ANNALS OF MATHEMATICS
Vol. 50, No. 2, April, 1949

AN ELEMENTARY PROOF OF THE PRIME-NUMBER THEOREM

ATLE SELBERG

(Received October 14, 1948)

1. Introduction

In this paper will be given a new proof of the prime-number theorem, which is elementary in the sense that it uses practically no analysis, except the simplest properties of the logarithm.

We shall prove the prime-number theorem in the form

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1 \tag{1.1}$$

where for $x > 0$, $\vartheta(x)$ is defined as usual by

$$\vartheta(x) = \sum_{p \leqq x} \log p, \tag{1.2}$$

p denoting the primes.

The basic new thing in the proof is a certain assymptotic formula (2.8), which may be written

$$\vartheta(x) \log x + \sum_{p \leqq x} \log p \, \vartheta\left(\frac{x}{p}\right) = 2x \log x + O(x). \tag{1.3}$$

From this formula there are several ways to deduce the prime-number theorem. The way I present §§2–4 of this paper, is chosen because it seems at the present to be the most direct and most elementary way.[1] But for completeness it has to be mentioned that this was not my first proof. The original proof was in fact rather different, and made use of the following result by P. Erdös, that for an arbitrary, positive fixed number δ, there exist a $K(\delta) > 0$ and an $x_0 = x_0(\delta)$ such that for $x > x_0$, there are more than

$$K(\delta)\, x/\log x$$

primes in the interval from x to $x + \delta x$.

My first proof then ran as follows: Introducing the notations

$$\varliminf \frac{\vartheta(x)}{x} = a, \qquad \varlimsup \frac{\vartheta(x)}{x} = A,$$

one can easily deduce from (1.3), using the well-known result

$$\sum_{p \leqq x} \frac{\log p}{x} = \log x + O(1), \tag{1.4}$$

[1] Because it avoids the concept of lower and upper limit. It is in fact easy to modify the proof in a few places so as to avoid the concept of limit at all, of course (1.1) would then have to be stated differently.

that

$$a + A = 2. \tag{1.5}$$

Next, taking a large x, with

$$\vartheta(x) = ax + o(x),$$

one can deduce from (1.3) in the modified form

$$(\vartheta(x) - ax) \log x + \sum_{p \leq x} \log p \left(\vartheta\left(\frac{x}{p}\right) - A \frac{x}{p} \right) = O(x), \tag{1.6}$$

that, for a fixed positive number δ, one has

$$\vartheta\left(\frac{x}{p}\right) > (A - \delta) \frac{x}{p}, \tag{1.7}$$

except for an exceptional set of primes $\leqq x$ with

$$\sum \frac{\log p}{p} = o(\log x).$$

Also one easily deduces that there exists an x' in the range $\sqrt{x} < x' < x$, wit

$$\vartheta(x') = Ax' + o(x').$$

Again from (1.6) with a and A interchanged, and x' instead of x, one deduces tha

$$\vartheta\left(\frac{x'}{p}\right) < (a + \delta) \frac{x'}{p}, \tag{1.8}$$

except for an exceptional set of primes $\leqq x'$ with

$$\sum \frac{\log p}{p} = o(\log x).$$

From Erdös' result it is then possible to show that one can chose primes p and p' not belonging to any of the exceptional sets, with

$$\frac{x}{p} < \frac{x'}{p'} < (1 + \delta) \frac{x}{p}.$$

Then we get from (1.7) and (1.8) that

$$(A - \delta) \frac{x}{p} < \vartheta\left(\frac{x}{p}\right) \leqq \vartheta\left(\frac{x'}{p'}\right) < (a + \delta) \frac{x'}{p'} < (a + \delta)(1 + \delta) \frac{x}{p},$$

so that

$$A - \delta < (a + \delta)(1 + \delta).$$

or making δ tend to zero

$$A \leqq a.$$

Hence since also $A \geqq a$ and $a + A = 2$ we have $a = A = 1$, which proves our theorem.

Erdös' result was obtained without knowledge of my work, except that it is based on my formula (2.8); and after I had the other parts of the above proof. His proof contains ideas related to those in the above proof, at which related ideas he had arrived independently.

The method can be applied also to more general problems. For instance one can prove some theorems proved by analytical means by Beurling, but the results are not quite as sharp as Beurlings.[2] Also one can prove the prime-number theorem for arithmetic progressions, one has then to use in addition ideas and results from my previous paper on Dirichlets theorem.[3]

Of known results we use frequently besides (1.4) also its consequence

$$\vartheta(x) = O(x). \tag{1.9}$$

Throughout the paper p, q and r denote prime numbers. $\mu(n)$ denotes Möbius' number-theoretic function, $\tau(n)$ denotes the number of divisors of n. The letter c will be used to denote absolute constants, and K to denote absolute positive constants. Some of the more trivial estimations are not carried out but left to the reader.

