ON THE DIFFERENCE OF CONSECUTIVE PRIMES

P. ERDŐS

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

\[
d_{k+1} > (1 + c_1)d_k, \quad k \leq n,
\]

and the number of \( l \)'s satisfying both

\[
d_{l+1} < (1 - c_1)d_l, \quad l \leq n,
\]

are each greater than \( c_2n \).

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

\[
\left( \frac{p_{k+1} + p_k}{2} \right)^{1/t} > p_k, \quad k \leq n; \quad \left( \frac{p_{l+1} + p_l}{2} \right)^{1/t} < p_l, \quad l \leq n,
\]

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

Let \( \epsilon \) be sufficiently small but fixed. It is well known that \( p_n < 2 \cdot n \log n \). Thus the number of \( k \leq n \), 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

\[
p_{k+1} < (1 + \epsilon)p_k, \quad d_k > (1 + c_1)d_{k-1}, \quad k \leq n,
\]

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, that²

Received by the editors October 17, 1947.


² This result was also stated in the above paper.
\[ \limsup_{k \to \infty} \frac{d_{k+1}}{d_k} > 1, \quad \liminf_{k \to \infty} \frac{d_{k+1}}{d_k} < 1. \]

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

\[ 1 + c_1 > \frac{d_{k+1}}{d_k} > 1 - c_1, \quad k \leq n, \]

is less than \( n/4 \).

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

\[ d_{k+1} = a, \quad d_k = b, \quad k \leq n. \]

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

\[ g(n; a, b) \leq V. \]

Now let \( c_1 > 0 \) be sufficiently small and \( q_1, q_2, \ldots \) run through the primes less than \( n^{10} \). Then \( V \) is not greater than \( n^{10} \) plus the number \( U \) of integers \( m < 2 \cdot n \cdot \log n \), which satisfy, for all \( i, \)

\[ m \not\equiv 0 \pmod{q_i}, \quad m \not\equiv -a \pmod{q_i}, \quad m \not\equiv -(a + b) \pmod{q_i}. \]

If \( q | a \cdot b \cdot (a+b) \) then these three residues are all different. In a previous paper\(^3\) I stated the following theorem: Let \( q_1, q_2, \ldots \) be primes all less than \( n^{10} \). Associate with each \( q_i \) \( t \) distinct residues \( r_1^{(i)}, \ldots, r_t^{(i)} \). Then the number of integers \( m \leq n \) for which

\[ m \not\equiv r_j^{(i)} \pmod{q_i}, \quad j = 1, 2, \ldots, t; \ i = 1, 2, \ldots, \]

is less than

\[ cn \prod_{i=1}^{t} \left(1 - \frac{1}{q_i}\right). \]

The proof of this theorem follows easily from Brun's method.\(^8\) Thus

we have

\[ U < c_4 n \log n \prod_{q \leq n} \left(1 - \frac{3}{q}\right), \quad q < n^a, \quad q \mid a \cdot b \cdot (a + b). \]

It is well known that

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

Thus

\[ U < c_6 \frac{n}{(\log n)^2} \prod_{q} \left(1 + \frac{3}{q}\right), \quad q \mid a \cdot b \cdot (a + b). \]

Hence finally from (6) and \( V \leq U + n^a, \)

\[ g(n; a, b) < c_6' \frac{n}{(\log n)^2} \prod_{q} \left(1 + \frac{3}{q}\right), \quad q \mid a \cdot b \cdot (a + b). \]

Now we split the \( k \)'s satisfying (5) into two classes. In the first class put the \( k \)'s with \( d_k > 2 \cdot \log n \) 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

\[ \sum' g(n; a, b) < c_6' \frac{n}{(\log n)^2} \sum' \prod_{q} \left(1 + \frac{3}{q}\right), \quad q \mid a \cdot b \cdot (a + b), \]

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

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

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

\[ \sum_2 \prod_{q \mid b \cdot (a+b)} \left(1 + \frac{3}{q}\right) < \sum_2 \left(\prod_{q \mid b} \left(1 + \frac{3}{q}\right)^2 + \prod_{q \mid a+b} \left(1 + \frac{3}{q}\right)^2\right) \]

\[ < \sum_2 \left(\prod_{q \mid b} \left(1 + \frac{15}{q}\right) + \prod_{q \mid a+b} \left(1 + \frac{15}{q}\right)\right) \]

\[ < \sum_{m < 3a} 2 \left(1 + \frac{2c_1 a}{m}\right) \frac{15^r(m)}{m} < c_6 c_1 a, \]

\[ ^4 \text{See, for example, Hardy-Wright, p. 349.} \]
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 $b \equiv 0 (\bmod m)$ 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 $a + b \equiv 0 (\bmod m)$. ($v(m)$ denotes the number of prime factors of $m$.) Thus

$$\sum' \prod_{q \mid a \cdot b \cdot (a+b)} \left( 1 + \frac{3}{q} \right) < c_6 c_1 \sum_1 a \prod_{q \mid a} \left( 1 + \frac{3}{q} \right)$$

$$< 20 c_6 c_1 \log n \sum_1 \prod_{q \mid a} \left( 1 + \frac{3}{q} \right)$$

$$< 20 c_6 c_1 \log n \sum_{m=1}^{\infty} \frac{(20 \log n)^{3v(m)}}{m^2}$$

$$< c_7 c_1 (\log n)^2 < \frac{1}{10 c_1^2} (\log n)^2$$

if $c_1 < 1/10 \cdot c_7 \cdot c_6$. 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 $k \leq n$ satisfying

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

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

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

$$d_k \geq \ell^{1/2} \cdot \log n \quad \text{or} \quad d_{k+1} \leq \ell^{1/2} \cdot \log n.$$ 

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

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

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 $4 \cdot n/\ell^{1/2}$.

As in the proof of Lemma 1 we obtain from our result proved in a previous paper that the number $Z$ of solutions of $d_u = a$, $u \leq n$ is less than
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
$$\frac{2c_{10}}{\log n} \sum_{a < \log n} \prod_{p \mid a} \left(1 + \frac{2}{p}\right) < \frac{2c_{10}}{\log n} \sum_{p=1}^{\infty} \frac{(\log n)2^p(d)}{d^{1/2}d^2} < \frac{c_{11}n}{d^{1/2}},$$

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

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

is less than \(\epsilon \cdot n\). Consider the product
$$\frac{d_n}{d_1} = \frac{d_2}{d_1} \frac{d_3}{d_2} \ldots \frac{d_n}{d_{n-1}}.$$

By Lemma 2 the number of \(k \leq n\) satisfying \(d_{k+1}/d_k > 2^i\) is less than \(c_{6n}/2^i\). Thus by Lemma 1 and (10) we have for every \(u\)
$$d_n/d_1 < 2^{2u\epsilon n} \prod_{l=2^u} \left(2^{2l}c_{6n}/2^l\right)(1 + c_1)^n/(1 + c_1)^{n/2}$$
$$< 2^{2u\epsilon n} \exp \sum_{l \leq u} \frac{c_{6i}n \log 4}{2^i} \cdot (1 - c_1)^{n/2}.$$

If \(\epsilon\) is sufficiently small there is a suitable choice of \(u\) such that
$$2^{2u\epsilon n} < (1 + c_1)^{n/8}$$
and
$$\exp \sum_{l \leq u} \frac{c_{6i}n \log 4}{2^i} < (1 + c_1)^{n/8}.$$

Thus \(d_n/d_1 < (1 - c_1^{n/8}) < 1/n\) for arbitrarily large \(n\), an evident contradiction. This proves (1) and completes the proof of Theorem 1.