BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 7, Number 3, November 1982

ELEMENTARY METHODS IN THE STUDY OF THE DISTRIBUTION OF PRIME NUMBERS[1]

BY HAROLD G. DIAMOND[2]

Table of Contents

1. Historical survey. The question of how primes are distributed among all natural numbers has long been a source of fascination. The oldest known result in this area is the classical theorem that there are an infinite number of primes. The proof of this theorem recorded by Euclid [**Euc**] has stood as one of the most beautiful in mathematics. No more detailed information about the distribution of primes was known in ancient times, for their apparently random occurrence frustrated all attempts at obtaining simple precise formulas.

We take up our account at the end of the 18th century, when the "right question" about the distribution of primes was asked and a conjectured answer offered by A. M. Legendre [**Leg**] and by C. F. Gauss [**Gau**]. They recast the prime distribution question in a statistical form: About how many of the first N positive integers are primes? Examination of tables of prime numbers led them to conjecture that the answer was, in some sense, $N/\log N$.

If we let $\pi(x)$ denote the number of primes in the interval $[1, x]$, the assertion of Gauss and Legendre can be expressed concisely as the celebrated Prime Number Theorem (P.N.T.).

Received by the editors January 10, 1982.

1980 *Mathematics Subject Classification*. Primary 10H15, 10A25.

[1]This article is based on an A.M.S. One-Hour Address given at the University of Notre Dame in March 1981.

[2]Research supported in part by a grant from the National Science Foundation.

0273-0979/82/0000-0351/$10.25

THEOREM 1.1.

$$\pi(x) \sim x/\log x \qquad (x \to \infty). \tag{1.1}$$

That is, the ratio $\pi(x)$: $x/\log x$ converges to 1 as x grows without bound.

This theorem was first proved in 1896, about a century after the observations of Gauss and Legendre. Before coming to the proof of the P.N.T. we shall discuss some significant contributions to prime number theory made earlier in the nineteenth century.

The first person to establish estimates of the true order of magnitude of $\pi(x)$ was P. L. Chebyshev. In the mid nineteenth century, he proved [**Chb**] that

$$.92 < \pi(x)/(x/\log x) < 1.11 \tag{1.2}$$

for all sufficiently large numbers x. Chebyshev's argument was based on an identity he discovered and a clever real variable method of mimicking the Möbius μ function. We shall describe Chebyshev's argument in §3.

Within a decade of the work of Chebyshev, there appeared the extraordinary memoir of G. F. B. Riemann [**Rie**] on $\pi(x)$. Riemann proposed to attack the problem indirectly by studying properties of a certain analytic function in the complex domain. This function, which is today called the Riemann zeta function, is defined for $\operatorname{Re} s > 1$ by the formula

$$\zeta(s) = 1 + 2^{-s} + 3^{-s} + 4^{-s} + \cdots. \tag{1.3}$$

The connection with prime numbers is revealed by a second representation of $\zeta(s)$, valid in the same region,

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}. \tag{1.4}$$

Here p runs through the prime numbers. The product formula is the analytic expression of the unique factorization theorem. We can formally verify (1.4) by expanding each factor $1/(1 - p^{-s})$ as a geometric series and multiplying together the factors.

We remark that the zeta function appears first to have been used by L. Euler [**Ayo 2**], over a century earlier, but only in the real domain. Euler used the two representations of zeta to show that the sum of the reciprocals of the primes diverges and this in turn gives another proof of the infinitude of primes.

The approach to the prime number problem proposed by Riemann, using a function of a complex variable generated by arithmetic data, came to be called *analytic*. On the other hand, direct real variable treatment of arithmetic data, such as the method of Chebyshev, came to be called *elementary*. We shall follow this usage here. To avoid confusion, we shall (with apologies to Sherlock Holmes) use the word *simple* for "easy to understand". It will be seen that some elementary arguments are far from simple.

We mention here some other contributions to prime number theory from the period after Chebyshev. F. Mertens [**Mer**] established the handsome relations

$$\sum_{p \leq x} \frac{\log p}{p} - \log x = O(1) \tag{1.5}$$

and

$$\prod_{p \leqslant x} (1 - 1/p) \sim e^{-\gamma}/\log x. \tag{1.6}$$

In both of these expressions p runs through prime numbers. The number γ is Euler's constant ($\gamma = .577215\ldots$).

The expression $f(x) = O(g(x))$ is to be read as "$|f(x)| \leqslant Bg(x)$ for some positive constant B and all values of $x \geqslant 1$". Also, we shall write $f(x) = o(g(x))$ if $f(x)/g(x) \to 0$ as $x \to \infty$. This notation was introduced by P. Bachmann, cf. [**Lnd 1**, p. 883], and popularized by E. Landau, who will enter our account shortly.

E. Meissel [**Mei, Mat**, pp. 273–278] calculated $\pi(x)$ for various large values of x by a method inspired by the inclusion-exclusion approach of Legendre [**Leg**, pp. 12–15]. Meissel's identity was to have subsequent applications in sieve theory [**Buc 1, 2, HaRi**].

Various refinements were made upon the method of Chebyshev. The best known work of this sort was that of J. J. Sylvester [**Syl 1, 2**], who showed that

$$.956 < \pi(x)/(x/\log x) < 1.045 \tag{1.7}$$

holds for all sufficiently large x. The approach of Sylvester was *ad hoc* and computationally complex; it offered no hope of leading to a proof of the P.N.T. Indeed, Sylvester concluded his article [**Syl 1**] with the lament that "...we shall probably have to wait [for a proof of the P.N.T.] until someone is born into the world as far surpassing Tchebycheff in insight and penetration as Tchebycheff has proved himself superior in these qualities to the ordinary run of mankind."