2. Proof of the basic formulas

We write, when x is a positive number and d a positive integer,

$$\lambda_d = \lambda_{d,x} = \mu(d) \log^2 \frac{x}{d}, \tag{2.1}$$

and if n is a positive integer,

$$\theta_n = \theta_{n,x} = \sum\nolimits_{d/n} \lambda_d . \tag{2.2}$$

Then we have

$$\theta_n = \begin{cases} \log^2 x, \text{ for } n = 1, \\ \log p \log x^2/p, \text{ for } n = p^\alpha, \alpha \geqq 1, \\ 2 \log p \log q, \text{ for } n = p^\alpha q^\beta, \alpha \geqq 1, \beta \geqq 1, \\ 0, \text{ for all other } n. \end{cases} \tag{2.3}$$

The first three of these statements follow readily from (2.2) and (2.1), the fourth is easily proved by induction. Clearly it is enough to consider n square-free, then if $n = p_1 p_2 \cdots p_k$,

$$\theta_{n,x} = \theta_{n/p_k,x} - \theta_{n/p_k,x/p_k} .$$

From this the remaining part of (2.3) follows.

[2] A. Beurling: *Analyse de la loi asymptotique de la distribution des nombres premiers généralisés*, Acta Math., vol. 68, pp. 255–291 (1937).

[3] These Annals this issue, pp. 297–304.

Now consider the expression

$$(2.4)\quad \begin{aligned}\sum_{n\leqq x}\theta_n &= \sum_{n\leqq x}\sum_{d/n}\lambda_d = \sum_{d\leqq x}\lambda_d\left[\frac{x}{d}\right] = x\sum_{d\leqq x}\frac{\lambda_d}{d} + O\,(\sum_{d\leqq x}|\lambda_d|)\\ &= x\sum_{d\leqq x}\frac{\mu(d)}{d}\log^2\frac{x}{d} + O\left(\sum_{d\leqq x}\log^2\frac{x}{d}\right) = x\sum_{d\leqq x}\frac{\mu(d)}{d}\log^2\frac{x}{d} + O(x).\end{aligned}$$

This on the other hand is equal to, by (2.3),

$$(2.5)\quad \begin{aligned}\sum_{n\leqq x}\theta_n = \log^2 x + \sum_{p^\alpha\leqq x}\log p\log\frac{x^2}{p}\\ + 2\sum_{\substack{p^\alpha q^\beta\leqq x\\ p<q}}\log p\log q = \sum_{p\leqq x}\log^2 p\\ + \sum_{pq\leqq x}\log p\log q + O\left(\sum_{p\leqq x}\log p\log\frac{x}{p}\right)\\ + O\,(\sum_{\substack{p^\alpha\leqq x\\ \alpha>1}}\log^2 x) + O\,(\sum_{\substack{p^\alpha q^\beta\leqq x\\ \alpha>1}}\log p\log q)\\ + \log^2 x = \sum_{p\leqq x}\log^2 p + \sum_{pq\leqq x}\log p\log q + O(x).\end{aligned}$$

The remainder term being obtained by use of (1.4) and (1.9). Hence from (2.4) and (2.5),

$$(2.6)\quad \sum_{p\leqq x}\log^2 p + \sum_{pq\leqq x}\log p\log q = x\sum_{d\leqq x}\frac{\mu(d)}{d}\log^2\frac{x}{d} + O(x).$$

It remains now to estimate the sum on the right-hand-side. To this purpose we need the formulas

$$(2.7)\quad \sum_{\nu\leqq z}\frac{1}{\nu} = \log z + c_1 + O(z^{-\frac{1}{2}}),$$

and

$$(2.7')\quad \sum_{\nu\leqq z}\frac{\tau(\nu)}{\nu} = \tfrac{1}{2}\log^2 z + c_2\log z + c_3 + O(z^{-\frac{1}{2}})$$

where the c's are absolute constants, (2.7) is well known, and (2.7′) may be easily derived by partial summation from the well-known result

$$\sum_{\nu\leqq z}\tau(\nu) = z\log z + c_4 z + O(\sqrt{z}).$$

From (2.7) and (2.7′) we get

$$\log^2 z = 2\sum_{\nu\leqq z}\frac{\tau(\nu)}{\nu} + c_5\sum_{\nu\leqq z}\frac{1}{\nu} + c_6 + O(z^{-\frac{1}{2}}).$$

