ON SOME NEW QUESTIONS ON THE DISTRIBUTION OF PRIME NUMBERS

P. ERDÖS AND P. TURÁN

1. **Introduction.** In connection with some recent unpublished investigations concerning the Riemann hypothesis one of us raised the question whether $\log p_n$ is convex for sufficiently large n, or at least whether it has few points of inflexion. (Throughout this paper $p_1=2$, $p_2=3, \cdots, p_n, \cdots$ denotes the sequence of primes.) In other words: Is it true that the inequalities

$$p_{n-1}\cdot p_{n+1} > p_n^2, \qquad p_{m+1}p_{m-1} < p_m^2 \tag{1}$$

both have infinitely many solutions? We shall show that the answer is affirmative.

A still simpler question is whether the sequence of primes itself is convex or concave from a certain n on. We shall prove that this is not so, that is, the equations

$$\frac{p_{n-1}+p_{n+1}}{2} > p_n, \qquad \frac{p_{m-1}+p_{m+1}}{2} < p_m \tag{2}$$

have infinitely many solutions.[1]

If the well known hypothesis about prime twins is true, that is, if the equation $p_{n+1}-p_n=2$ has infinitely many solutions, (1) and (2) of course are trivially satisfied.

The first inequality of (2) is inserted only for the sake of completeness. It follows from the well known fact that $\limsup (p_{n+1}-p_n)=\infty$ (since $n!+2, n!+3, \cdots, n!+n$ are all composite). The proof of the other inequalities will be simple, but less trivial.

Clearly $p_{n-1}p_{n+1}>p_n^2$ implies $(p_{n-1}+p_{n+1})/2>p_n$ and $p_m>(p_{m-1}+p_{m+1})/2$ implies $p_m^2>p_{m-1}p_{m+1}$. The well known relations between the various mean values suggest the following questions: Is it true that for every t the inequalities

$$\left(\frac{p_{n-1}^t+p_{n+1}^t}{2}\right)^{1/t} > p_n \tag{3}$$

and

Received by the editors May 12, 1947, and, in revised form, July 3, 1947.

[1] Professor G. Pólya and Mr. P. Ungár communicated to us subsequently a proof very similar to our own.

$$\left(\frac{p_{m-1}^t + p_{m+1}^t}{2}\right)^{1/t} < p_m \tag{4}$$

have infinitely many solutions? By the well known relations between the means it follows that (1) and (2) is a consequence of (3) and (4), and that it suffices to prove (3) for $t<0$ and (4) for $t>0$. (The inequality about means states that $((a^t+b^t)/2)^{1/t}$ is an increasing function of t.)[2]

An elementary proof of (3) and (4) is given in §2. The only result we use about primes is that

$$\pi(x) > c_1\, x/\log x. \tag{5}$$

This can be found in the first pages of Ingham's book *The distribution of prime numbers*. ($\pi(x)$ denotes the number of primes not exceeding x.)

All these questions can be investigated by a method which is less elementary than that given in §2, but which perhaps can be used to attack some of the unsolved problems which can be raised here. Only (2) is treated by this method (in §3).

In §4 we state without proof some results about the number of solutions of (3) and (4). Finally we state some unsolved problems, which are natural generalizations of our theorems.

2. Elementary proofs.

THEOREM 1. *The inequalities* (3) *and* (4) *have infinitely many solutions.*

We need the following lemma.

LEMMA. *Let $A>0$ be any constant. Then the inequalities*

$$p_k - p_{k-1} < p_{k+1} - p_k, \qquad p_k - p_{k-1} < A p_k^{1/2}, \tag{6}$$

$$p_{k+1} - p_k < p_k - p_{k-1}, \qquad p_{k+1} - p_k < A p_k^{1/2} \tag{7}$$

have infinitely many solutions.

The proof of (6) is quite trivial. It follows from (5) that for infinitely many m and a suitable c_2, $p_{m+1}-p_m<c_2 \log p_m$. Determine the least $k>m$ for which $p_{k+1}-p_k>p_{m+1}-p_m$. Then clearly p_{k-1}, p_k, p_{k+1} satisfy (6).