A decade after Sylvester wrote these words, J. Hadamard embarked upon his prime number studies along the analytic lines sketched by Riemann. This research culminated in the successful proof of the P.N.T. by Hadamard [**Had**] and C. de la Vallée Poussin [**laV 1**], independently, at the end of the nineteenth century. The period of the next fifty years—roughly covering the career of G. H. Hardy—was the Golden Age of analytic methods. Other triumphs of this time include the oscillation result for $\pi(x)$ of J. E. Littlewood [**Lit, Ing 2, Leh, Dia 1**] and the error terms in the P.N.T. of de la Vallée Poussin [**laV 2**], Littlewood [**Lnd 3**. vol. 2, pp. 31–47], and the school of I. M. Vinogradov [**Wal, Tit, Cha 2, ViI**].

The development of prime number theory during this time was influenced by a number of factors. One, of course, was the great success of Hadamard, H. von Mangoldt [**Man 1, 2**], and others in carrying out much of Riemann's program. Another important influence was the *Handbuch* [**Lnd 1**] of Landau, which led to the wide dissemination of analytic ideas.

The general Tauberian theory of N. Wiener [**Wie 1, 2**] showed an "equivalence" between the P.N.T. and the nonvanishing of the Riemann zeta function on the line $\operatorname{Re} s = 1$ in the complex plane (cf. §§4, 7). The role of analytic methods was actively promoted by influential mathematicians such as Hardy, who frequently opined that a nonanalytic proof of the P.N.T. was unlikely to be found [**Har, Boh, Ing 1**].

The accomplishments of elementary methods during this period were relatively meager. This reinforced the belief that they did not have the power to achieve deep results. Elementary methods became unfashionable in number theory.

Although in eclipse, elementary methods did retain a certain appeal in the eyes of some. One reason for this was that they, unlike analytic methods, do not require the introduction of ideas so remote from the arithmetic questions under consideration. A more widely acceptable rationale—success where analytic methods failed—emerged from the work of V. Brun done around 1920. Brun showed by an elementary sieve argument [**Bru 1**, **HaRi**, **HaRo**] that the twin primes (numbers p and $p + 2$ both of which are prime) are so sparse that the sum of their reciprocals converges, in contrast to the sum of the reciprocals of all primes. This was followed by other sieve results, of which we briefly note the Brun-Titchmarsh inequality, Schnirelmann's representation of integers as a sum of a bounded number of primes, and the sieve methods of Selberg and Rosser. A third reason for pursuing an elementary approach to the P.N.T. was the hope that such a result would shed new light on the structure of the primes.

In 1948 A. Selberg [**Sel 1**] discovered the formula

$$\sum_{p \leq x} \log^2 p + \sum_{pq \leq x} \log p \log q = 2x \log x + O(x), \tag{1.8}$$

which is a kind of "weighted" version of the Chebyshev relation which will be given at the end of §2. (Here p and q run through primes.) Selberg's formula at once raised hope and despair. On the positive side, it immediately implied the nontrivial fact that relatively short intervals could not have too great a concentration of primes (cf. (5.5)). On the other hand the formula contains in a balanced way both primes and numbers expressible as a product of two primes, and it was not obvious how one would pry these quantities apart.

A century had elapsed since the work of Chebyshev, a period of numerous unsuccessful elementary attempts at proving the P.N.T. Clearly, some daring was required to invest much hope and energy in mounting another elementary assault on the P.N.T. What was needed in the present case was a kind of nonlinear Tauberian theorem to wrest the primes apart from the numbers expressible as a product of two primes in Selberg's formula.

P. Erdös noted that Selberg's formula also yields a lower bound for the number of primes in certain intervals. With this observation Selberg found an elementary Tauberian argument that gave the P.N.T. Subsequently, Selberg and Erdös each discovered other, more direct Tauberian arguments [**Sel 1**, **Erd 1**].

The elementary proof of the P.N.T. was a sensation. It helped earn Selberg a Fields Medal [**Boh**] and Erdös a Cole Prize [**Coh**]. Their articles were given a detailed and incisive review by A. E. Ingham [**Ing 3**], who hailed the proof as "a discovery of the first importance...." There soon appeared variant proofs of Selberg's formula, e.g. [**Tl**, **Sha 1**, **Pop**], and of the Tauberian argument, e.g. [**Crp 1**, **Wri**, **Bre 1**]. Generalizations were made to primes in arithmetic progressions [**Sel 2**, **Sha 2**], algebraic number fields [**Ayo 1**], and abstract number systems [**FS**, **Ami 1, 2**, **Seg**]. There appeared versions in which *all* analysis was purged [**Eda, Fog**]. And, most telling, the theorem soon took its place alongside the aforementioned theorem of Euclid in many books on number theory, e.g. [**Gio, Cha 2, HW, Nag, Sch**].

It was now natural to ask, How good an error term in the P.N.T. could be produced by elementary methods? As an index for comparison, we note that de la Vallée Poussin [**laV 2**] had shown by analytic means in 1899 that

$$\pi(x) - \lim_{\varepsilon \to 0+} \left\{ \int_0^{1-\varepsilon} + \int_{1+\varepsilon}^{x} \right\} \frac{dt}{\log t} = O\{x \exp(-c \log^{\alpha} x)\} \tag{1.9}$$

for $\alpha = 1/2$ and some positive constant c.

The integral in (1.9), which is usually denoted by $\operatorname{li} x$, can be shown by m integrations by parts to equal

$$\frac{x}{\log x} + \frac{1!x}{\log^2 x} + \cdots + \frac{(m-1)!x}{\log^m x} + O_m\left(\frac{x}{\log^{m+1} x}\right)$$

for any positive integer m. (The subscript on the O symbol is to remind us that the implied constant depends on m.) Also, it follows from (1.9) that

$$\pi(x) - \operatorname{li} x = O_m(x/\log^m x) \tag{1.10}$$

holds for any positive number m.

The first elementary P.N.T. error estimates [**Crp 2, Kuh, Bre 2, Wir 1**] had the form

$$\pi(x) - x/\log x = O(x/\log^c x)$$

with c a constant between 1 and 2. Error terms of the type (1.10) for arbitrarily large values of m were achieved by E. Bombieri [**Bom 1**] and E. Wirsing [**Wir 2**] in the early 1960's.