By taking here $z = x/d$, we get

$$\sum_{d \leqq x} \frac{\mu(d)}{d} \log^2 \frac{x}{d} = 2 \sum_{d \leqq x} \frac{\mu(d)}{d} \sum_{\nu \leqq x/d} \frac{\tau(\nu)}{\nu} + c_5 \sum_{d \leqq x} \frac{\mu(d)}{d} \sum_{\nu \leqq x/d} \frac{1}{\nu}$$

$$+ c_6 \sum_{d \leqq x} \frac{\mu(d)}{d} + O(x^{-\frac{1}{2}} \sum_{d \leqq x} d^{-\frac{1}{2}}) = 2 \sum_{d\nu \leqq x} \frac{\mu(d)\tau(\nu)}{d\nu}$$

$$+ c_5 \sum_{d\nu \leqq x} \frac{\mu(d)}{d\nu} + c_6 \sum_{d \leqq x} \frac{\mu(d)}{d} + O(1)$$

$$= 2 \sum_{n \leqq x} \frac{1}{n} \sum_{d/n} \mu(d) \tau\left(\frac{n}{d}\right) + c_5 \sum_{n \leqq x} \frac{1}{n} \sum_{d/n} \mu(d)$$

$$+ O(1) = 2 \sum_{n \leqq x} \frac{1}{n} + c_5 + O(1) = 2 \log x + O(1).$$

We used here that $\sum_{d/n} \mu(d) \tau(n/d) = 1$, and the well-known $\sum_{d \leqq x} (\mu(d))/d = O(1)$. Now (2.6) yields

$$\sum_{p \leqq x} \log^2 p + \sum_{pq \leqq x} \log p \log q = 2x \log x + O(x). \tag{2.8}$$

This formula may also be written in the form given in the introduction

$$\vartheta(x) \log x + \sum_{p \leqq x} \log p \, \vartheta\left(\frac{x}{p}\right) = 2x \log x + O(x), \tag{2.9}$$

by noticing that

$$\sum_{p \leqq x} \log^2 p = \vartheta(x) \log x + O(x).$$

By partial summation we get from (2.8)

$$\sum_{p \leqq x} \log p + \sum_{pq \leqq x} \frac{\log p \log q}{\log pq} = 2x + O\left(\frac{x}{\log x}\right). \tag{2.10}$$

This gives

$$\sum_{pq \leqq x} \log p \log q = \sum_{p \leqq x} \log p \sum_{q \leqq x/p} \log q = 2x \sum_{p \leqq x} \frac{\log p}{p}$$

$$- \sum_{p \leqq x} \log p \sum_{qr \leqq x/p} \frac{\log q \log r}{\log qr} + O\left(x \sum_{p \leqq x} \frac{\log p}{p\left(1 + \log \dfrac{x}{p}\right)}\right)$$

$$= 2x \log x - \sum_{qr \leqq x} \frac{\log q \log r}{\log qr} \vartheta\left(\frac{x}{qr}\right) + O(x \log \log x).$$

Inserting this for the second term in (2.8) we get

$$\vartheta(x) \log x = \sum_{pq \leqq x} \frac{\log p \log q}{\log pq} \vartheta\left(\frac{x}{pq}\right) + O(x \log \log x). \tag{2.11}$$

Writing now

$$\vartheta(x) = x + R(x)$$

, (2.9) easily gives

$$R(x) \log x = -\sum_{p \leq x} \log p \, R\left(\frac{x}{p}\right) + O(x), \tag{2.12}$$

and (2.11) yields in the same manner

$$R(x) \log x = \sum_{pq \leq x} \frac{\log p \log q}{\log pq} R\left(\frac{x}{pq}\right) + O(x \log \log x), \tag{2.13}$$

since

$$\sum_{pq \leq x} \frac{\log p \log q}{pq \log pq} = \log x + O(\log \log x),$$

which follows by partial summation from

$$\sum_{pq \leq x} \frac{\log p \log q}{pq} = \tfrac{1}{2} \log^2 x + O(\log x),$$

which again follows easily from (1.4).

