ON THE ELEMENTARY PROOF OF THE PRIME NUMBER THEOREM

by N. LEVINSON†
(Received 7th May 1966)

1. Perhaps the simplest elementary proof of the prime number theorem, see Erdös (2) and Selberg (5), is Wright's modification (8), (3, p. 362) of Selberg's original proof (5). Another variant is due to V. Nevanlinna (4). Wright's proof uses Selberg's idea of smoothing the weighting process which occurs in the Selberg inequality, (1.2) below, by iterating this inequality. Here it will be shown that the proof requires less ingenuity if use is made of a further smoothing operation, namely first integrating the Selberg inequality itself. Integration has been used on a related inequality by Breusch (1) to obtain a remainder term. This method also makes proof by contradiction unnecessary.

Recall that by definition \( \Lambda(n) = \log p \) for \( n = p^j \), \( p \) a prime number and \( j \) a positive integer, and \( \Lambda(n) = 0 \) otherwise. As customary let

\[
\psi(x) = \sum_{n \leq x} \Lambda(n) \quad (\psi(x) = 0, \ x < 2).
\]

The prime number theorem is equivalent [3, p. 345] to

\[
\lim_{x \to \infty} \frac{\psi(x)}{x} = 1. \quad (1.1)
\]

The Selberg inequality [3, p. 359] is

\[
\psi(x) \log x + \sum_{n \leq x} \Lambda(n) \psi \left( \frac{x}{n} \right) = 2x \log x + O(x) \quad (1.2)
\]

for large \( x \), where the summation is finite since \( \psi(x) = 0 \) for \( x < 2 \). The simplest proof of (1.2) is probably that of Tatuzawa and Iseki (7), see also Shapiro (6). Since the terms in the sum in (1.2) are non-negative, deleting the sum leads to an inequality that implies

\[
0 \leq \lim_{x \to \infty} \sup \frac{\psi(x)}{x} \leq 2. \quad (1.3)
\]

If \( f(t) \) has a continuous derivative and if \( c_n, n \geq 1 \), are constants then partial summation leads readily [3, p. 346] to

\[
\sum_{n \leq x} c_n f(n) = C(x) f(x) - \int_1^x C(t) f'(t) \, dt; \quad C(t) = \sum_{n \geq t} c_n. \quad (1.4)
\]

† The preparation of this work was supported in part by the Office of Naval Research and by the National Science Foundation NSF Gp-4364.
With \( c_n = \Lambda(n) \), (1.4) and (1.3) yield
\[
\sum_{n \leq x} \Lambda(n) \log n = \psi(x) \log x + O(x). \tag{1.5}
\]
Also
\[
\sum_{j \leq x} \Lambda(j) \psi \left( \frac{x}{j} \right) = \sum_{j \leq x} \Lambda(j) \sum_{k \leq x/j} \Lambda(k) = \sum_{j \leq x} \Lambda(j) \Lambda(k). \tag{1.6}
\]
Thus, if
\[
\Lambda_2(n) = \Lambda(n) \log n + \sum_{jk \leq n} \Lambda(j) \Lambda(k), \tag{1.7}
\]
then (1.5) and (1.6) in (1.2) yield
\[
\sum_{n \leq x} \Lambda_2(n) = 2 \log x + O(x)
\]
as an equivalent to (1.2). Using (1.4) with \( c_n = 1 \)
\[
\sum_{n \leq x} \log n = x \log x + O(x). \tag{1.8}
\]
Combining the above two inequalities
\[
Q(n) = \sum_{k \leq n} (\Lambda_2(k) - 2 \log k) = O(n). \tag{1.9}
\]
The basic property of \( \Lambda(n) \) is \( \sum_{d \mid n} \Lambda(d) = \log n \). Summing this for \( n \leq x \) and using (1.8) yields the well known
\[
\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + O(1), \tag{1.10}
\]
which by (1.4) is equivalent to
\[
\int_2^x \frac{\psi(t)}{t^2} \, dt = \log x + O(1). \tag{1.11}
\]
2. If \( R(x) = \psi(x) - x, x \geq 2; \) \( R(x) = 0, x<2, \) then, from (1.2) and (1.10),
\[
R(x) \log x + \sum \Lambda(n) R \left( \frac{x}{n} \right) = O(x). \tag{2.1}
\]
From (1.3)
\[
\lim \sup_{x \to \infty} \left| \frac{R(x)}{x} \right| = 1 \tag{2.2}
\]
so that there exists \( c < \infty \) such that
\[
\left| R(x) \right| \leq c \left| x \right|. \tag{2.3}
\]
Let \( S(y) = 0, y \leq 2, \) and for \( y > 2 \) let
\[
S(y) = \int_y^\infty \frac{R(x)}{x} \, dx. \tag{2.4}
\]
Then except at \( y = p^l \), where \( R(y) \) is discontinuous, \( S'(y) = R(y)/y \) and so by (2.3),
\[
\left| S'(y) \right| \leq c, \quad y \neq p^l. \tag{2.5}
\]
Hence first, for \( y_1 \) and \( y_2 \) in an interval not containing any \( p^j \) in its interior,