A few years earlier A. O. Gelfond [**Gel 1**] had obtained a de la Vallée Poussin type error estimate (1.9) with $\alpha = 1/9$ by a method which was partially elementary. An elementary proof of (1.9) with a positive α was announced by W. B. Jurkat at the 1962 International Congress of Mathematicians [**Jur**]. In 1970 the author and J. Steinig gave an elementary proof of (1.9) with $\alpha = 1/7 - \varepsilon$ [**DS**]. We shall describe this work in §8. The current record holders for an elementary P.N.T. error term are A. F. Lavrik and A. Š. Sobirov [**LS**] and A. V. Sokolovski [**Sok**], who sharpened a lemma in the Diamond-Steinig article to obtain (1.9) with $\alpha = 1/6 - \varepsilon$.

2. Arithmetic functions and convolutions. Our subsequent work will be expressed in terms of arithmetic functions and a method of combining functions. We introduce these notions here.

An *arithmetic function* is a map from **N**, the positive integers, to **R**, the reals, or **C**, the complex numbers. We set $1(n) \equiv 1$, $L(n) = \log n$, and for r a positive integer we set

$$e_r(n) = \begin{cases} 1, & \text{if } n = r, \\ 0, & \text{if } n \neq r. \end{cases}$$

We define the Möbius function by

$$\mu(n) = \begin{cases} (-1)^k, & \text{if } n \text{ is a product of } k \text{ distinct primes,} \\ 0, & \text{if } p^2 \mid n \text{ for some prime } p, \end{cases}$$

and von Mangoldt's function by

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^\alpha, p \text{ prime and } \alpha \in \mathbf{N}, \\ 0, & \text{otherwise.} \end{cases}$$

Associated with each arithmetic function f is a summatory function F defined on $\mathbf{R}$ by

$$F(x) = \sum_{n \leqslant x} f(n).$$

The following summatory functions will be of interest to us:

$$[x] = \sum_{n \leqslant x} 1, \qquad M(x) = \sum_{n \leqslant x} \mu(n),$$
$$\Psi(x) = \sum_{n \leqslant x} \Lambda(n) = \sum_{p^\alpha \leqslant x} \log p,$$

and of course $\pi(x)$, the summatory function of the arithmetic function 1_P, which is 1 at primes and 0 elsewhere.

Many relations in prime number theory are best expressed in terms of *multiplicative convolution*. This is an operation which combines two arithmetic functions f and g to form a new one, called $f*g$, according to the rule

$$f*g(n) = \sum_{ij=n} f(i)g(j) = \sum_{i \mid n} f(i)g(n/i).$$

In the middle expression the sum extends over all ordered pairs of positive integers whose product equals n; in the last the sum extends over all positive integers which divide n. Some convolution relations that we shall use are

(2.1) $f*e_1 = f$, for any arithmetic function f,

(2.2) $1*\mu = e_1$,

(2.3) $\Lambda*1 = L$.

The first formula follows immediately from the definitions of $*$ and e_1, and asserts that e_1 is the unity element of convolution. The second identity is called the Möbius inversion formula. It is proved by first noting that $1*\mu(1) = 1 = e_1(1)$. Then for $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} > 1$ one has

$$\sum_{d \mid n} \mu(d) = \sum_{d \mid p_1 p_2 \cdots p_r} \mu(d),$$

since $\mu(d) = 0$ if d contains any square factors, and the last expression is an expanded form of $(1-1)^r = 0$. We verify (2.3) by noting that each side vanishes at $n = 1$, and for $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} > 1$ we have

$$\sum_{d \mid n} \Lambda(d) = \sum_{i=1}^{r} \sum_{\beta=1}^{\alpha_i} \log p = \sum_{i=1}^{r} \alpha_i \log p_i = \log n,$$

since $\Lambda(d) = 0$ unless d is of the form p_i^β.

Convolution is an associative and commutative operation. A number of messy arithmetical calculations in number theory are in fact verifications of associativity.

The operation of pointwise multiplying an arithmetic function by L is a derivation. It follows immediately from the definition of convolution that

$$L(f*g) = (Lf)*g + f*(Lg) \tag{2.4}$$

holds for any arithmetic functions f and g.

The summatory function of a convolution $f*g$ is related to the summatory functions F and G of the factors by the following relations:

$$\sum_{k \leq x} f*g(k) = \sum_{n \leq x} F\left(\frac{x}{n}\right)g(n) = \sum_{m \leq x} G\left(\frac{x}{m}\right)f(m). \tag{2.5}$$

We show this by writing the definition of $*$ and reordering the resulting sum, which extends over lattice points in a hyperbolic region. We have

$$\sum_{k \leq x} \sum_{mn=k} f(m)g(n) = \sum_{mn \leq x} f(m)g(n) = \sum_{n \leq x} \sum_{m \leq x/n} f(m)g(n),$$

and similarly for the expression involving G.

Examples of (2.5) which will be of use to us are the formulas

$$\sum_{k \leq x} \Lambda*1(k) = \sum_{n \leq x} \Psi\left(\frac{x}{n}\right) = \sum_{m \leq x} \left[\frac{x}{m}\right]\Lambda(m).$$

If we recall (2.3) and note that

$$\left|\int_1^x \log t\,dt - \sum_{n \leq x} \log n\right| \leq \log x,$$

then we obtain

$$\sum_{n \leq x} \Psi\left(\frac{x}{n}\right) = x\log x - x + O(\log ex). \tag{2.6}$$

In some estimations it is useful to combine the two representations in (2.5). We then obtain the following formula.

