P. Erdos

SOME PROBLEMS ON THE DISTRIBUTION OF PRIME NUMBERS

In this lecture I will talk about some recent questions on the distribution of primes. It is not claimed that the problems I will discuss are necessarily important ones. I will mainly speak about problems which have occupied me a great deal in the last few years.

Denote by $\pi(X)$ the number of primes not exceeding x. The prime number theorem states that $\pi(X)/ X/_{\log X} \to 1$ as $x \to \infty$. It is well known that $\sum_{k=2}^{x} 1/\log k$ gives a much better approximation to $\pi(X)$ than $x/\log x$. The statement that for $X > X(\varepsilon)$

$$(1) \qquad \left| \pi(x) - \sum_{k=2}^{x} \frac{1}{\log k} \right| < x^{1/2+\varepsilon}$$

is equivalent to the Riemann hypothesis. In our lectures we will not deal with problems like (1) but will rather consider problems of distribution of primes in the small (e.g. problems on consecutive prime numbers).

An old and very difficult problem on prime numbers requires to find to every integer n a prime number $p > n$, or in other words to be able to write down arbitrarily large prime numbers. The largest prime number which is known is $2^{2281}-1$, this number was proved to be a prime number by the electronic computer of the Institute for Numerical Analysis the S.W.A.C. in the spring of 1953. It is an unsolved problem whether there are infinitely many primes of the form 2^p-1.

A few years ago Mills[1] proved that there exists a constant $c > 1$ so that for all integers n, $[c^{3^n}]$ is a prime. Unfortunately it seems impossible to determine c explicitely. Very likely for every $c > 1$ there are infinitely many integers n for which $[c^n]$ is composite.

Put $d_n=p_{n+1}-p_n$. It immediately follows from the prime number theorem that $\overline{\lim}\, d_n/\log n \geq 1$. Backlund[2] proved that the above limit is ≥ 2 and Brauer and Zeitz[3] proved that it is ≥ 4. Westzynthius[4] proved that $\overline{\lim}\, d_n/\log n = \infty$. I proved[5] using Brun's method that for infinitemy many n

$$d_n > c \log n \log\log n / (\log\log\log n)^2$$

and When[6] proved the same result much simpler without the use of Brun's method. Finally Rankin[7] proved that for infinitely many n

$$d_n > c \log n \log\log n \log\log\log\log n / (\log\log\log n)^2. \tag{2}$$

It seems very hard to improve (2). In the opposite direction Ingham[8] proved that for $n > n_0$ $d_n > n^{5/8}$. Cramer[9] conjectured that

$$\overline{\lim}\, d_n / (\log n)^2 = 1 .$$

Using the Riemann hypothesis Cramer[9] proved that

$$\sum_{k=1}^{n} d_k^2 < c\, n (\log n)^4.$$

I conjectured[10] that

$$\sum_{k=1}^{n} d_k^2 < c\, n (\log n)^2. \tag{3}$$

$\sum_{k=1}^{n} d_k^2 > c' n (\log n)^2$ is an immediate consequence of the prime number theorem, thus (3) if true is best possible. A simpler conjecture[10] which I also can not prove is the following: Let n be any integer, denote by $a_1, a_2, \dots a_{\varphi(n)}$ the integers $\leq$ n which are relatively prime to n. Then there exists a constant C independent of n so that

$$\sum_{i=2}^{\varphi(n)} (a_i - a_{i-1})^2 < c \frac{n^2}{\varphi(n)} \tag{4}$$

Another conjecture which seems very deep is that $d_n/\log n$ has a continuous distribution function i.e. that for every $c \geqslant 0$ the density $f(c)$ of integers n for which $d_n/\log n < c$ exists and is a continuous function of c, for which $f(0)=0$, $f(\infty)=1$. Here is another problem which has the same relation to the previous one as (4) has to (3): Put $n_k=2.3.5\ldots p_k, a_1, a_2, \ldots a_{\varphi(n_k)}$ are the integers, $< n_k$ relatively prime to n_k. Denote by $A(c, n_k)$ the number of a's satisfying

$$a_i - a_{i-1} < c\, n_k/\varphi(n_k), \quad 2 \leq i \leq \varphi(n_k).$$

Then $\lim_{k=\infty} A(c,n_k)/\varphi(n_k)$ exists and is a continuous function of c.