\[
\| S(y_2) - S(y_1) \| \leq c \| y_2 - y_1 \|
\]  

(2.6)

which however then implies the above for all \( y_1 \) and \( y_2 \) since \( S \) is continuous. But (2.6) implies

\[
\| S(y_2) - |S(y_1)| \| \leq c \| y_2 - y_1 \|.
\]  

(2.7)

Replacing \( n \) by \( j \) in (2.1), dividing by \( x \), and integrating gives

\[
\int_2^y \frac{R(x)}{x} \log x \, dx + \Sigma \Lambda(j) \int_2^y R \left( \frac{x}{j} \right) \frac{dx}{x} = O(y).
\]

Integrating the first term by parts and using (2.3) this becomes

\[
S(y) \log y + \Sigma \Lambda(j) S \left( \frac{y}{j} \right) = O(y).
\]  

(2.8)

Replacing \( y \) by \( y/k \) in (2.8), multiplying by \( \Lambda(k) \) and summing for \( k \leq y \)

\[
\Sigma \Lambda(k) S \left( \frac{y}{k} \right) \log \frac{y}{k} + \Sigma \Sigma \Lambda(k) \Lambda(j) S \left( \frac{y}{jk} \right) = O(y) \sum_{k \leq y} \frac{\Lambda(k)}{k}.
\]

Using \( \log \frac{y}{k} = \log y - \log k \) and replacing this \( k \) by \( m \),

\[
\log y \Sigma \Lambda(k) S \left( \frac{y}{k} \right) - \Sigma S \left( \frac{y}{m} \right) (\Lambda(m) \log m - \sum_{j \leq m} \Lambda(j) \Lambda(k)) = O(y \log y).
\]

Replacing the first sum by use of (2.8) and using (1.7) the above implies

\[
\log^2 y \mid S(y) \mid \leq \Sigma \mid S \left( \frac{y}{m} \right) \mid \Lambda_2(m) + K_1 y \log y
\]  

(2.9)

for some constant \( K_1 \).

Consider the identity

\[
\Sigma \left| S \left( \frac{y}{m} \right) \right| \Lambda_2(m) = 2 \sum_{m \leq y} \left| S \left( \frac{y}{m} \right) \right| \log m + J(y),
\]  

(10.10)

where, recalling the definition of \( Q(n) \) in (1.9),

\[
J(y) = 2 \sum_{m \leq y} (Q(m) - Q(m-1)) \left| S \left( \frac{y}{m} \right) \right|, \quad Q(1) = 0;
\]

\[
J(y) = \Sigma Q(m) \left( \left| S \left( \frac{y}{m} \right) \right| - \left| S \left( \frac{y}{m+1} \right) \right| \right).
\]

