and Harrington on Brun's method will soon appear which will contain a detailed proof of this result.

we have

$$U < c_4 n \log n \prod_q \left(1 - \frac{3}{q}\right), \quad q < n^{c_3}, \qquad q \nmid a \cdot b \cdot (a + b).$$

It is well known that[4]

$$\prod_{q<x} \left(1 - \frac{3}{q}\right) < \frac{c}{(\log x)^3} \quad \text{and} \quad \prod_q \left(1 - \frac{q}{q^2}\right) > 0.$$

Thus

$$U < c_5 \frac{n}{(\log n)^2} \prod_q \left(1 + \frac{3}{q}\right), \qquad q \mid a \cdot b \cdot (a + b).$$

Hence finally from (6) and $V \leqq U + n^{c_3}$,

$$g(n; a, b) < c_5' \frac{n}{(\log n)^2} \prod_q \left(1 + \frac{3}{q}\right), \qquad q \mid a \cdot b \cdot (a + b). \tag{7}$$

Now we split the k's satisfying (5) into two classes. In the first class put the k's with $d_k > 20 \log n$ and in the second class the other k's. From $p_n < 2 \cdot n \cdot \log n$ we deduce that the number of k's of the first class is less than $n/10$.

The number of the k's of the second class is not greater than

$$\sum{}' g(n; a, b) < c_5' \frac{n}{(\log n)^2} \sum{}' \prod_q \left(1 + \frac{3}{q}\right), \qquad q \mid a \cdot b \cdot (a + b), \tag{8}$$

where the prime indicates that the summation is extended over those a and b with $a < 20 \cdot \log n$, $1 + c_1 > b/a > 1 - c_1$. Now

$$\sum{}' \prod_{q \mid a \cdot b(a+b)} \left(1 + \frac{3}{q}\right) \leqq \sum_1 \left(\prod_{q \mid a} \left(1 + \frac{3}{q}\right) \sum_2 \prod_{q \mid b(a+b)} \left(1 + \frac{3}{q}\right)\right)$$

where in $\sum_1$, $a < 20 \log n$ and in $\sum_2$, $1 + c_1 > b/a > 1 - c_1$. We have

$$\begin{aligned} \sum_2 \prod_{q \mid b\,(a+b)} \left(1 + \frac{3}{q}\right) &< \sum_2 \left(\prod_{q \mid b} \left(1 + \frac{3}{q}\right)^2 + \prod_{q \mid a+b} \left(1 + \frac{3}{q}\right)^2 \right) \\ &< \sum_2 \left(\prod_{q \mid b} \left(1 + \frac{15}{q}\right) + \prod_{q \mid a+b} \left(1 + \frac{15}{q}\right) \right) \\ &< \sum_{m<3a} 2 \left(1 + \frac{2c_1 a}{m}\right) \frac{15^{V(m)}}{m} < c_6 c_1 a, \end{aligned}$$

[4] See, for example, Hardy-Wright, p. 349.

by interchanging the order of summation and by observing that the number of b's satisfying $1+c_1>b/a>1-c_1$ and $b\equiv 0 \pmod m$ is less than $1+(2\cdot c_1\cdot a/m)$. The same holds for the b's satisfying $1+c_1>b/a>1-c_1$ and $a+b\equiv 0 \pmod m$. ($v(m)$ denotes the number of prime factors of m.) Thus

$$\begin{aligned}\sum{}' \prod_{q\mid a\cdot b\cdot(a+b)}\left(1+\frac{3}{q}\right) &< c_6c_1\sum_1 a\prod_{q\mid a}\left(1+\frac{3}{q}\right)\\ &< 20c_6c_1\log n\sum_1\prod_{q\mid a}\left(1+\frac{3}{q}\right)\\ &< 20c_6c_1\log n\sum_{m=1}^{\infty}\frac{(20\log n)3^{V(m)}}{m^2}\\ &< c_7c_1(\log n)^2 < \frac{1}{10c_5'}(\log n)^2\end{aligned}$$

if $c_1<1/10\cdot c_7\cdot c_5'$. Hence finally from (8) the number of solutions of (5) is less than

$$n/10+n/10<n/4,$$

which proves Lemma 1.

