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

$$\psi(x) = \sum_{n \leqq x} \Lambda(n) \quad (\psi(x) = 0,\ x < 2).$$

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

$$\lim_{x \to \infty} \psi(x)/x = 1. \tag{1.1}$$

The Selberg inequality [**3**, p. 359] is

$$\psi(x) \log x + \Sigma \Lambda(n) \psi\left(\frac{x}{n}\right) = 2x \log x + O(x) \tag{1.2}$$

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

$$0 \leqq \lim_{x \to \infty} \sup \psi(x)/x \leqq 2. \tag{1.3}$$

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

$$\sum_{n \leqq x} c_n f(n) = C(x) f(x) - \int_1^x C(t) f'(t) dt;\ C(t) = \sum_{n \leqq t} c_n. \tag{1.4}$$

† 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

$$\sum_{n \leqq x} \Lambda(n) \log n = \psi(x) \log x + O(x). \tag{1.5}$$

Also

$$\sum_{j \leqq x} \Lambda(j)\psi\left(\frac{x}{j}\right) = \sum_{j \leqq x} \Lambda(j) \sum_{k \leqq x/j} \Lambda(k) = \sum_{jk \leqq x} \Lambda(j)\Lambda(k). \tag{1.6}$$

Thus, if

$$\Lambda_2(n) = \Lambda(n) \log n + \sum_{jk = n} \Lambda(j)\Lambda(k), \tag{1.7}$$

then (1.5) and (1.6) in (1.2) yield

$$\sum_{n \leqq x} \Lambda_2(n) = 2x \log x + O(x)$$

as an equivalent to (1.2). Using (1.4) with $c_n = 1$

$$\sum_{n \leqq x} \log n = x \log x + O(x). \tag{1.8}$$

Combining the above two inequalities

$$Q(n) = \sum_{k \leqq n} (\Lambda_2(k) - 2 \log k) = O(n). \tag{1.9}$$

The basic property of $\Lambda(n)$ is $\sum_{d \mid n} \Lambda(d) = \log n$. Summing this for $n \leqq x$ and using (1.8) yields the well known

$$\sum_{n \leqq x} \frac{\Lambda(n)}{n} = \log x + O(1), \tag{1.10}$$

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 $R(x) = \psi(x) - x$, $x \geqq 2$; $R(x) = 0$, $x < 2$, then, from (1.2) and (1.10),

$$R(x) \log x + \Sigma\Lambda(n)R\left(\frac{x}{n}\right) = O(x). \tag{2.1}$$

From (1.3)

$$\limsup_{x \to \infty} |R(x)|/x \leqq 1 \tag{2.2}$$

so that there exists $c < \infty$ such that

$$|R(x)| \leqq c\,|x|. \tag{2.3}$$

Let $S(y) = 0$, $y \leqq 2$, and for $y > 2$ let

$$S(y) = \int_2^y \frac{R(x)}{x}\, dx. \tag{2.4}$$

Then except at $y = p^j$, where $R(y)$ is discontinuous, $S'(y) = R(y)/y$ and so by (2.3),

$$|S'(y)| \leqq c, \quad y \neq p^j. \tag{2.5}$$

https://doi.org/10.1017/S0013091500011470 Published online by Cambridge University Press

Hence first, for y_1 and y_2 in an interval not containing any p^j in its interior,

$$| S(y_2)-S(y_1)| \leqq c \mid y_2-y_1 \mid \tag{2.6}$$

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

$$\| S(y_2)|-|S(y_1)\| \leqq c \mid y_2-y_1 \mid. \tag{2.7}$$

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

$$S(y) \log y + \Sigma\Lambda(j) S\left(\frac{y}{j}\right) = O(y). \tag{2.8}$$

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

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

Using $\log y/k = \log y - \log k$ and replacing this k by m,

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

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

$$\log^2 y \mid S(y)| \leqq \Sigma \left| S\left(\frac{y}{m}\right)\right| \Lambda_2(m) + K_1 y \log y \tag{2.9}$$

for some constant K_1.