Using (1.9) and (2.7) there is a \( K_2 \) such that

\[
J(y) \leq K_2 \sum_{m \leq y} m \left( \frac{y}{m} - \frac{y}{m+1} \right) \leq K_2 y \log y.
\]  

(2.11)
Returning to the terms in the sum in (2.10)

\[ \log m \left| \frac{y}{m} \right| \leq \int_{m}^{m+1} \log u \left| S \left( \frac{y}{u} \right) \right| \, du + J_m, \]

\[ J_m = \int_{m}^{m+1} \log u \left( \left| S \left( \frac{y}{m} \right) \right| - \left| S \left( \frac{y}{u} \right) \right| \right) \, du. \]

By (2.7)

\[ J_m \leq c \left( \frac{y}{m} - \frac{y}{m+1} \right) \int_{m}^{m+1} \log u \, du \leq \frac{cy \log (m+1)}{m(m+1)}, \]

Using \( \log (m+1) \leq m, \)

\[ \log m \left| \frac{y}{m} \right| \leq \int_{m}^{m+1} \log u \left| S \left( \frac{y}{u} \right) \right| \, du + \frac{cy \log (m+1)}{m+1}. \]  

Using (2.11) and (2.12) in (2.10) there follows from (2.9) the key inequality

\[ \log^2 y \left| S(y) \right| \leq 2 \int_{\frac{y}{2}}^{y} \log u \left| S \left( \frac{y}{u} \right) \right| \, du + K_3 y \log y. \]  

3. From \( R(x) = \psi(x) - x \) and (1.11)

\[ \int_{\frac{y}{2}}^{y} R(u) \frac{du}{u^2} = O(1). \]

From this and (2.3) it can be easily verified that

\[ \int_{\frac{y}{2}}^{y} \frac{S(t)}{t^2} \, dt = O(1). \]  

Let \( T(x) = e^{-x}S(e^x). \) By (2.2) and (2.4)

\[ \alpha = \limsup_{x \to \infty} \left| T(x) \right| \leq 1. \]  

By (2.5), if \( k = 2c, \)

\[ \left| T'(x) \right| \leq k, \quad x \neq j \log p, \]

and by the argument following (2.5) this implies

\[ \| T(x_2) - T(x_1) \| \leq k \left| x_2 - x_1 \right|. \]

By (3.1)

\[ \int_{\log 2}^{x} T(u) \, du = O(1), \]

which shows there is a constant \( M \) such that

\[ \left| \int_{x_1}^{x_2} T(u) \, du \right| \leq M, \]

and one can assume \( M \) so large that \( Mk \geq 1. \)
The key inequality (2.13) with \( x = \log y, s = \log y/u \) (noting that \( T(x) = 0 \) for \( x < \log 2 \)) becomes
\[
| T(x) | \leq \frac{2}{x^2} \int_0^x (x-s) | T(s) | \, ds + \frac{K_3}{x}
\]
or, equivalently,
\[
| T(x) | \leq \frac{2}{x^2} \int_0^x \frac{u}{v} \left( \frac{1}{v} \int_0^u | T(s) | \, ds \right) + \frac{K_3}{x}.
\] (3.5)

From the definition of \( \alpha \) in (3.2),
\[
\gamma = \limsup_{x \to \infty} \frac{1}{x} \int_0^x | T(s) | \, ds \leq \alpha.
\]

But (3.5) shows that \( \alpha \leq \gamma \) so that indeed \( \alpha = \gamma \). Next (3.3), (3.4) and \( \alpha = \gamma \) will be used to show that \( \alpha = 0 \), which leads easily to (1.1).

From (3.2), given \( \beta > \alpha \), there exists \( \xi \) such that
\[
| T(x) | \leq \beta \text{ for } x \geq \xi.
\] (3.6)

If \( T(x) \) has no zeros for large \( x \), (3.4) shows that \( \gamma = 0 \) and hence that \( \alpha = 0 \). Suppose then that \( T(x) \) has arbitrarily large zeros. Let \( a \) and \( b \) be successive zeros of \( T(x) \) for \( x > \xi \).