The (2.12) and (2.13) yield

$$2 \mid R(x) \mid \log x \leqq \sum_{p \leq x} \log p \left| R\left(\frac{x}{p}\right) \right| + \sum_{pq \leq x} \frac{\log p \log q}{\log pq} \left| R\left(\frac{x}{pq}\right) \right| + O(x \log \log x).$$

From this, by partial summation,

$$2 \mid R(x) \mid \log x \leqq \sum_{n \leq x} \left\{ \sum_{p \leq n} \log p + \sum_{pq \leq n} \frac{\log p \log q}{\log pq} \right\} \cdot \left\{ \left| R\left(\frac{x}{n}\right) \right| - \left| R\left(\frac{x}{n+1}\right) \right| \right\} + O(x \log \log x),$$

or by (2.10)

$$\begin{aligned}
2 \mid R(x) \mid \log x &\leqq 2 \sum_{n \leq x} n \left\{ \left| R\left(\frac{x}{n}\right) \right| - \left| R\left(\frac{x}{n+1}\right) \right| \right\} \\
&\quad + O\left(\sum_{n \leq x} \frac{n}{1 + \log n} \left| R\left(\frac{x}{n}\right) - R\left(\frac{x}{n+1}\right) \right| \right) + O(x \log \log x) \\
&= 2 \sum_{n \leq x} \left| R\left(\frac{x}{n}\right) \right| + O\left(\sum_{n \leq x} \frac{n}{1 + \log n} \left\{ \vartheta\left(\frac{x}{n}\right) - \vartheta\left(\frac{x}{n+1}\right) \right\} \right) \\
&\quad + O\left(x \sum_{n \leq x} \frac{1}{n(1 + \log n)} \right) + O(x \log \log x) = 2 \sum_{n \leq x} \left| R\left(\frac{x}{n}\right) \right| \\
&\quad + O\left(\sum_{n \leq x} \frac{1}{1 + \log n} \vartheta\left(\frac{x}{n}\right) \right) + O(x \log \log x) \\
&= 2 \sum_{n \leq x} \left| R\left(\frac{x}{n}\right) \right| + O(x \log \log x),
\end{aligned}$$

or

$$(2.14) \qquad | R(x) | \leqq \frac{1}{\log x} \sum_{n \leqq x} \left| R\left(\frac{x}{n}\right) \right| + O\left(x \frac{\log \log x}{\log x}\right),$$

which is the result we will use in the following.[4]

3. Some properties of $R(x)$

From (1.4) we get by partial summation that

$$\sum_{n \leqq x} \frac{\vartheta(n)}{n^2} = \log x + O(1),$$

or

$$\sum_{n \leqq x} \frac{R(n)}{n^2} = O(1).$$

This means there exists an absolute positive constant K_1, so that for all $x > 4$ and $x' > x$,

$$(3.1) \qquad \left| \sum_{x \leqq n \leqq x'} \frac{R(n)}{n^2} \right| < K_1 .$$

Accordingly we have, if $R(n)$ does not change its sign between x and x', that there is a y in the interval $x \leqq y \leqq x'$, so that

$$(3.2) \qquad \left| \frac{R(y)}{y} \right| < \frac{K_2}{\log \frac{x'}{x}}, \qquad K_2 \geqq 1.$$

This is easily seen to hold true if $R(n)$ changes the sign also.[5]

Thus for an arbitrary fixed positive $\delta < 1$ and $x > 4$, there will exist a y in the interval $x \leqq y \leqq e^{K_2/\delta} x$, with

$$(3.3) \qquad | R(y) | < \delta y.$$

From (2.10) we see that for $y < y'$,

$$0 \leqq \sum_{y < p \leqq y'} \log p \leqq 2(y' - y) + O\left(\frac{y'}{\log y'}\right),$$

from which follows that

$$| R(y') - R(y) | \leqq y' - y + O\left(\frac{y'}{\log y'}\right).$$

---

[4] Apparently we have here lost something in the order of the remainder-term compared to (2.8). Actually we could instead of (2.14) have used the inequality

$$| R(x) | \leqq \frac{2}{\log^2 x} \sum_{n \leqq x} \frac{\log n}{n} \left| R\left(\frac{x}{n}\right) \right| + O\left(\frac{x}{\log x}\right),$$

which can be proved in a similar way.

[5] Because there will then be a $| R(y) | < \log y$.

