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

\[
| \ N(-) - \sum_{k=2}^{x} \frac{1}{\log k} | < x^{1/2} + \varepsilon
\]

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 proved that there exists a constant \( c > 1 \) so that for all integers \( n \), \( \lceil c^n \rceil \) is a prime. Unfortunately it seems impossible to determine \( c \) explicitly. Very likely for every \( z > 1 \) there are infinitely many integers \( n \) for which \( \lceil c^n \rceil \) is composite.

Put \( d_n = p_{n+1} - p_n \). It immediately follows from the prime
number theorem that \( \lim_{n \to \infty} \frac{n}{\log n} = 2 \). Backlund proved that the above limit is \( \geq 2 \) and Brauer and Zeitz proved that it is \( \geq 4 \). Westzynthius proved that \( \lim d_n/\log n = \infty \). I proved using Brun’s method that for infinitely many \( n \)

\[ d_n > C \frac{\log n \log \log n}{(\log \log \log n)^2} \]

and Whan proved the same result much simpler without the use of Brun’s method. Finally Rankin proved that for infinitely many \( n \)

\[ d_n > C \frac{\log n \log \log n \log \log \log n}{(\log \log \log n)^2} \]

It seems very hard to improve (2). In the opposite direction Ingham proved that for \( n > n_0 \), \( d_n > n^{5/6} \). Cramer conjectured that

\[ \lim_{n \to \infty} \frac{d_n}{(\log n)^2} = 1 \]

Using the Riemann hypothesis Cramer proved that

\[ \sum_{k=1}^{n} d_k^2 < C n (\log n)^4 \]

I conjectured that

\[ \sum_{k=1}^{n} d_k^2 < C n (\log n)^2 \]

\[ \sum_{k=1}^{n} d_k^2 \frac{\log n}{(\log \log n)^2} \]

Is an immediate consequence of the prime number theorem, thus (3) if true is best possible. A simpler conjecture which I also cannot prove is the following: Let \( n \) be any integer, denote by \( a_1, a_2, \ldots, q(n) \) the integers \( \leq n \) which are relatively prime to \( n \). Then there exists a constant \( C \) independent of \( n \) so that

\[ \sum_{i=2}^{q(n)} (a_i - a_{i-1})^2 < C \frac{n^2}{q(n)} \]

Another conjecture which seems very deep is that \( d_n/\log n \) has
a continuous distribution function i.e. that for every $c > 0$ 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

$$\frac{a_i - a_{i-1}}{n_k / \varphi(n_k)}, \ 3 \leq i \leq \varphi(n_k).$$

Then $\lim A(c, n_k) / \varphi(n_k)$ exists and is a continuous function of $c$. It follows from the prime number theorem that $\lim d_n / \log n < 1$, and I proved that $\lim d_n / \log n < 1$. It is of course to be expected that $\lim d_n / \log n = 0$, in fact a well known ancient conjecture states that $d_n = 2$ for infinitely many $n$. I further proved that

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

I could not prove that

$$\lim \left( \min \left( d_n, d_{n+1}, d_{n+2} \right) \right) / \log n = \alpha, \text{ or that }$$

$$\lim \left( \max(d_n, d_{n+1}) \right) / \log n = 1.$$

Turan and I proved that $d_{n+1} > (1 + \varepsilon)d_n$ and $d_{n+1} < (1 - \varepsilon)d_n$ have both more than $c.n$ solutions in integers $m \leq n$. One would of course conjecture that $\lim d_{n+1} / d_n = \infty$ and that $\lim d_{n+1} / d_n = 0$ 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. Rényi and I proved that the number of solution of $d_{m+1} = d_m$, $m \leq n$ is less than $c.n / (\log n)^{3/2}$, the right order of magnitude for the number of solutions 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 Turán [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