**Case 1.** \( b-a \leq 2\beta/k \). In this case it follows from (3.3) that if the graph of \( | T(x) | \) rises as rapidly as possible going right from \( x = a \) and left from \( x = b \), it cannot lie above a triangle with altitude \( k(b-a)/2 \leq \beta \), so that
\[
\int_a^b | T(x) | \, dx \leq \frac{1}{2}(b-a)\beta.
\] (3.7)

**Case 2.** \( 2\beta/k < b-a \leq 2M/\beta \). Reasoning much as in Case 1 for a distance \( \beta/k \) from each end point, and otherwise using (3.6),
\[
\int_a^b | T(x) | \, dx \leq \frac{\beta^2}{k} + \left( b-a - \frac{2\beta}{k} \right) \beta = (b-a)\beta \left( 1 - \frac{\beta}{k(b-a)} \right)
\]
\[
\leq (b-a)\beta \left( 1 - \frac{\beta^2}{2Mk} \right) < (b-a)\beta \left( 1 - \frac{\alpha^2}{2Mk} \right).
\] (3.8)

**Case 3.** \( b-a > 2M/\beta \). By (3.4), since \( T(x) \neq 0 \) for \( a < x < b \),
\[
\int_a^b | T(x) | \, dx \leq M \leq \frac{1}{2}(b-a)\beta.
\] (3.9)

Since \( Mk \geq 1 \) and \( \alpha \leq 1 \), (3.8) is also valid where (3.7) and (3.9) hold. If \( x_{\xi} \) is the first zero of \( T(x) \) to the right of \( \xi \) and \( \tilde{x} \) the largest zero to the left of \( x_{\xi} \), (3.8) and (3.4) imply
\[
\frac{1}{x} \int_0^x | T(u) | \, du \leq \frac{1}{x} \left( \int_0^{x_{\xi}} | T(u) | \, du + (\tilde{x} - x_{\xi})\beta \left( 1 - \frac{\alpha^2}{2Mk} \right) + M \right).
\]
Hence letting $x \to \infty$ above and using $\gamma = a$,
\[
\alpha \leq \beta \left( 1 - \frac{\alpha^2}{2Mk} \right).
\]
Since this holds for every $\beta > \alpha$, it holds for $\beta = \alpha$ and so $\alpha^3 \leq 0$. Thus $\alpha = 0$ and so $S(x)/x \to 0$ as $x \to \infty$. Hence, given a small $\varepsilon > 0$, if $x$ is large enough,
\[
|S(x)| \leq \frac{1}{2}\varepsilon^2 x,
\]
so that
\[
S(x(1+\varepsilon)) - S(x) \leq \frac{1}{2}\varepsilon^2 (x+x(1+\varepsilon)) = \varepsilon^2 x,
\]
or,
\[
\int_x^{x(1+\varepsilon)} \frac{R(u)}{u} \, du \leq \varepsilon^2 x.
\]
Since $R(u) = \psi(u) - u$ and $\psi$ is non-decreasing,
\[
\frac{\psi(x)}{x(1+\varepsilon)} \int_x^{x(1+\varepsilon)} \frac{R(u)}{u} \, du \leq \int_x^{x(1+\varepsilon)} \frac{R(u)}{u} \, du + \varepsilon x \leq (\varepsilon + \varepsilon^2)x,
\]
so that, for large enough $x$,
\[
\psi(x)/x \leq (1+\varepsilon)^2.
\]
Similarly $S(x) - S(x(1-\varepsilon)) \geq -\varepsilon^2 x$ for large enough $x$ leads to
\[
\psi(x)/x \geq (1-\varepsilon)^2,
\]
which proves that $\psi(x)/x \to 1$ as $x \to \infty$.

REFERENCES


Massachusetts Institute of Technology Cambridge, Massachusetts