LEMMA 2.1 (DIRICHLET HYPERBOLA METHOD). *Let $1 < y < x$. Then*

$$\sum_{n \leq x} f*g(n) = \sum_{n \leq y} F\left(\frac{x}{n}\right)g(n) + \sum_{m \leq x/y} G\left(\frac{x}{m}\right)f(m) - F\left(\frac{x}{y}\right)G(y).$$

PROOF. We have

$$\begin{aligned}\sum_{y<n \leq x} F\left(\frac{x}{n}\right)g(n) &= \sum_{y<n \leq x} \sum_{m \leq x/n} f(m)g(n)\\ &= \sum_{m \leq x/y} \sum_{y<n \leq x/m} g(n)f(m) = \sum_{m \leq x/y} \left(G\left(\frac{x}{m}\right) - G(y)\right)f(m). \quad \square\end{aligned}$$

The most familiar occurrence of this identity is Dirichlet's original application to estimate the summatory function of $d(n)$, the number of positive integer divisors of n [Dir]. We have $d(n) = 1*1(n)$, and the method with $y = \sqrt{x}$ and some familiar estimations yields

$$\sum_{n\leq x} d(n) = x\log x + (2\gamma - 1)x + O\left(\sqrt{x}\right). \tag{2.7}$$

3. Chebyshev's estimates. In this section we shall deduce the bounds (1.2) for $\pi(x)$ found by Chebyshev. We shall use the Ψ function instead of π, because (1) Ψ satisfies relation (2.6) while π does not occur in such a simple way and (2) Ψ is essentially a weighted version of π. We begin by showing how an estimate of $\Psi(x) - x$ yields an estimate of $\pi(x) - \operatorname{li} x$.

It is convenient to introduce here the function

$$\vartheta(x) = \sum_{p\leq x} \log p = \sum_{n\leq x} L1_P(n),$$

and first show that ϑ and Ψ are close. We have

$$\Psi(x) = \sum_{p^\alpha \leq x} \log p = \vartheta(x) + \vartheta(x^{1/2}) + \cdots + \vartheta(x^{1/k}),$$

where $k = [\log x/\log 2]$, since $x^{1/\alpha} < 2$ for $\alpha > \log x/\log 2$. Now $0 \leq \vartheta(x^{1/\alpha}) \leq x^{1/\alpha}\log x$ for $\alpha = 2, 3, \ldots$, so we have

$$\Psi(x) - \vartheta(x) = O(x^{1/2}\log x). \tag{3.1}$$

The ϑ function is a weighted form of π. We invert the relation by writing

$$\pi(x) = \sum_{p\leq x} 1 = \sum_{n\leq x} \frac{\vartheta(n) - \vartheta(n-1)}{\log n},$$

$$\pi(x) - \sum_{2\leq n\leq x} \frac{1}{\log n} = \sum_{2\leq n\leq x} \frac{\vartheta(n) - n - \{\vartheta(n-1) - (n-1)\}}{\log n},$$

and obtain by partial summation

$$\pi(x) - \sum_{2\leq n\leq x} \frac{1}{\log n} = \frac{\vartheta(x) - [x]}{\log[x+1]} + \frac{1}{\log 2} + \sum_{2\leq n\leq x} (\vartheta(n) - n)\left(\frac{1}{\log n} - \frac{1}{\log(n+1)}\right).$$

If we change from ϑ to Ψ and introduce li in place of $\sum 1/\log n$, we obtain

$$\pi(x) - \operatorname{li} x = \sum_{2\leq n\leq x} (\Psi(n) - n)\left(\frac{1}{\log n} - \frac{1}{\log(n+1)}\right) + \frac{\Psi(x) - [x]}{\log[x+1]} + O\left(\sqrt{x}\right). \tag{3.2}$$

LEMMA 3.1. *Suppose that* $|\Psi(x) - x| \leq F(x)$, *where F is a positive nondecreasing function on* $[1, \infty)$. *Then*

$$|\pi(x) - \operatorname{li} x| \leq 2F(x)/\log x + 2F\left(\sqrt{x}\right) + B\sqrt{x}$$

holds for all $x > 1$. Here B is a positive constant.

PROOF. We have from (3.2)

$$|\pi(x)-\operatorname{li}x| \leqslant F([x]) \sum_{\sqrt{x}<n\leqslant x}\left(\frac{1}{\log n}-\frac{1}{\log(n+1)}\right)$$
$$+F\left(\sqrt{x}\right)\sum_{2\leqslant n\leqslant\sqrt{x}}\left(\frac{1}{\log n}-\frac{1}{\log(n+1)}\right)$$
$$+\frac{F([x])}{\log[x+1]}+O\left(\sqrt{x}\right) \leqslant \frac{F([x])}{\log\sqrt{x}}+\frac{F\left(\sqrt{x}\right)}{\log 2}+B\sqrt{x}. \quad \square$$

We shall henceforth work to show that $\Psi(x)$ is close to x.

We now describe—in convolution language—Chebyshev's strategy [**Dia 2**]. We begin with an instructive false start. If we convolve both sides of (2.3) by μ and note that

$$\Lambda * 1 * \mu = \Lambda * e_1 = \Lambda,$$

we obtain the representation

$$\Lambda = L*\mu. \tag{3.3}$$

This identity can be used to obtain a bound $\Psi(x) \leqslant cx$, but with a comparatively large value of c. The trouble stems from the fact that we know little about μ at the outset.

We shall instead seek a manageable function ν which in some sense approximates μ. We form

$$\Lambda * 1 * \nu = L*\nu \tag{3.4}$$

and use estimates for

$$E(x) := \sum_{m\leqslant x} 1*\nu(m) = \sum_{n\leqslant x}\left[\frac{x}{n}\right]\nu(n) \tag{3.5}$$

to obtain bounds for Ψ.