Hence, if $y/2 \leqq y' \leqq 2y$, $y > 4$,

$$| R(y') - R(y) | \leqq | y' - y | + O\left(\frac{y'}{\log y'}\right),$$

or

$$| R(y') | \leqq | R(y) | + | y' - y) + O\left(\frac{y'}{\log y'}\right).$$

Now consider an interval $(x, e^{K_2/\delta} x)$, according to (3.3) there exists a in this interval with

$$| R(y) | < \delta y.$$

Thus for any y' in the interval $y/2 \leqq y' \leqq 2\, y$, we have

$$| R(y') | \leqq \delta y + | y' - y | + \frac{K_3 y'}{\log x},$$

or

$$\left|\frac{R(y')}{y'}\right| < 2\delta + \left|1 - \frac{y'}{y}\right| + \frac{K_3}{\log x}.$$

Hence if $x > e^{K_3/\delta}$ and $e^{-(\delta/2)} \leqq y'/y \leqq e^{\delta/2}$, we get

$$\left|\frac{R(y')}{y'}\right| < 2\delta + (e^{\delta/2} - 1) + \delta < 4\delta.$$

Thus for $x > e^{K_3/\delta}$ the interval $(x, e^{K_2/\delta}x)$ will always contain a sub-interv $(y_1, e^{\delta/2}y_1)$, such that $| R(z) | < 4\delta z$ if z belongs to this sub-interval.

4. Proof of the prime-number theorem

We are now going to prove the

THEOREM.

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1.$$

Obviously this is equivalent to

$$\lim_{x\to\infty} \frac{R(x)}{x} = 0. \tag{4.1}$$

We know that for $x > 1$,

$$| R(x) | < K_4 x. \tag{4.2}$$

Now assume that for some positive number $\alpha < 8$,

$$| R(x) | < \alpha x, \tag{4.3}$$

holds for all $x > x_0$. Taking $\delta = \alpha/8$, we have according to the precedir section (since we may assume that $x_0 > e^{K_3/\delta}$), that all intervals of the type $(x, e^{K_2/\delta}x)$ with $x > x_0$, contain an interval $(y, e^{\delta/2}y)$ such that

$$|R(z)| < \alpha z/2, \tag{4.4}$$

for $y \leqq z \leqq e^{\delta/2}y$.

The inequality (2.14) then gives, using (4.2),

$$\begin{aligned}|R(x)| &\leqq \frac{1}{\log x}\sum_{n\leqq x}\left|R\left(\frac{x}{n}\right)\right| + O\left(\frac{x}{\sqrt{\log x}}\right)\\ &< K_4\frac{x}{\log x}\sum_{(x/x_0)<n\leqq x}\frac{1}{n} + \frac{x}{\log x}\sum_{n\leqq(x/x_0)}\frac{1}{n}\left|\frac{n}{x}R\left(\frac{x}{n}\right)\right| + O\left(\frac{x}{\sqrt{\log x}}\right),\end{aligned}$$

writing now $\rho = e^{K_2/\delta}$, we get further, using (4.3) and (4.4),

$$\begin{aligned}|R(x)| &< \frac{\alpha x}{\log x}\sum_{n\leqq(x/x_0)}\frac{1}{n} - \frac{\alpha x}{2\log x}\sum_{1\leqq\nu\leqq(\log(x/x_0)/\log\rho)}\\ &\sum_{\substack{y_\nu\leqq n\leqq y_\nu e^{(\delta/2)}\\ \rho^{\nu-1}<y_\nu\leqq\rho^\nu e^{-(\delta/2)}}}\frac{1}{n} + O\left(\frac{x}{\sqrt{\log x}}\right) = \alpha x - \frac{\alpha x}{2\log x}\sum_{1\leqq\nu\leqq(\log(x/x_0)/\log\rho)}\frac{\delta}{2}\\ &\quad + O\left(\frac{x}{\sqrt{\log x}}\right) = \alpha x - \frac{\alpha\delta}{4\log\rho}x + O\left(\frac{x}{\sqrt{\log x}}\right)\\ &\quad = \alpha\left(1 - \frac{\alpha^2}{256K_2}\right)x + O\left(\frac{x}{\sqrt{\log x}}\right) < \alpha\left(1 - \frac{\alpha^2}{300K_2}\right)x,\end{aligned}$$

for $x > x_1$. Since the iteration-process

$$\alpha_{n+1} = \alpha_n\left(1 - \frac{\alpha_n^2}{300K_2}\right),$$

obviously converges to zero if we start for instance with $\alpha_1 = 4$ (one sees easily that then $\alpha_n < K_5/\sqrt{n}$), this proves (4.1) and thus our theorem.

Final remark. As one sees we have actually never used the full force of (2.8) in the proof, we could just as well have used it with the remainder term $o(x \log x)$ instead of $O(x)$. It is not necessary to use the full force of (1.4) either, if we have here the remainder-term $o(\log x)$ but in addition knowing that $\vartheta(x) > Kx$ for $x > 1$ and some positive constant K, we can still prove the theorem. However, we have then to make some change in the arguments of §3.

The Institute for Advanced Study
and
Syracuse University