Now we prove (7). Assume that (7) has only a finite number of solutions (that is, there are no solutions for $p>p_0$). Let m be large

[2] See, for example, Hardy-Littlewood-Pólya, *Inequalities*, p. 26.

and $p_{m+1}-p_m<c_2 \log p_m$. Let p_r be the smallest prime greater than $p_m^{1/2}$. Then clearly

$$(8) \qquad p_{r+1}-p_r \leqq p_{r+2}-p_{r+1}<\cdots \leqq p_{m+1}-p_m<c_2 \log p_m.$$

For if not let k $(r<k \leqq m)$ be the greatest index for which $p_{k+1}-p_k<p_k-p_{k-1}$. Then clearly p_{k-1}, p_k, p_{k+1} satisfy (7) (since $p_{k+1}-p_k \leqq p_{m+1}-p_m<c_2 \log p_m<c_2(p_m)^{1/4}<c_2(p_k)^{1/2}$). This proves (8). But if $p_{t+1}-p_t=p_{t+2}-p_{t+1}=\cdots=p_{t+s+1}-p_{t+s}=d$ we evidently have $s \leqq d$ (since the integers x, $x+d$, $\cdots$, $x+xd=x(d+1)$ can not all be primes). Hence we obtain from (8) that

$$m-r \leqq 1+2+\cdots+[c_2 \log p_m]<(c_2 \log p_m)^2$$

or

$$m=\pi(p_m) \leqq r+(c_2 \log p_m)^2 \leqq p_m^{1/2}+(c_2 \log p_m)^2$$

which contradicts (5), and completes the proof of the lemma.

Now we can prove Theorem 1. Since, for $a>0, b>0$, $((a^t+b^t)/2)^{1/t}$ is an increasing function of t, it suffices to prove (3) if t is a negative integer not greater than -2, say $t=-l$. Let p_{k-1}, p_k, p_{k+1} satisfy (6) with $A<1/2l^2$. Then we show that they also satisfy (3). Put $p_k-p_{k-1}=u$; since $((a^t+b^t)/2)^{1/t}$ is an increasing function of a and b it will clearly be sufficient to show that (3) is satisfied in case $p_{k+1}-p_k=w+1$. Thus we have to show that $(t=-l \leqq -2)$

$$\left(\frac{(p_k-u)^{-l}+(p_k+u+1)^{-l}}{2}\right)^{-1/l}>p_k$$

or

$$(p_k-u)^{-l}+(p_k+u+1)^{-l}<2p_k^{-l}$$

or

$$(p_k+u+1)^l(2(p_k-u)^l-p_k^l)>p_k^l(p_k-u)^l.$$

Now clearly for $u<p_k^{1/2}/2l^2$

$$(9) \qquad \begin{aligned} p_k^l-ulp_k^{l-1}<(p_k-u)^l &<p_k^l-ulp_k^{l-1}+\binom{l}{2}u^2p_k^{l-2}+\cdots \\ &<p_k^l-(ul-1/2)p_k^{l-1}. \end{aligned}$$

Thus it suffices to show that

$$(p_k^l+(u+1)lp_k^{l-1})(p_k^l-2ulp_k^{l-1})>p_k^l(p_k^l-(ul-1/2)p_k^{l-1}),$$

or

$$(l - 1/2)p_k^{l-1} > 2l^2u(u+1)p_k^{l-2},$$

which is clearly satisfied for $u < p_k^{1/2}/2l^2$, which proves (3).

Now we prove (4). Assume that p_{k-1}, p_k, p_{k+1} satisfy (7) with $A < 1/2t^2$. Put $p_{k+1} - p_k = u$. As before it suffices to consider the case $p_k - p_{k-1} = u+1$ and $t \geqq 2$. Then we have to show that

$$(p_k - (u+1))^t + (p_k + u)^t - 2p_k^t < 0.$$

We have as in (8) for $u < (1/2t^2)p_k^{1/2}$ $(t \geqq 2)$

$$(p_k + u)^t < p_k^t + (tu + 1/2)p_k^{t-1}. \tag{10}$$

Thus from (8) and (9)

$$(p_k - (u+1))^t + (p_k + u)^t - 2p_k^t < 2p_k^t - ((u+1)t - 1/2)p_k^{t-1} + (ut + 1/2)p_k^{t-1} - 2p_k^t < 0,$$

which proves (4) and completes the proof of Theorem 1.