It follows from the prime number theorem that $\underline{\lim}\, d_n/\log n \leqslant 1$, and I[10] proved that $\underline{\lim}\, d_n/\log n < 1$. It is of course to be expected that $\underline{\lim}\, d_n/\log n=0$, in fact a well known ancient conjecture states that $d_n=2$ for infinitely many n. I[11] further proved that

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

I could not prove that

$$\overline{\lim}\, (\min(d_n, d_{n+1}, d_{n+2}))/\log n = \infty, \text{ or that } \underline{\lim}\, (\max(d_n, d_{n+1}))/\log n < 1.$$

Turan and I[12] proved that $d_{n+1} > (1+\varepsilon)d_n$ and $d_{n+1} < (1-\varepsilon)d_n$ have both more than $c_1 n$ solutions in integers $m \leqslant n$. One would of course conjecture that $\overline{\lim}\, d_{n+1}/d_n = \infty$ and that $\underline{\lim}\, d_{n+1}/d_n=0$ and that the density of integers n for which $d_{n+1} > d_n$ is 1/2, but these conjectures seem very difficult to prove. Renyi and I[13] proved that the number of solution of $d_{m+1}=d_m$, m n is less than $c.n/(\log n)^{3/2}$, the right order of magnitude for the number of solutionw is very likely $c.n/\log n$, but it is not even known that the equation $d_{n+1}=d_n$ has infinitely many solutions. All the results mentioned here are obtained by Brun's method.

In our paper with Turan[12] we remark that we are unable to prove that the inequalities $d_{n+2} > d_{n+1} > d_n$ have infinitely many solutions. In fact we are even unable to prove that at least one of the inequalities $d_{n+2} > d_{n+1} > d_n$, $d_{n+2} < d_{n+1} < d_n$ have infinitely many solutions.

It seems certain that $d_n/\log n$ is everywhere dense in the interval $(0,\infty)$. We prove the following

<u>Theorem.</u>[14] <u>The set of limit points of $d_n/\log n$ form a set of positive measure.</u>

Remark. Despite the fact that the set of limit points of $d_n/\log n$ has positive measure, I can not decide about any given number whether it actually belongs to our set, thus in particular I do not know if 1 is a limit point of the numbers $d_n/\log n$.

First of all it follows from the prime number theorem that there are at least $x/4.\log x$ values of m for which

$$\frac{x}{3} < p_m < x, \quad d_m < 2 \log x. \tag{5}$$

We shall show that the set of limit points S of $d_m/\log m$ (or what is the same thing of $d_m/\log x$) of the m satisfying (5) have positive measure. (A point z is in S iff there exists an infinite sequence $x_k \to \infty$ and $m_k \to \infty$

$$\frac{1}{3} x_k < p_{m_k} < x_k, \; d_{m_k} < 2\log x_k, \text{ so that } d_{m_k}/\log x_k \to z).$$ The set S is clearly in $(0,2)$.

If our Theorem is false then by the Heine-Borel theorem S can be covered for every ε by a finite number of intervals the sum of the length of which is less than ε. Let (a_k, b_k), $k=1,2,\dots,N$, $\sum_{i=1}^{N}(b_i-a_i) < \varepsilon$ be the intervals which cover S. Let $\eta < \varepsilon/2N$. Clearly for all sufficiently large x the d_m which satisfy for some $1 \leq k \leq N$

(6) $$a_k - \eta < d_m / \log x < b_k + \eta .$$

Now we estimate the number of integers $m(p_m \leq x)$ which satisfy /6/.

A well known result of Schnirelmann[15] states that the number of solutions of $d_m = 1$, $p_m \leq x$ is less than

$$c_1 \frac{x}{(\log x)^2} \prod_{p/l} (1 + 1/p) .$$

Thus the number of solutions is m of $p_m \leq x$, d_m satisfies /6/; is less than

(7) $$c_1 \frac{x}{(\log x)^2} \sum_{k=1}^{N} \sum_{(a_k - \eta)\log x}^{(b_k+\eta)\log x} \prod_{p/l} \left(1 + \frac{1}{p}\right) < c_2 \frac{x}{(\log x)^2} \sum_{k=1}^{N} (b_k - a_k + 2\eta) \log x <$$

$$< 2 \varepsilon c_2 \frac{x}{\log x} .$$

since it is well known that

$$\sum_{ay}^{by} \prod_{p/l} \left(1 + \frac{1}{p}\right) = c(b-a)y + o(y) .$$

For sufficiently small ε /7/ contradicts /5/, thus our Theorem is proved.

Before finishing my lecture I mention a few problems on additive number theory.

Romanoff[16] proved that the density of integers of the form $a^k + p$ a integer, is positive. This result is surprising since the number of solutions of $a^k + p \leq x$ is less than $\frac{x}{c}$ x.

