

Problems and results on the differences of consecutive primes.

By P. ERDŐS in Syracuse (U. S. A.).

Let $p_1 < p_2 < \dots$ be the sequence of consecutive primes. Put $d_n = p_{n+1} - p_n$. The sequence d_n behaves extremely irregularly. It is well known that $\overline{\lim} d_n = \infty$ (since the numbers $n! + 2, n! + 3, \dots, n! + n$ are all composite). It has been conjectured that $d_n = 2$ for infinitely many n (i. e. there are infinitely many prime twins). This conjecture seems extremely difficult. In fact not even $\underline{\lim} d_n < \infty$, or even $\underline{\lim} \frac{d_n}{\log n} = 0$ has ever been proved. A few years ago I proved¹⁾ by using Bruns's method that

$$(1) \quad \underline{\lim} \frac{d_n}{\log n} < 1.$$

$\underline{\lim} \frac{d_n}{\log n} \leq 1$ is an immediate consequence of the prime number theorem. WESTZYNTHIUS²⁾ proved in the other direction that

$$(2) \quad \overline{\lim} \frac{d_n}{\log n} = \infty.$$

In fact he show that for infinitely many n ,

$$d_n > \log n \cdot \log \log \log n / \log \log \log \log n.$$

I proved³⁾ using Brun's method that for infinitely many n

$$(3) \quad d_n > c \frac{\log n \cdot \log \log n}{(\log \log \log n)^2}.$$

CHEN⁴⁾ proved (3) very much simpler without using Brun's method,

¹⁾ *Duke Math. Journal*, Vol. 6 (1940), p. 438–441.

²⁾ *Comm. Phys. Math. Soc. Sci. Fenn.*, Helsingfors, Vol. 5 (1931), No. 25. p. 1–37.

³⁾ *Quarterly Journal of Math.*, Vol. 6 (1935), p. 124–128. In this paper one can find some more literature on the difference of consecutive primes.

⁴⁾ *Schriften des Math. Seminars und des Instituts für angewandte Math. der Univ. Berlin*, 4 (1938), p. 35–55.

and RANKIN⁵⁾ proved that

$$(4) \quad d_n > c \frac{\log n \cdot \log \log n \cdot \log \log \log n}{(\log \log \log n)^2}$$

In the present note I prove the following

Theorem:

$$(5) \quad \overline{\lim} \frac{\min(d_n, d_{n+1})}{\log n} = \infty.$$

In other words to every c there exist values of n satisfying the inequalities $d_n > c \log n$, $d_{n+1} > c \log n$.

It can be conjectured that $\overline{\lim} \left(\frac{\min(d_n, d_{n+1}, \dots, d_{n+k})}{\log n} \right) = \infty$ for every k , but I cannot prove this for $k > 1$.

It can also be conjectured that $\underline{\lim} \frac{\max(d_n, d_{n+1})}{\log n} < 1$, but I cannot prove this either.

Proof of the Theorem⁶⁾. Let n be a large integer, $m = \varepsilon \cdot \log n$, where ε is a small but fixed number, $f(m)$ tends to infinity together with m and $f(m) = o(\log m)^{1/3}$, $N = \prod_{p_i \leq m} p_i$, q_i denotes the primes $\leq (\log m)^2$, r_i the primes of the interval $[(\log m)^2, m^{1/100} \log \log m]$, s_i the primes of the interval $(m^{1/100} \log \log m, \frac{m}{2})$, and t_i the primes satisfying $\frac{m}{2} \leq t_i \leq m$.