THEOREM 2. *Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers which do not form an arithmetic progression from a certain point on. Let $t < 1$ and $a_k < k^2/4(1-t) - ck$, for every c if k is sufficiently large. Then*

$$((a_{k-1}^t + a_{k+1}^t)/2)^{1/t} > a_k \tag{11}$$

have infinitely many solutions.

THEOREM 3. *Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers which do not form a convex sequence from a certain point on (that is, $a_k - a_{k-1} > a_{k+1} - a_k$ has infinitely many solutions). Let $t > 1$ and $a_k < k^2/4(1-t) - ck$ for every c if k is sufficiently large. Then*

$$((a_{k-1}^t + a_{k+1}^t)/2)^{1/t} < a_k \tag{12}$$

has infinitely many solutions.

The inequalities in both theorems are best possible in the following sense: For every c there exists a sequence $a_1 < a_2 < \cdots$ of integers with $a_k < k^2/4(1-t) - ck$ for all k, the a's not forming an arithmetic progression from a certain point on, and so that (11) has only a finite number of solutions. The same holds for (12).

REMARK. It follows from (5) and our lemma (in §2) that Theorem 1 is a consequence of Theorems 2 and 3.

We prove Theorem 2 only in the special case $t = 0$; the proof of the general case and that of Theorem 3 is similar but requires slightly

longer calculations. It is well known that for $t=0$ the left side of (11) becomes $(a_{k-1}a_{k+1})^{1/2}$. Thus we have to prove that if $a_k<k^2/4-ck$ for every c if k is sufficiently large, then

$$a_{k-1}a_{k+1} > a_k^2 \tag{13}$$

has finitely many solutions.

Suppose this is not true. Then for $k_0<k$, $a_{k-1}\cdot a_{k+1}\leqq a_k^2$. Since the a's do not form an arithmetic progression from a certain point on, it is clear that the equation $a_{k+2}-a_{k+1}>a_{k+1}-a_k$ has infinitely many solutions. Put $a_{k+1}-a_k=x$; we have $a_{k+2}\geqq a_k+2x+1$. Thus since $a_{k+1}^2\geqq a_k\cdot a_{k+2}$ we have

$$(a_k+x)^2 \geqq a_k(a_k+2x+1), \quad \text{or} \quad x^2 \geqq a_k. \tag{14}$$

Assume now that, for some $k>k_0$, $(a_{k+1}-a_k)^2<a_k$. Determine the least $l>k$ for which $a_{l+1}-a_l>a_l-a_{l-1}$. Then we have from (14)

$$(a_{k+1}-a_k)^2 \geqq (a_l-a_{l-1})^2 \geqq a_{l-1} \geqq a_k$$

an evident contradiction. But this means that, for $k>k_0$, $(a_{k+1}-a_k)^2 \geqq a_k$. Thus we clearly obtain that for large enough n the number of a's in the interval $(n^2, (n+1)^2)$ (where n^2 is counted in the interval but $(n+1)^2$ not) does not exceed 2. Thus we evidently have $a_k>k^2/4-ck$ for sufficiently large c, an evident contradiction. This completes the proof. The sequence n^2, $n(n+1)$ with an arbitrary finite set added to it shows that the result is best possible.

3. **Analytical proof.** Now we give an alternative proof of (2) which uses deeper tools. We use the prime number theorem for arithmetic progressions in the form given by A. Page[3]

$$\left|\pi(x, k, l) - \frac{1}{\phi(k)}\int_2^x \frac{dy}{\log y}\right| < \alpha_1 x(\exp(-\alpha_2(\log x)^{1/2}) + \frac{x}{\phi(k)}\exp\left(-\alpha_2\frac{\log x}{k^{1/2}(\log k)^2}\right) \tag{15}$$

where $\pi(x, k, l)$ denotes the number of primes not exceeding x which are congruent to l (mod k) (we assume $(k, l)=1$) and $\alpha_1, \cdots$ are independent of x, k and l. We also need the following result due to Kusmin:[4] Let $\beta_1, \beta_2, \cdots, \beta_n$ be real and $\nu\leqq\beta_2-\beta_1\leqq\cdots\leqq\beta_n-\beta_{n-1}\leqq 1-\nu$ $(0\leqq\nu\leqq 1/2)$. Then

[3] A. Page, Proc. London Math. Soc. (2) vol. 39 (1935) pp. 116–141.