As an example, consider $\nu = e_1 - 2e_2$. In this case $E(x) = [x] - 2[x/2]$, which equals 1 for $[x]$ odd and 0 for $[x]$ even. We have first

$$\sum_{m\leqslant x} L*\nu(m) = \sum_{n\leqslant x}\left\{\frac{x}{n}\log\frac{x}{n}-\frac{x}{n}+O\left(\log\frac{ex}{n}\right)\right\}\nu(n)$$
$$=(x\log x - x)\sum_{n\leqslant x}\frac{\nu(n)}{n} - x\sum_{n\leqslant x}\frac{\nu(n)\log n}{n}+O(\log ex)$$
$$= x\log 2 + O(\log ex).$$

On the other hand we have

$$\sum_{m\leqslant x}\Lambda*1*\nu(m) = \sum_{n\leqslant x}E\left(\frac{x}{n}\right)\Lambda(n)\begin{cases}\leqslant \displaystyle\sum_{n\leqslant x}\Lambda(n) = \Psi(x),\\ \geqslant \displaystyle\sum_{x/2<n\leqslant x}\Lambda(n) = \Psi(x)-\Psi\left(\frac{x}{2}\right).\end{cases}$$

These relations together imply that

$$\Psi(x) \geqslant x \log 2 + O(\log ex)$$

and

$$\Psi(x) - \Psi(x/2) \leqslant x \log 2 + O(\log ex).$$

In the last relation we replace x by $x/2, x/4, \dots, x/2^{k-1}$, where $k = [\log x/\log 2]$, and sum to obtain

$$\Psi(x) \leqslant 2x \log 2 + O(\log^2 ex).$$

Let us note some desirable properties for a function ν which aspires to be an "approximate convolution inverse of 1". First, it is necessary that $\sum \nu(n)/n = 0$ in order that we have

$$\sum_{m \leqslant x} L * \nu(m) = O(x).$$

(The Möbius function satisfies this condition, cf. §4.) Second, it is desirable that $\nu(n) \to 0$ as $n \to \infty$ so that we can carry out our estimations without excessive difficulty. (The Möbius function does not have this property!) In practice we shall ask that $\nu(n) = 0$ for all n's from some point onward. Third, in order that our estimates should be reasonably sharp, we ask that $E(x)$, defined by (3.5), should be near 1 for a reasonably long interval $[1, a)$ and never be too large or too negative. This insures that

$$\sum_{n \leqslant x} E(x/n)\Lambda(n) \approx \Psi(x).$$

(If $\nu = \mu$, the Möbius function, then $E(x) = 1$ for all $x \geqslant 1$.)

The choice made by Chebyshev was

$$\nu = e_1 - e_2 - e_3 - e_5 + e_{30}.$$

This function satisfies

$$\sum \nu(n)/n = 1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{5} + \frac{1}{30} = 0,$$

and $\nu(n) = 0$ for all $n > 30$. We have to check the values of $E(x)$. This has to be done only for $0 \leqslant x < 30$, since E has period 30, as one verifies immediately.

The calculations are made even simpler by the observation that for x nonintegral we have

$$\begin{aligned} E(x) + E(30 - x) &= E(x) + E(-x) \\ &= ([x] + [-x]) - ([x/2] + [-x/2]) - ([x/3] + [-x/3]) \\ &\quad - ([x/5] + [-x/5]) + ([x/30] + [-x/30]) = 1, \end{aligned}$$

since $[y] + [-y] = -1$ for any nonintegral y.

Now direct verification shows that

$$E(x) = \begin{cases} 1, & \text{for } 1 \leqslant x < 6, 7 \leqslant x < 10, 11 \leqslant x < 12, 13 \leqslant x < 15, \\ 0, & \text{for } 0 \leqslant x < 1, 6 \leqslant x < 7, 10 \leqslant x < 11, 12 \leqslant x < 13. \end{cases}$$

The relation $E(x) = 1 - E(30 - x)$ now tells us that E assumes only the values 0 and 1 for all nonintegral arguments between 15 and 30. Since E is continuous from the right, $E(x) = 0$ or 1 also at integers.

Arguing exactly as we did for $\nu = e_1 - 2e_2$, we find in the present case that

$$\sum_{m \leqslant x} L*\nu(m) = -\sum_{n \leqslant x} (\log n)\nu(n)/n + O(\log ex) = Ax + O(\log ex),$$

where

$$A = \frac{\log 2}{2} + \frac{\log 3}{3} + \frac{\log 5}{5} - \frac{\log 30}{30} = .92129\ldots,$$

and

$$\sum_{n \leqslant x} E\left(\frac{x}{n}\right)\Lambda(n) \begin{cases} \leqslant \sum_{n \leqslant x} \Lambda(n) = \Psi(x), \\ \geqslant \sum_{x/6 < n \leqslant x} \Lambda(n) = \Psi(x) - \Psi(x/6). \end{cases}$$

This yields the bounds

$$\Psi(x) \geqslant Ax + O(\log ex), \qquad \Psi(x) \leqslant (6/5)Ax + O(\log^2 ex).$$

These estimates and formula (3.2) give the estimates (1.2) for $\pi(x)$.

The refinements of Sylvester **[Syl 1, 2]** were made by introducing much more complicated functions ν. We shall consider improved Chebyshev type arguments in §9.

Chebyshev **[Chb]** used the estimates (1.2) to prove the proposition that was then, and even now is, called *Bertrand's Postulate* **[Ber]**. This is the assertion that if $x \geqslant 1$, then each interval $(x, 2x]$ contains at least one prime. Bertrand's postulate holds for all sufficiently large numbers x, since

$$\pi(2x) > (.92)\frac{2x}{\log 2x} > \frac{1.11x}{\log x} > \pi(x).$$

It can be established for small values of x by inspection. This type of argument and Chebyshev's estimates show that any interval $(x, Ax]$ contains primes provided that $A > 6/5$ and x is sufficiently large. The P.N.T. guarantees that each interval $(x, Ax]$ contains a prime provided that $A > 1$ and x is greater than some X depending on A.

4. Theorems "equivalent" to the P.N.T. Several asymptotic formulas involving prime numbers have "comparable depth" with the P.N.T., i.e. their truth can be deduced from the P.N.T. by elementary and rather simple methods, and conversely the P.N.T. can be deduced from any of these formulas by such methods. An example is the relation $\Psi(x) \sim x$, which we have shown implies the P.N.T. and is also a consequence of the P.N.T., as we shall indicate.