LEMMA 2. *There exists a constant c_8 so that the number of integers $k\leqq n$ satisfying*

$$d_{k+1}/d_k>t \quad\text{or}\quad d_{k+1}/d_k<1/t \tag{9}$$

is less than $c_8\cdot n/t^{1/2}$.

It suffices to prove the lemma for large t. We split the integers k satisfying (9) into two not necessarily disjoint classes. In the first class are the k's for which either

$$d_k\geqq t^{1/2}\cdot\log n \quad\text{or}\quad d_{k+1}\geqq t^{1/2}\cdot\log n.$$

In the second class are the k's for which either

$$d_k\leqq(\log n)/t^{1/2} \quad\text{or}\quad d_{k+1}\leqq(\log n)/t^{1/2}.$$

Clearly if (9) is satisfied then k is in one of these classes.

We obtain from $p_n<2\cdot n\cdot\log n$ that the number of k's of the first class is less than $4\cdot n/t^{1/2}$.

As in the proof of Lemma 1 we obtain from our result proved in a previous paper[3] that the number Z of solutions of $d_u=a$, $u\leqq n$ is less than

$$Z < c_9 n \log n \prod_q \left(1 - \frac{2}{q}\right), \qquad q \nmid a,\ q < n^{c_3}.$$

Thus as in Lemma 1

$$Z < c_{10} \frac{n}{\log n} \prod_{p \mid a} \left(1 + \frac{2}{p}\right).$$

Thus the number of k's of the second class is less than

$$2c_{10} \frac{n}{\log n} \sum_{a < \log n / t^{1/2}} \prod_{p \mid a} \left(1 + \frac{2}{p}\right) < 2c_{10} \frac{n}{\log n} \sum_{p=1}^{\infty} \frac{(\log n) 2^{V(d)}}{t^{1/2} d^2}$$
$$< \frac{c_{11} n}{t^{1/2}},$$

which proves Lemma 2, with $c_8 = 2 + c_{11}$.

Now we can prove Theorem 1. It will suffice to prove (1). Suppose that (1) is not true. Then for every $c_1 > 0$ and $\epsilon > 0$ there exists an arbitrarily large n so that the number of solutions of

$$d_{k+1} > (1 + c_1) d_k \tag{10}$$

is less than $\epsilon \cdot n$. Consider the product

$$\frac{d_n}{d_1} = \frac{d_2}{d_1} \cdot \frac{d_3}{d_2} \cdots \frac{d_n}{d_{n-1}} \cdot$$

By Lemma 2 the number of $k \leqq n$ satisfying $d_{k+1}/d_k > 2^{2l}$ is less than $c_8 n / 2^l$. Thus by Lemma 1 and (10) we have for every u

$$d_n/d_1 < 2^{2u\epsilon n} \prod_{l \geqq 2^{2u}} (2^{2l})^{c_8 n / 2l} \cdot (1 + c_1)^{n/4} (1 - c_1)^{n/2}$$
$$< 2^{2u\epsilon n} \exp \sum_{l \geqq u} \frac{c_8 l n \log 4}{2^l} \cdot (1 - c_1)^{n/4}.$$

If ϵ is sufficiently small there is a suitable choice of u such that $2^{2u\epsilon n} < (1 + c_1)^{n/8}$ and

$$\exp \sum_{l \geqq u} \frac{c_8 l n \log 4}{2^l} < (1 + c_1)^{n/8}.$$

Thus $d_n/d_1 < (1 - c_1^2)^{n/4} < 1/n$ for arbitrarily large n, an evident contradiction. This proves (1) and completes the proof of Theorem 1.

SYRACUSE UNIVERSITY