Reprinted from Duke Mathematical Journal
Vol. 6, No. 2, June, 1940

THE DIFFERENCE OF CONSECUTIVE PRIMES

By P. Erdös

Let p_n denote the n-th prime. Backlund [1][1] proved that, for every positive ϵ and infinitely many n, $p_{n+1} - p_n > (2 - \epsilon) \log p_n$. Brauer and Zeitz [2, 10] proved that $2 - \epsilon$ can be replaced by $4 - \epsilon$. Westzynthius [9] proved that for an infinity of n

$$p_{n+1} - p_n > \frac{2 \log p_n \log \log \log p_n}{\log \log \log \log p_n},$$

and this was improved by Ricci [7] to

$$p_{n+1} - p_n > c_1 \log p_n \log \log \log p_n ,$$

where, as throughout the paper, the c's denote positive absolute constants. I [4] showed that

$$p_{n+1} - p_n > c_2 \frac{\log p_n \log \log p_n}{(\log \log \log p_n)^2},$$

and lately Rankin [6] proved

$$p_{n+1} - p_n > c_3 \frac{\log p_n \log \log p_n \log \log \log \log p_n}{(\log \log \log p_n)^2}.$$

In the other direction the best known result is that of Ingham [5] which states that for sufficiently large n

$$p_{n+1} - p_n < p_n^{\frac{38}{61}+\epsilon} < p_n^{\frac{5}{8}}.$$

Thus it is known that

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} = \infty.$$

Very much less is known about

$$A = \liminf \frac{p_{n+1} - p_n}{\log p_n}.$$

Hardy and Littlewood proved a few years ago, by using the Riemann hypothesis, that $A \leqq \frac{2}{3}$, and Rankin recently proved, again by using the Riemann hypothesis, that $A \leqq \frac{3}{5}$. In the present paper we are going to prove—without the Riemann hypothesis—that

$$A < 1 - c_4, \qquad \text{for a certain } c_4 > 0.$$

Received December 12, 1939.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

It seems extremely likely that $A = 0$. In fact, a well-known conjecture states that the equation $p_{n+1} - p_n = 2$ has infinitely many solutions (i.e., there are infinitely many prime twins).

We need two lemmas.

LEMMA 1. *The number of solutions of*

$$a = p_i - p_j, \qquad p_j, p_i \leqq n,$$

does not exceed

$$c_5 \prod_{p \mid a} \left(1 + \frac{1}{p}\right) \frac{n}{(\log n)^2}.$$

The proof is well known ([8], p. 670).

LEMMA 2. *Let c_4 be sufficiently small; then*

$$\sum' \prod_{p \mid a} \left(1 + \frac{1}{p}\right) < \frac{1}{6c_5} \log n,$$

where the prime indicates that the summation is extended over the a's of the interval

$$(1 - c_4) \log n \leqq a \leqq (1 + c_4) \log n.$$

Proof. We have

$$\sum' \prod_{p \mid a} \left(1 + \frac{1}{p}\right) \leqq \sum_{d < (1+c_4) \log n} \frac{1}{d} \left(\frac{2c_4 \log n}{d} + 1\right)$$

$$< c_6 \log n + \sum_{d < (1+c_4) \log n} \frac{1}{d} < \frac{1}{6c_5} \log n$$

for sufficiently small c_4, and the proof is complete.