**Theorem.** The set of limit points of $d_n/\log n$ form a set of positive measure.

**Remark.** Despite the fact that the set of limit points of $d_n/\log n$ has positive measure, I cannot 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

$$\frac{x}{4} < \mu_m < x, \quad d_{\mu_n} < 2 \log x.$$  

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 $S$ if there exists an infinite sequence $x_1, x_2, \ldots \to \infty$ and $m_1, m_2, \ldots \to \infty$,

$$\frac{1}{3} x_k < \mu_m < x_k, \quad d_{\mu_n} < 2 \log x, \quad \text{so that } d_{\mu_n}/\log x \to 2).$$  

If our Theorem is false then by the Heine-Borel theorem $S$ can be covered for every $\varepsilon$ by a finite number of intervals the sum of the lengths of which is less than $\varepsilon$. Let \((a_k, b_k)_1, b_k, \ldots, b_k, a_k, a_k, \ldots, a_k\) be the intervals which cover $S$. Let $\eta < \varepsilon/2N$. Clearly for all sufficiently large $x$ the $d_m$ which
satisfy for some $1 \leq k \leq N$

\begin{equation}
(6)\quad a_k - \eta < d_m / \log x < b_k + \eta
\end{equation}

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

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

\[
C_1 \frac{x}{(\log x)^2} \prod_{\nu} (1 + 1/\nu).
\]

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

\begin{equation}
(7)\quad C_1 \frac{x}{(\log x)^2} \sum_{\nu} \sum_{k+x} \left( \prod_{\nu} \left(1 + \frac{1}{\nu}\right) \right) < c \frac{x}{(\log x)^2} \sum_{\nu} \left( b_k - a_k + 2 \eta \right) \log x
\end{equation}

\[
< 2 \varepsilon C_2 \frac{x}{(\log x)^2}
\]

since it is well known that

\[
\sum_{\nu} \prod_{\nu} \left( 1 + \frac{1}{\nu} \right) = (e - 1) \zeta(\nu) + O(1).\]

For sufficiently small $\varepsilon$ /7/ contradicts /5/, thus our Theorem is proved.

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

Romanoff\textsuperscript{16}) proved that the density of integers of the form $a^k + p$ is integer, is positive. This result is surprising since the number of solutions of $a^k + p \leq x$ is less than $g(x)$. Generalising this result I proved the following result:

Let $a_1, a_2, \ldots$ be an infinite sequence of integers satisfying $a_k \mid a_{x+1}$, then the necessary and sufficient condition that the density of $p + a_k$ is positive is that for a certain $c$ and all $k$

\[
a_k < c^k \prod_{p \mid a_k} \left( 1 + \frac{1}{p} \right) < \infty.
\]
Kalmar\(^{18}\) raised the problem whether for every \(a > 1\) the density of integers of the form \(p + \left[ a^k \right]\) is positive. At present I cannot answer this question.

Following a question of Turán I proved\(^{17}\) that if \(f(n)\) denotes the number of solutions of \(n = p + 2^k\) then \(\lim f(n) = \infty\), 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 \(a_1 < a_2 < ...\) 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 \(\lim f(n) = \infty\).

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

\[
(8) \quad a \equiv c \pmod{\lambda_i} \quad \gamma \leq \lambda_1 < \lambda_2 < \cdots < \lambda_k
\]

so that every integer satisfies at least one of the congruences \((8)\). The simplest such system is \(0 \pmod{2}\), \(0 \pmod{3}\), \(1 \pmod{4}\), \(5 \pmod{6}\), \(7 \pmod{12}\).

Dean Swift constructed such a system with \(c = 5\), 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 \(p_1 + p_2 + 2^k + \cdots + 2^r\). It seems very
likely that for every \( r \) there are infinitely many integers not of the form \( p+2 + \ldots + 2^{kr} \).

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 \( c_2^2 \)? If \( c_1 \) can be chosen sufficiently small this is well known.

---

**Footnotes**

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, 285-276. In fact Schnirelmann proves that the number of solutions of \( p_1 - p_2 = \ell \), \( p_1 < x \)
is less than \( c \frac{x}{(\log x)^2} \prod_{p \leq x} \left(1 + \frac{1}{p}\right) \).


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

18) Oral Communication.