These relations are all given in Landau's *Handbuch* **[Lnd 1]**. Shortly after the *Handbuch's* appearance, A. Axer **[Axe]** and Landau **[Lnd 2]** showed how to deduce the various propositions from one another by elementary arguments and these results came to be described as "equivalent" to the P.N.T. The notion of equivalence never had a clear logical basis, and of course it was rendered void by the discovery of an elementary proof of the P.N.T. Nevertheless, it survives today as a convenient way of grouping together a number of statements about primes.

The following asymptotic formulas are equivalent to the P.N.T. in the foregoing sense. There exist in the literature direct elementary and analytic proofs of all these relations except (to the knowledge of this writer) the last.

(4.1) $\Psi(x) \sim x$,

(4.2) $\Sigma_{n\leq x} \Lambda(n)/n = \log x + c + o(1)$,

(4.3) $\Sigma_{p\leq x} \log p/p = \log x + c' + o(1)$,

(4.4) $M(x) = \Sigma_{n\leq x} \mu(n) = o(x)$,

(4.5) $\Sigma_{n\leq x} \mu(n)/n = o(1)$,

(4.6) $p_n \sim n \log n$, p_n the nth prime.

We remark that replacement of the o by an O in any of (4.2) – (4.5) results in a relation which is easily established. For example, the Mertens relation (1.5) is the O-analogue of (4.3), and the O-analogue of (4.4) is trivial.

We shall indicate how several of these relations can be deduced from the P.N.T. The remaining cases and converse implications can be supplied by the interested reader by analogous methods.

P.N.T. $\Rightarrow$ (4.1). We have

$$\Psi(x) \sim \vartheta(x) = \sum_{n\leq x} \log n(\pi(n) - \pi(n-1))$$
$$= \pi(x)\log[x] - \sum_{n\leq x-1} \log\left(\frac{n+1}{n}\right)\pi(n) \sim x,$$

since the last sum is at most

$$c \sum_{2\leq n\leq x} \frac{1}{n}\frac{n}{\log n} = O(x/\log x). \quad \square$$

(4.1) $\Rightarrow$ (4.4). We shall use a variant form of (2.3) which we now give. From (2.2) and (2.4) we obtain

$$0 = Le_1 = L(1*\mu) = L1*\mu + 1*L\mu.$$

We replace $L1*\mu$ by Λ (by (3.3)) and convolve both sides of the resulting equation by μ obtaining

$$L\mu = -\Lambda*\mu. \tag{4.7}$$

If we add the identity $0 = 1*\mu - e_1$ to (4.7), we find for any $x \geq 1$ that

$$\sum_{n\leq x} \mu(n)\log n = \sum_{n\leq x} (1-\Lambda)*\mu(n) - 1. \tag{4.8}$$

Summation by parts brings the left side of (4.8) to the form

$$M(x)\log[x] - \sum_{n\leq x-1} M(n)\log(1 + 1/m) = M(x)\log x + O(x).$$

We write the sum on the right side of (4.8) as

$$\sum_{n\leq x}\left(\left[\frac{x}{n}\right] - \Psi\left(\frac{x}{n}\right)\right)\mu(n).$$

We estimate this sum with the aid of (4.1). For any $\varepsilon > 0$ we have

$$|[x/n] - \Psi(x/n)| < \varepsilon x/n,$$

provided that x/n exceeds some $A = A(\varepsilon)$. Also, Chebyshev's estimate for Ψ insures that

$$|[x/n] - \Psi(x/n)| < Bx/n$$

holds for some B and all values of x/n. Thus

$$\begin{aligned}\left|\sum_{n \leq x}\left(\left[\frac{x}{n}\right] - \Psi\left(\frac{x}{n}\right)\right)\mu(n)\right| &\leq \sum_{n \leq x/A} \varepsilon\frac{x}{n} + \sum_{x/A < n \leq x} B\frac{x}{n} \\ &\leq \varepsilon x(1 + \log(x/A)) + Bx(A/x + \log A) \\ &\leq 2\varepsilon x \log x,\end{aligned}$$

provided that x is sufficiently large.

Combining the estimates of the two sides of (4.8), we find that $|M(x)|\log x \leq 3\varepsilon x \log x$ for all sufficiently large x. Thus (4.4) holds. □

(4.4) ⇒ (4.5). We shall show that

$$\sum_{n \leq x} \mu(n)\frac{x}{n} = o(x). \tag{4.9}$$

Let $S(x)$ denote the left side of (4.9). For $x \geq 1$ we have

$$S(x) = \sum_{n \leq x} \mu(n)\left(\frac{x}{n} - \left[\frac{x}{n}\right]\right) + 1,$$

and we apply the *idea* of the Dirichlet hyperbola method to evaluate the sum. For A a large fixed number we obtain

$$\begin{aligned}S(x) = 1 &+ \sum_{n \leq x/A}\left(\frac{x}{n} - \left[\frac{x}{n}\right]\right)\mu(n) + \int_1^A M\left(\frac{x}{t}\right) dt \\ &- \sum_{2 \leq n \leq A} M\left(\frac{x}{n}\right) - (A - [A])M(x/A).\end{aligned}$$

If we now use the fact that $M(y) = o(y)$ as $y \to \infty$, we find that

$$|S(x)| \leq 1 + \frac{x}{A} + \int_1^A o\left(\frac{x}{t}\right) dt + \sum_{n \leq A} o\left(\frac{x}{n}\right) + o\left(\frac{x}{A}\right) \leq \frac{x}{A} + o(x).$$

(The detailed ε estimates are made as in the demonstration of (4.4).) Since A is arbitrary, we have $S(x) = o(x)$. □

5. Selberg's formula. The starting point in the elementary proof of the P.N.T. was the estimate

$$\sum_{p \leq x} \log^2 p + \sum_{pq \leq x} \log p \log q = 2x \log x + O(x) \tag{1.8}$$

found by A. Selberg [**Sel 1**]. (Here p and q run through primes.) In this section we shall derive Selberg's formula and observe some connections with the P.N.T.