Our aim will be to determine a residue class $x \pmod{N}$ so that

$$(6) \quad (x+1, N) = 1 \text{ and } (x+k, N) \neq 1 \text{ for all } |k| \leq m f(m) \text{ and } k \neq \pm 1.$$

Suppose we already determined an x satisfying (6). Then we complete the proof as follows: Consider the arithmetic progression $(x+1) + dN$, $d=1, \dots$. Since $(x+1, N) = 1$ it represents infinitely many primes, in fact by a theorem of LINNIK⁷⁾ the least prime it represents does not exceed N^{c_1} where c_1 is an absolute constant independent of N . Now by the prime number theorem, or by the more elementary results of TCHEBICHEFF, we have

$$N^{c_1} = \left(\prod_{p_i \leq m} p_i \right)^{c_1} < e^{2m c_1} = n^{2\varepsilon c_1} < n^{1/2}$$

for $\varepsilon < \frac{1}{4c_1}$, or there exists a prime p_j satisfying

$$(7) \quad p_j < n^{1/2}, \quad p_j = (x+1) + dN.$$

⁵⁾ *Journal of the London Math. Soc.*, Vol. 13 (1938), p. 242—247. For further results on the difference of consecutive primes see P. ERDŐS and P. TURÁN, *Bull. Amer. Math. Soc.*, Vol. 54 (1948).

⁶⁾ We use the method of CHEN.

⁷⁾ On the least prime in an arithmetical progression, I. The basic theorem, *Math. Sbornik*, Vol. 15 (57), No 2, p. 139—178. II. The Deuring—Heilbronn phenomenon, *Math. Sbornik*, Vol. 15 (57), p. 347—368.

It follows from (6) that

$$(8) \quad p_{j+1} - p_j \geq mf(m), \quad p_j - p_{j-1} \geq mf(m).$$

Thus from (7) and (8)

$$(9) \quad \frac{p_{j+1} - p_j}{\log p_j} \geq \frac{mf(m)}{\log n} = \varepsilon f(m) \rightarrow \infty, \quad \frac{p_j - p_{j-1}}{\log p_j} \geq \frac{mf(m)}{\log n} = \varepsilon f(m) \rightarrow \infty,$$

which proves (5) and our Theorem is proved.

Now we only have to find an x satisfying (6). Put

$$(10) \quad x \equiv 0 \pmod{q_i}, \quad x \equiv 0 \pmod{s_i}.$$

Let $|k| \leq mf(m)$, have no factor among the q 's and s 's. Then we assert that k is either ± 1 or a prime $> \frac{m}{2}$ or has all its prime factors among the r 's. For if not then k would be greater than the product of the least r and the least t , i. e.

$$k \geq \frac{m}{2} (\log m)^2 > mf(m); \quad (f(m) = o(\log m))$$

an evident contradiction.

Denote by $u_1, u_2, \dots, u_\xi$ the integers $\leq |mf(m)|$ all whose prime factors are r 's. We estimate ξ as follows: We split the u 's into two classes. In the first class are the u 's which have less than $10 \cdot \log \log m$ different prime factors. The number of these u 's is clearly less than

$$(11) \quad (m^{1/100 \log \log m} \cdot \log m)^{10 \log \log m} < m^{2/3}$$

(since the number of integers of the form p^α , $p^\alpha < mf(m)$, $p < m^{1/10 \log \log m}$ is less than $m^{1/100 \log \log m} \cdot \log m$).

For the u 's of the second class $v(u) \geq 10 \cdot \log \log m$ ($v(u)$ denotes the number of different prime factors of u). Thus from

$$\sum 2^{v(u)} < 2 \sum_{b=1}^{mf(m)} 2^{v(b)} < cmf(m) \cdot \log m < m (\log m)^2$$

we obtain that the number of the u 's of the second class is less than

$$(12) \quad \frac{m (\log m)^2}{2^{10 \log \log m}} < \frac{m}{(\log m)^2}.$$

Hence finally from (11) and (12)

$$(13) \quad \xi = o\left(\frac{m}{\log m}\right).$$

Denote now by $v_1, v_2, \dots, v_\eta$ the integers of absolute value $\leq mf(m)$ which do not satisfy the congruence (10). Then the v 's are either -1 or are u 's, or of the form $\pm p$, $\frac{m}{2} < p \leq mf(m)$. Thus by (13) and the results

of TCHEBICHEFF about primes

$$(14) \quad \eta < c \frac{mf(m)}{\log m}.$$

Suppose we already determined for $i < j$ a residue class $\lambda^{(i)} \pmod{r_i}$ so that

$$(15) \quad x \equiv \lambda^{(i)} \pmod{r_i}, \quad \lambda^{(i)} \not\equiv -1, \quad i = 1, 2, \dots, (j-1).$$