[4] R. O. Kusmin, Zhurnal Leningradskoe Fiziko-Matematicheskoe obshchestvo vol. 1 (1927) pp. 233–239.

$$\left|\sum_{\nu=1}^{n} e^{2\pi i\beta_\nu}\right| < \frac{2}{\pi\nu}. \tag{16}$$

We are going to show that for sufficiently large x there exist primes

$$x \leqq p_{k-1} < p_k < p_{k+1} \leqq 4x \text{ so that } p_{k+1} - p_k < p_k - p_{k-1}. \tag{17}$$

Suppose (17) is not satisfied. Let x be sufficiently large, and consider the primes

$$\begin{aligned} p_j < x \leqq p_{j+1} < \cdots < p_{j+H} \leqq 2x < p_{j+H+1} < \cdots \\ < p_{j+H+E} \leqq 4x. \end{aligned} \tag{18}$$

Since we assume that (17) is false, we have

$$p_{j+2} - p_{j+1} \leqq p_{j+3} - p_{j+2} < \cdots < p_{j+H+E} - p_{j+H+E-1}. \tag{19}$$

We evidently have

$$p_{j+r+1} - p_{j+r} < (3/2)\log x \qquad \text{for } r \leqq H. \tag{20}$$

For if not, then, by (19), $\pi(4x) - \pi(2x) = E < (4x/3\log x) + 1$ which contradicts the prime number theorem.

Put

$$S(\gamma) = \sum_{\nu=j+1}^{j+H} e^{2\pi i\gamma p_\nu}$$

where $(40\log x)^{-1} < \gamma < (12\log x)^{-1}$. We have by (20)

$$\frac{1}{20\log x} < \gamma(p_{\nu+1} - p_\nu) < \frac{1}{2}.$$

Thus from (16)

$$|S(\gamma)| < 20\log x. \tag{21}$$

Let q be any prime satisfying $12\log x < q < 40\log x$ (such a q exists for sufficiently large x). We evidently have

$$\left|S\left(\frac{1}{q}\right)\right| = \sum_{l=1}^{q-1} e^{2\pi i l/q} \sum\nolimits' 1 \tag{22}$$

where the prime indicates that the summation is extended over the $p_\nu \equiv l \pmod q$ with $j+1 \leqq \nu \leqq j+H$. We have by (15)

$$\left|\sum\nolimits' 1 - \frac{1}{q-1}\int_{p_{j+1}}^{p_{j+H}} \frac{dy}{\log y}\right| < \alpha_3 x \exp\left(-\alpha_4 \frac{(\log x)^{1/2}}{(\log\log x)^2}\right). \tag{23}$$

Thus from (23) and (22)

$$\left| S\left(\frac{1}{q}\right) - \frac{1}{q-1}\int_{p_{j+1}}^{p_{j+H}} \frac{dy}{\log y} \right| < \alpha_3 x q \exp\left(-\alpha_4 \frac{(\log x)^{1/2}}{(\log\log x)^2}\right).$$

We have from the prime number theorem $p_{j+1}<1.01x$, $p_{j+H}>1.99x$. Thus

$$\left| S\left(\frac{1}{q}\right) \right| > \frac{1}{40 \log x}\int_{1.01x}^{1.99x} \frac{dy}{\log y} - \alpha_5 x \log x \exp\left(-\alpha_4 \frac{(\log x)^{1/2}}{(\log\log x)^2}\right) > \frac{1}{80}\frac{x}{(\log x)^2}, \tag{24}$$

which contradicts (20) and proves (16). Hence (2) follows immediately since, as remarked in the introduction, $p_{k+1}-p_k>p_k-p_{k-1}$ has (trivially) infinitely many solutions.