Generalising this result I proved [17] the following result : Let $a_1, a_2, \ldots$ be an infinite sequence of integers satisfying $a_k \mid a_{k+1}$, then the necessary and sufficient condition that the density of $p + a_k$ is positive is that for a certain c and all k

$$a_k < c^k , \quad \prod_{p/a_k} \left(1 + \frac{1}{p}\right) < c .$$

Kalmar[18] raised the problem whether for every $a > 1$ the density of integers of the form $p+[a^k]$ is positive. At present I cannot answer this question.

Following a question of Turan I proved[17] that if $f(n)$ denotes the number of solutions of $n=p+2^k$ then $\overline{\lim}\, f(n)=\infty$, in fact for infinitely many n $f(n) > c.\log\log n$. One would guess that $f(n)/\log n \to 0$; but I can not even prove that there do not exist infinitely many integers n so that for all $2^k < n$, $n-2^k$ is a prime, it seems that $n=105$ is the largest such integer.

Let $a_1 < a_2 < \dots$ be an infinite sequence of integers satisfying $a_k < c^k$. Denote again by $f(n)$ the number of solutions of $n=p+a_k$. It seems likely that $\overline{\lim}\, f(n)=\infty$.

Van der Corput[19] and I[18] proved that there exist infinitely many odd integers n not of the form 2^k+p, in fact I[18] proved that there exists an arithmetic progression consisting only of even numbers no term of which is of the form 2^k+p. I also wanted to prove that for every r there exists an arithmetic progression no term of which is of the form $p+q_r$ where q_r has not more than r prime factors. This result would easily follow if I could prove the following conjecture on congruences which seems interesting in itself: To every constant c there exists a system of congruences

$$(8) \qquad a_i \pmod{n_i}, \; c \le n_1 < n_2 < \dots < n_k$$

so that every integer satisfies at least one of the congruences (8). The simplest such system is 0(mod 2), 0(mod 3), 1(mod 4), 5(mod 6), 7(mod 12).
Dean Swift constructed such a system with $c=6$, but the general question seems very difficult.

Linnik[20] recently proved that there exists a number r so that every integer is of the form $p_1+p_2+2^{k_1}+\dots+2^{k_r}$. It seems very likely that for every r there are infinitely many integers not of the form $p+2 \quad +\ldots+2^{k_r}$.

One final problem: Is it true that to every c_1, c_2 and sufficiently large x there exist more than $c_1 \log x$ consecutive primes $\leq x$ so that the difference between any two is greater than c_2? If c_1 can be chosen sufficiently small this is well known[21].

F O O T N O T E S

1) Bull.Amer.Math.Soc. 53, 604 (1947).

2) Commentationes in honorem. Ernst Leonardi lindelöf, Helsinki 1929.

3) Sitz. Berliner Math. Ges.29, 1930, 116-125. See also H.Zeitz. Elementare Betrachtungen über eine zahlenteoretische Behauptung von Legendre.

4) Comm.Phys.-Math., Helsingfors, 5/25/1931.

5) Quarterly Journal of Math. Oxford Series , 1935, 124-128.

6) Schriften des Math. Seminars und des Institutes für angewandte Math. Der Univ.Berlin 4, 1938, 35-55.

7) London Math.Soc. Journal, 13, 1938, 242-247.

8) Quarterly Journal of Math. Oxford Series 8, 1937, 255-266.

9) Acta Arithmetica, 2, 1937, 23-45;

10) Duke Math.Journal 6, 1940, 438-441.

11) Publ.Math. 1, 1949, 33-37.

12) Bull.Amer.Math.Soc. 54, 1948, 371-378. See further P.Erdős ibid 54, 1948, 885-889.

13) Simon Stevin,Wis. Natuurk. Tijdschr. 27, 1950, 115-125.

14) Professor Ricci informed me in the discussion following my lecture, that this result is an easy consequence of some previous results of this. (Rivista di Mat.Univ.Parma, 5 (1954) 3-54).

15) E.Landau, die Goldbachsche Vermutung und der Schnirelmannsche Satz, Göttinger Nachrichten, 1930, 255-276. In fact Schnurelmann proves that the number of solutions of $p_1 - p_2 = \ell$, $p_1 < x$

is less than $c\, x/(\log x)^2 \prod_{p|l} (1+\frac{1}{p})$.

16) Math.Annalen, 109 (1934) 668-678.

17) Summa Brasiliensis Mathematicae, II (1950) 113-123.

18) Oral Communication.

19) Van der Corput, Simon Stevin, 27 (1950) 99-105,(in Dutch).
P.Erdös (in Hungarian) Math.Lapok III (1952) 122-128.

20) Trudy Math. Inst.Steklov, 38, 1951, 152-169.

21) P.Erdös, Acta Szeged, 15 (1949), 57-63. See also M. Cugiani
Annali di Matematica Serie IV, 38 (1955), 309-320.