Denote by $v_1^{(j)}, \dots, v_{\eta_j}^{(j)}$ the v 's which do not satisfy any of the congruences (15). There clearly exists a residue class $\pmod{r_j}$ which contains at least η_j/r_j of the v 's. Denote this residue class by $\lambda_1^{(j)}$. If $\lambda_1^{(j)} \not\equiv -1 \pmod{r_j}$ we put

$$(16) \quad x \equiv \lambda_1^{(j)} \pmod{r_j}.$$

If on the other hand $\lambda_1^{(j)} \equiv -1 \pmod{r_j}$ we distinguish two cases: In the first case the residue class $\lambda_1^{(j)} \pmod{r_j}$ contains less than $\frac{1}{2} \eta_j$ of the $v^{(j)}$'s.

Then there clearly exists a residue class $\lambda_2^{(j)} \not\equiv \lambda_1^{(j)} \pmod{r_j}$ which contains more than $\eta_j/2r_j$ of the $v^{(j)}$'s. Put for these r_j 's

$$(17) \quad x \equiv \lambda_2^{(j)} \pmod{r_j}.$$

We continue this operation for all the r 's and let us first assume that for every r_j either $\lambda_1^{(j)} \not\equiv -1 \pmod{r_j}$ or that the first case occurs. Denote by $V_1, V_2, \dots, V_\varrho$ the v 's which do not satisfy the congruences (16) and (17). Clearly

$$(18) \quad \varrho \leq \eta \prod \left(1 - \frac{1}{2r_j}\right) < c \frac{mf(m) \log \log m}{(\log m)^{3/2}} = o\left(\frac{m}{\log m}\right)$$

since

$$\frac{c_1}{\sqrt{\log z}} < \prod_{p \leq z} \left(1 - \frac{1}{2p}\right) < \frac{c_2}{\sqrt{\log z}}.$$

Put now

$$(19) \quad x \equiv -V_i \pmod{t_i}, \quad 1 \leq i \leq \varrho,$$

where t_i is chosen so that $V_i - 1 \not\equiv 0 \pmod{t_i}$ and the different V_i correspond different t_i . This is always possible since the number of prime factors of $V_i - 1$ is less than $c \log m$ and number of t 's equals $\pi(2m) - \pi(m)$, and we have by (18) and the results of TCHEBICHEFF

$$\pi(2m) - \pi(m) > c_1 \frac{m}{\log m} > \varrho + c \log m.$$

For the t 's not used in (19) we put

$$(20) \quad x \equiv 0 \pmod{t_i}.$$

The congruences (10), (16), (17) and (10) determine $x \pmod{N}$ so that (6) is clearly satisfied, which proves our Theorem in case the second case never occurs.

Assume now that the second case occurs for some r 's. Let r_j be the r of smallest index for which the second case occurs. Consider the congruences (16) and (17) for $i < j$, and denote again by $v_1^{(j)}, \dots, v_{\eta_j}^{(j)}$ the v 's which do not satisfy (16) and (17) for $i < j$. By our assumption the arithmetic progression $-1 + dr_j$ contains at least $\frac{1}{2}\eta_j$ of the $v^{(j)}$'s and since all $v^{(j)}$'s are less than $mf(m)$ in absolute value, we obtain

$$\eta_j < \frac{4mf(m)}{r_j} = o\left(\frac{m}{\log m}\right)$$

and the proof is completed as in the previous case, thus our Theorem is completely proved.

Added in proof (1. Aug. 1949). By a slight modification of the last step of the proof we can show the existence of infinitely many n so that

$$\min(d_n, d_{n+1}) > c \frac{\log n \log \log n \log \log \log n}{(\log \log \log n)^2}.$$

(Received January 31, 1949.)