Consider the identity

$$\Sigma \left| S\left(\frac{y}{m}\right)\right| \Lambda_2(m) = 2 \sum_{m \leqq y} \left| S\left(\frac{y}{m}\right)\right| \log m + J(y), \tag{2.10}$$

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

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

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

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

$$J(y) \leqq K_2 \sum_{2 \leqq m \leqq y} m \left(\frac{y}{m} - \frac{y}{m+1}\right) \leqq K_2 y \log y. \tag{2.11}$$

https://doi.org/10.1017/S0013091500011470 Published online by Cambridge University Press

Returning to the terms in the sum in (2.10)

$$\log m\left|S\left(\frac{y}{m}\right)\right| \leqq \int_m^{m+1} \log u\left|S\left(\frac{y}{u}\right)\right| du+J_m,$$

$$J_m=\int_m^{m+1} \log u\left(\left|S\left(\frac{y}{m}\right)\right|-\left|S\left(\frac{y}{u}\right)\right|\right) du.$$

By (2.7)

$$J_m \leqq c\left(\frac{y}{m}-\frac{y}{m+1}\right)\int_m^{m+1} \log u\,du \leqq \frac{cy\log(m+1)}{m(m+1)},$$

Using $\log(m+1) \leqq m$,

$$\log m\left|S\left(\frac{y}{m}\right)\right| \leqq \int_m^{m+1} \log u\left|S\left(\frac{y}{u}\right)\right| du+\frac{cy}{m+1}. \tag{2.12}$$

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

$$\log^2 y\,|S(y)| \leqq 2\int_2^y \log u\left|S\left(\frac{y}{u}\right)\right| du+K_3 y\log y. \tag{2.13}$$

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

$$\int_2^x \frac{R(u)}{u^2}du=O(1).$$

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

$$\int_2^y \frac{S(t)}{t^2}dt=O(1). \tag{3.1}$$

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

$$\alpha=\limsup_{x\to\infty}|T(x)| \leqq 1. \tag{3.2}$$

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

$$|T'(x)| \leqq k, \quad x \neq j\log p,$$

and by the argument following (2.5) this implies

$$\|T(x_2)|-|T(x_1)\| \leqq k\,|x_2-x_1|. \tag{3.3}$$

By (3.1)

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

which shows there is a constant M such that

$$\left|\int_{x_1}^{x_2} T(u)du\right| \leqq M, \tag{3.4}$$

and one can assume M so large that $Mk \geqq 1$.

https://doi.org/10.1017/S0013091500011470 Published online by Cambridge University Press

The key inequality (2.13) with $x = \log y$, $s = \log y/u$ (noting that $T(x) = 0$ for $x < \log 2$) becomes

$$|T(x)| \leqq \frac{2}{x^2}\int_0^x (x-s)|T(s)|\,ds + \frac{K_3}{x}$$

or, equivalently,

$$|T(x)| \leqq \frac{2}{x^2}\int_0^x v\,dv\left(\frac{1}{v}\int_0^v |T(s)|\,ds\right) + \frac{K_3}{x}. \tag{3.5}$$

From the definition of α in (3.2),

$$\gamma = \limsup_{x\to\infty} \frac{1}{x}\int_0^x |T(s)|\,ds \leqq \alpha.$$

But (3.5) shows that $\alpha \leqq \gamma$ 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 x_β such that

$$|T(x)| \leqq \beta \text{ for } x \geqq x_\beta. \tag{3.6}$$

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 $x > x_\beta$.