THEOREM 5.1. *Let $x \geqslant 1$. Then* (1.8) *holds,*

$$\vartheta(x)\log x + \sum_{p \leqslant x} \vartheta(x/p)\log p = 2x\log x + O(x), \tag{5.1}$$

and

$$\sum_{n \leqslant x} \Lambda(n)\log n + \sum_{n \leqslant x} \Lambda{*}\Lambda(n) = 2x\log x + O(x). \tag{5.2}$$

We first show that (1.8), (5.1) and (5.2) are equivalent (in the sense of §4). Indeed, formula (2.5) for the sum of a convolution and the definition of ϑ yield

$$\sum_{pq \leqslant x} \log p \log q = \sum_{p \leqslant x} \vartheta(x/p)\log p$$

and summation by parts along with the Chebyshev upper bound for $\vartheta(x)/x$ gives

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

To show the equivalence of (5.1) and (5.2) we first sum by parts to obtain

$$\sum_{n \leqslant x} \Lambda(n)\log n = \Psi(x)\log x + O(x)$$

and then use (3.1) to replace Ψ by ϑ. Next, we have

$$\begin{aligned} 0 \leqslant \sum_{n \leqslant x} (\Lambda{*}\Lambda - L1_P{*}L1_P)(n) &= \sum_{n \leqslant x} (\Lambda - L1_P){*}(\Lambda + L1_P)(n) \\ &= \sum_{n \leqslant x} (\Psi(x/n) - \vartheta(x/n))(\Lambda(n) + 1_P(n)\log n) \\ &\leqslant \sum_{n \leqslant x} O\{(x/n)^{2/3}\}2\Lambda(n) = O(x) \end{aligned}$$

by summation by parts and the bound $\Psi(x) = O(x)$. Thus the corresponding terms of (5.1) and (5.2) agree with those of (1.8) to within $O(x)$ error terms.

PROOF OF THEOREM 5.1. We show (5.2) by means of an identity and an estimate. The starting point is Chebyshev's formula (2.3). We apply L to each side of this relation and use the derivation property (2.4) to obtain

$$(L\Lambda){*}1 + \Lambda{*}L = L^2.$$

We apply Chebyshev's identity again, replacing $\Lambda{*}L$ by $\Lambda{*}\Lambda{*}1$. If we convolve each side of the resulting identity by μ, we find that

$$L\Lambda + \Lambda{*}\Lambda = L^2{*}\mu. \tag{5.3}$$

We then sum each side of (5.3) and note that the left side of the summed equation is precisely the left side of (5.2). To treat the right side we first note that, since log is monotone increasing, we have

$$\begin{aligned} \sum_{n \leqslant x} \log^2 n &= \int_1^x \log^2 t\,dt + O(\log^2 ex) \\ &= x\log^2 x - 2x\log x + 2x + O(\log^2 ex). \end{aligned}$$

In §2 we noted Dirichlet's estimate

$$\sum_{n \leq x} 1*1(n) = x \log x + cx + O\left(\sqrt{x}\right).$$

This result and a further application of the hyperbola method (Lemma 2.1) yield the estimate

$$\sum_{n \leq x} 1*1*1(n) = \frac{1}{2} x \log^2 x + c_1 x \log x + c_2 x + O\left(x^{2/3} \log ex\right).$$

Our strategy for treating the expression

$$\sum_{n \leq x} L^2 * \mu(n) = \sum_{n \leq x} \left(\sum_{m \leq x/n} \log^2 m \right) \mu(n),$$

occurring on the right side of the summed form of (5.3), is to approximate $\Sigma \log^2 m$ by an expression

$$G\left(\frac{x}{n}\right) = \sum_{m \leq x/n} (2 \cdot 1*1*1 + a \cdot 1*1 + b \cdot 1)(n)$$

for suitable constants a and b. The sum with G in place of $\Sigma \log^2 m$ is

$$\begin{aligned} \sum_{n \leq x} G\left(\frac{x}{n}\right)\mu(n) &= \sum_{n \leq x} (2 \cdot 1*1*1 + a \cdot 1*1 + b \cdot 1)*\mu(n) \\ &= \sum_{n \leq x} 2 \cdot 1*1(n) + a \cdot 1(n) + b e_1(n) \\ &= 2x \log x + O(x), \end{aligned}$$

which is the right side of (5.2).

It remains to show that a suitable choice of constants a and b will yield

$$\sum_{n \leq x} \left\{ \left(\sum_{m \leq x/n} \log^2 m \right) - G\left(\frac{x}{n}\right) \right\} \mu(n) = O(x). \tag{5.4}$$

We have

$$\begin{aligned} \sum_{n \leq y} \log^2 n &= y \log^2 y - 2y \log y + 2y + O\left(\log^2 ey\right), \\ 2 \sum_{n \leq y} 1*1*1(n) &= y \log^2 y + 2c_1 y \log y + 2c_2 y + O\left(y^{2/3} \log ey\right), \\ a \sum_{n \leq y} 1*1(n) &= ay \log y + acy + O\left(\sqrt{y}\right), \\ b \sum_{n \leq y} 1 &= by + O(1). \end{aligned}$$

If we choose $a = -2 - 2c_1$ and then $b = 2 - 2c_2 - ac$, we obtain

$$\sum_{m \leq y} \log^2 m - G(y) = O\left(y^{3/4}\right).$$

The truth of (5.4) follows immediately from this relation, and the proof of Selberg's formula is complete. □

We note briefly another derivation of Selberg's formula whose starting point is an approximate Möbius inversion in Chebyshev's identity somewhat akin to the method of §3. For $x > 1$ convolve each side of (2.3) by $\nu = (1 - L/\log x)\mu$ and sum over all integers up to x. One side of the resulting equation is

$$\begin{aligned}\sum_{n\leqslant x} \Lambda*1*\nu &= \sum_{n\leqslant x} \Lambda(n) - \frac{1}{\log x} \sum_{n\leqslant x} \Lambda*1*L\mu(n)\\ &= \Psi(x) + \frac{1}{\log x} \sum_{n\leqslant x} \Lambda*\Lambda(n).\end{aligned}$$

The other side of the equation can be treated by first subtracting $1*1 - 2\gamma \cdot 1$ from $L1$ and adding it back on. We then estimate the resulting expressions as we have done above to get $2x + O(x/\log x)$. The method is carried out by Edwards **[Edw]**.