4. **Problems and conjectures.** In connection with the Riemann hypothesis the question arose how often the expression

$$p_{k-1}p_{k+1} - p_k^2$$

changes its sign. We can show by using Brun's method that for $k \leqq n$

$$\left(\frac{p_{k-1}^t + p_{k+1}^t}{2}\right)^{1/t} - p_k \tag{25}$$

changes its sign cn times (as remarked before if $t=0$, (25) becomes $p_{k-1}p_{k+1}-p_k^2$).

The inequalities (2) can be stated as follows: The inequalities

$$\frac{p_{n+1}-p_n}{p_n-p_{n-1}} > 1 \quad \text{and} \quad \frac{p_{n+1}-p_n}{p_n-p_{n-1}} < 1$$

have infinitely many solutions. By Brun's method we can show that

$$\limsup \frac{p_{n+1}-p_n}{p_n-p_{n-1}} > 1 \quad \text{and} \quad \liminf \frac{p_{n+1}-p_n}{p_n-p_{n-1}} < 1.$$

It is very probable that the lim sup is infinite and the lim inf is 0.

(2) can be generalized as follows: Let $\sum_{k=1}^{n} a_k x_k$ be any linear form. What is the necessary and sufficient condition that both inequalities

$$\sum_{k=1}^{n} a_k x_k > 0 \quad \text{and} \quad \sum_{k=1}^{n} a_k x_k < 0 \tag{26}$$

have infinitely many solutions in consecutive primes $p_{u+1}, \cdots, p_{u+n}$? From the prime number theorem we obtain the necessary

condition $\sum_{k=1}^{n} a_k = 0$. But $x_2 - x_1$ shows that this condition is not sufficient. Pólya remarked that if (25) has infinitely many solutions we can not have $a_1 \geqq 0$, $a_1 + a_2 \geqq 0$, $\cdots$, $a_1 + a_2 + \cdots + a_n \geqq 0$. The characterization of the forms which satisfy (26) seems a difficult problem.

Finally we mention two more questions:

(1) Can the inequalities $p_{n+1} - p_n < p_{n+2} - p_{n+1} < \cdots < p_{n+k} - p_{n+k-1}$ have infinitely many solutions for every fixed k?

(2) Is it true that the number of solutions of $p_{k+1} - p_k > p_k - p_{k-1}$, $k \leqq n$ is $n/2 + o(n)$? As we already have stated we can show that the number of solutions in question is between $c_1 n$ and $(1 - c_1)n$.

SYRACUSE UNIVERSITY AND
UNIVERSITY OF BUDAPEST

ON MERSENNE'S NUMBER M_{227} AND COGNATE DATA

H. S. UHLER

When p equals one of the 55 primes 2, 3, 5, $\cdots$, 257 then, strictly speaking, $M_p = 2^p - 1$ is called a Mersenne number. To obtain a clear perspective of the history of this special subject the reader may consult the interesting accurate paper by R. C. Archibald.[1] Without any superior value of p, it has been shown by E. Lucas that the prime or composite character of a number of the form $2^p - 1$ (p prime) may be investigated by employing the sequence 3, 7, 47, 2207, $\cdots$ when p is of the form $4n - 1$, and the sequence 4, 14, 194, 37634, $\cdots$ when $p = 4n + 1$. In both cases the law of formation of the terms is $s_k = s_{k-1}^2 - 2$. However, it is no longer necessary to use the $4n - 1$ Lucasian series since D. H. Lehmer[2] stated and proved the following theorem: "*The number $N = 2^n - 1$, where n is an odd prime, is a prime if, and only if, N divides the $(n-1)$st term of the series*

$$S_1 = 4,\ S_2 = 14,\ S_3 = 194, \cdots, S_k, \cdots,$$

where $S_k = S_{k-1}^2 - 2$." This justifies the use by the present writer of the second progression although 227 falls in the $4n - 1$ class.

Received by the editors July 7, 1947.

[1] R. C. Archibald, *Mersenne's numbers*, Scripta Mathematica vol. 3 (1935) pp. 112–119.

[2] D. H. Lehmer, *On Lucas's test for the primality of Mersenne's numbers*, J. London Math. Soc. vol. 9–10 (1934–1935) pp. 162–165.