Case 1. $b-a \leqq 2\beta/k$. 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 $k(b-a)/2 \leqq \beta$, so that

$$\int_a^b |T(x)|\,dx \leqq \tfrac{1}{2}(b-a)\beta. \tag{3.7}$$

Case 2. $2\beta/k < b-a \leqq 2M/\beta$. Reasoning much as in Case 1 for a distance β/k from each end point, and otherwise using (3.6),

$$\int_a^b |T(x)|\,dx \leqq \frac{\beta^2}{k} + \left(b-a-\frac{2\beta}{k}\right)\beta = (b-a)\beta\left(1-\frac{\beta}{k(b-a)}\right)$$
$$\leqq (b-a)\beta\left(1-\frac{\beta^2}{2Mk}\right) < (b-a)\beta\left(1-\frac{\alpha^2}{2Mk}\right). \tag{3.8}$$

Case 3. $b-a > 2M/\beta$. By (3.4), since $T(x) \neq 0$ for $a < x < b$,

$$\int_a^b |T(x)|\,dx \leqq M \leqq \tfrac{1}{2}(b-a)\beta. \tag{3.9}$$

Since $Mk \geqq 1$ and $\alpha \leqq 1$, (3.8) is also valid where (3.7) and (3.9) hold. If x_1 is the first zero of $T(x)$ to the right of x_β and $\tilde{x}$ the largest zero to the left of x, (3.8) and (3.4) imply

$$\frac{1}{x}\int_0^x |T(u)|\,du \leqq \frac{1}{x}\left(\int_0^{x_1} |T(u)|\,du + (\tilde{x}-x_1)\beta\left(1-\frac{\alpha^2}{2Mk}\right) + M\right).$$

https://doi.org/10.1017/S0013091500011470 Published online by Cambridge University Press

Hence letting $x \to \infty$ above and using $\tilde{x} \leqq x$ and $\gamma = \alpha$,

$$\alpha \leqq \beta\left(1 - \frac{\alpha^2}{2Mk}\right).$$

Since this holds for every $\beta > \alpha$, it holds for $\beta = \alpha$ and so $\alpha^3 \leqq 0$. 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,

$$| S(x)| \leqq \tfrac{1}{3}\varepsilon^2 x,$$

so that

$$S(x(1+\varepsilon)) - S(x) \leqq \tfrac{1}{3}\varepsilon^2(x + x(1+\varepsilon)) < \varepsilon^2 x,$$

or,

$$\int_x^{x(1+\varepsilon)} \frac{R(u)}{u}\, du \leqq \varepsilon^2 x.$$

Since $R(u) = \psi(u) - u$ and ψ is non-decreasing,

$$\frac{\psi(x)}{x(1+\varepsilon)} \int_x^{x(1+\varepsilon)} du \leqq \int_x^{x(1+\varepsilon)} \frac{R(u)}{u}\, du + \varepsilon x \leqq (\varepsilon + \varepsilon^2)x,$$

so that, for large enough x,

$$\psi(x)/x \leqq (1+\varepsilon)^2.$$

Similarly $S(x) - S(x(1-\varepsilon)) \geqq -\varepsilon^2 x$ for large enough x leads to

$$\psi(x)/x \geqq (1-\varepsilon)^2,$$

which proves that $\psi(x)/x \to 1$ as $x \to \infty$.

REFERENCES

(1) R. Breusch, An elementary proof of the prime number theorem with remainder term, *Pacific J. Math.* **10** (1960), 487-497.

(2) P. Erdös, On a new method in elementary number theory which leads to an elementary proof of the prime number theorem, *Proc. Nat. Acad. Sci. U.S.A.* **35** (1949), 374-384.

(3) G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford, 1960).

(4) V. Nevanlinna, Über den elementaren Beweis des Primzahlsatzes, *Soc. Sci. Fenn. Comment. Phys.-Math.* **27** (1962), No. 3, 8 pp.

(5) A. Selberg, An elementary proof of the prime number theorem, *Ann. of Math.* (2), **50** (1949), 305-313.

(6) H. N. Shapiro, On a theorem of Selberg and generalizations, *Ann. of Math.* (2), **51** (1950), 485-497.

(7) T. Tatuzawa and K. Iseki, On Selberg's elementary proof of the prime-number theorem, *Proc. Japan Acad.* **27** (1951), 340-342.

(8) E. M. Wright, The elementary proof of the prime number theorem, *Proc. Roy. Soc. Edinburgh* (A), **63** (1951), 257-267.

Massachusetts Institute of Technology
Cambridge, Massachusetts