Now we can prove our theorem. Denote by $p_1, p_2, \cdots, p_x$ the primes of the interval $\frac{1}{2}n, n$. It follows from the prime number theorem that, for suffiiently large n, $x > (\frac{1}{2} - \epsilon)n/\log n$. It suffices to prove that if n is sufficiently large, then for at least one i

$$p_{i+1} - p_i < (1 - c_4) \log n \qquad (i \leqq x - 1).$$

For then we have

$$\liminf_{r \to \infty} \frac{p_{r+1} - p_r}{\log p_r} \leqq \frac{(1 - c_4) \log n}{\log \frac{1}{2}n} \to 1 - c_4.$$

Write

$$b_1 = p_2 - p_1, b_2 = p_3 - p_2, \cdots, b_{x-1} = p_x - p_{x-1}.$$

Evidently

$$\sum_{i=1}^{x-1} b_i \leqq \tfrac{1}{2}n.$$

From Lemmas 1 and 2 it follows that the number of b's lying in the interval

$$(1 - c_4) \log n \leqq b \leqq (1 + c_4) \log n$$

does not exceed

$$c_5 \frac{n}{(\log n)^2} \sum{}' \prod_{p | a} \left(1 + \frac{1}{p}\right) < \frac{n}{6 \log n}.$$

Hence if $b_i < (1 - c_4) \log n$ had no solution, we should obtain

$$\sum_{i=1}^{x-1} b_i > \frac{n}{6 \log n} (1 - c_4) \log n + (\tfrac{1}{3} - \epsilon) \frac{n}{\log n} (1 + c_4) \log n$$
$$= \tfrac{1}{2}n(1 - 2\epsilon) + (\tfrac{1}{6} - \epsilon)c_4 n > \tfrac{1}{2}n.$$

This is an evident contradiction and the theorem is proved.

Denote by $q_1 < q_2 < \cdots < q_y$ the primes not exceeding n. Cramér [3] proved by aid of the Riemann hypothesis that

$$\sum_{i=1}^{y-1} (q_{i+1} - q_i) = O\left(\frac{n}{\log \log n}\right) \quad (q_{i+1} - q_i > (\log q_i)^3).$$

It might be conjectured that the following stronger result also holds:

$$\sum_{i=1}^{y-1} (q_{i+1} - q_i)^2 = O(n \log n).$$

This result if true must be very deep. I could not even prove the following very much more elementary conjecture: Let n be any integer and let $0 < a_1 < a_2 < \cdots < a_x < n$ be the $\varphi(n)$ integers relatively prime to n; then

$$\sum_{i=1}^{x-1} (a_{i+1} - a_i)^2 < c_6 \frac{n^2}{\varphi(n)}.$$

BIBLIOGRAPHY

1. R. J. Backlund, *Über die Differenzen zwischen den Zahlen die zu den n ersten Primzahlen teilerfremd sind*, commemoration volume in honor of E. L. Lindelöf, Helsingfors, 1929.
2. A. Brauer and H. Zeitz, *Über eine zahlentheoretische Behauptung von Legendre*, Sitz. Berliner Math. Ges., vol. 29(1930), pp. 116–125.
3. H. Cramér, *On the difference between consecutive primes*, Acta Arithmetica, vol. 2(1937), pp. 23–45.
4. P. Erdös, *On the difference of consecutive primes*, Quarterly Journal of Math., Oxford Series, vol. 6(1935), pp. 124–128.
5. A. E. Ingham, *On the difference between consecutive primes*, Quarterly Journal of Math., Oxford Series, vol. 8(1937), pp. 255–266.
6. R. A. Rankin, *The difference between consecutive primes*, Journal London Math. Soc., vol. 13(1938), pp. 242–247.

7. G. Ricci, *Ricerche aritmetiche sui polinomi*, II: *Intorno a una proposizione non vera di Legendre*, Rend. Circ. Mat. di Palermo, vol. 58(1934).
8. L. Schnirelmann, *Über additive Eigenschaften von Zahlen*, Math. Annalen, vol. 107(1933), pp. 649–690.
9. E. Westzynthius, *Über die Verteilung der Zahlen die zu den n ersten Primzahlen teilerfremd sind*, Comm. Phys. Math. Soc. Sci. Fenn., Helsingfors, vol. 5(1931), no. 25, pp. 1–37.
10. H. Zeitz, *Elementare Betrachtung über eine zahlentheoretische Behauptung von Legendre*, Berlin, 1930.

Institute for Advanced Study.