Features of Selberg's formula. Each of the arithmetic functions $L\Lambda$ and $\Lambda*\Lambda$ occurring in (5.2) is nonnegative. Also, we note that the truth of the P.N.T. implies that

$$\Psi(x)\log x \sim x\log x \sim \sum_{n\leqslant x} \Psi\left(\frac{x}{n}\right)\Lambda(n)$$

(the last by approximating $\Psi(x/n)$ by x/n and using Mertens' relation (1.5)), i.e. the two terms on the left side of Selberg's formula have equal "weight". The error term in the formula clearly has lower weight then the main terms. These features of Selberg's formula enable us to give various prime number estimates that were not accessible to Chebyshev.

For example, we can give nontrivial upper bounds for the number of primes lying in relatively short intervals. Since the weighted prime counting function ϑ satisfies $\vartheta(x) \sim x$ as $x \to \infty$ by the P.N.T., we expect that $\vartheta(y) - \vartheta(x) \sim y - x$ as $x \to \infty$, provided that $y - x$ is of sufficient size relative to x. Chebyshev's estimates (1.2) give

$$\frac{\vartheta(y) - \vartheta(x)}{y - x} < \frac{1.11y - .92x}{y - x} = \frac{.19x}{y - x} + 1.11$$

for large x, an estimate which becomes unbounded if $y - x = o(x)$.

We now apply Selberg's formula (1.8) once with argument y and once with x and subtract to obtain

$$\sum_{x<p\leqslant y} \log^2 p + \sum_{x<pq\leqslant y} \log p \log q = 2y\log y - 2x\log x + O(y).$$

If we drop the double sum and make some simple estimates, assuming that $x < y < 2x$, we obtain

$$0 \leqslant \vartheta(y) - \vartheta(x) \leqslant 2(y - x) + O(x/\log x).$$

Thus we have

$$\vartheta(y) - \vartheta(x) \leqslant c(y - x) \tag{5.5}$$

as an upper estimate for the contribution of primes in an interval $(x, y]$ provided that $2x > y > x + c'x/\log x$. This estimate holds for intervals that are significantly shorter than in the Chebyshev case.

Let a and A denote the limit inferior and limit superior respectively of $\Psi(x)/x$. Chebyshev's estimates insure that $.92129\ldots \leqslant a \leqslant A \leqslant 1.10555\ldots$. Selberg's formula shows that a and A are located symmetrically about 1: we have $A + a = 2$. This may be shown easily with the aid of the Mertens relation

$$\sum_{n\leqslant x} \frac{x}{n}\Lambda(n) \sim x\log x. \tag{5.6}$$

We can rewrite Selberg's formula in still another way as

$$\frac{\Psi(x)}{x} + \frac{1}{\log x}\sum_{n\leqslant x}\frac{\Psi(x/n)}{x/n}\frac{\Lambda(n)}{n} = 2 + O\left(\frac{1}{\log x}\right). \tag{5.7}$$

The second term on the left of (5.7) is a complicated average of the function $u \mapsto \Psi(u)/u$ by (5.6). The proof of the P.N.T. will proceed by deducing estimates of Ψ by using information about this average. Such an "unaveraging process" is called a Tauberian method. In the next section we shall describe two Tauberian arguments for deducing the P.N.T. from Selberg's formula.

6. Elementary proof of the P.N.T. There are a number of ways of deducing the P.N.T. from Selberg's formula. Here we shall sketch the methods of Erdös [**Erd 1**] and of Selberg [**Sel 1**], which are of intrinsic as well as historical interest.

We begin with Erdös' "method of compensation". The starting point is the Selberg formula in the form (5.7). We have noted above that $A + a = 2$, where A and a are the limits superior and inferior of $\Psi(x)/x$, and we suppose that $A > a$.

We take a large value of x for which $\Psi(x)/x$ is near A. It follows from (5.6) and (5.7) that $\Psi(x/p)/(x/p)$ is near a for a set S comprising "most" primes $p \leqslant x$. (Here "most" means that

$$\left.\sum_{\substack{n\leqslant x\\ n\notin S}}\frac{\Lambda(n)}{n}\Big/\sum_{n\leqslant x}\frac{\Lambda(n)}{n} \approx 0.\right)$$

Choose a small prime $p_1 \in S$. A similar argument shows that $\Psi(x/p_1p)/(x/p_1p)$ is near A for "most" primes $p \leqslant x/p_1$.

It is then shown to be impossible that $\Psi(x/p) \approx ax/p$ for most $p \leqslant x$ and that $\Psi(x/pp_1) \approx Ax/pp_1$ for most $p \leqslant x/p_1$. The contradiction argument is somewhat delicate; it depends on the fact that the values of the Ψ function do not change too swiftly.

Selberg's argument, like Erdös', depends on a recursion. This time, however, the recursion is built into a variant form of Selberg's formula.

LEMMA 6.1.

$$\sum_{n\leqslant x}(1-\Lambda)(n)\log n = \sum_{n\leqslant x}(1-\Lambda)*(1-\Lambda)(n) + O(x).$$

PROOF. We have

$$L(1-\Lambda) - (1-\Lambda)*(1-\Lambda) = L - L\Lambda - 1*1 + 2\Lambda*1 - \Lambda*\Lambda.$$

If we sum both sides of this equation and recall Selberg's formula, Chebyshev's relation

$$\sum_{n \leq x} \Lambda * 1(n) = \sum_{n \leq x} \log n = x \log x + O(x),$$

and the Dirichlet divisor estimate

$$\sum_{n \leq x} 1 * 1(n) = x \log x + O(x),$$

then the lemma follows. □

If we set $R(x) = [x] - \Psi(x)$ and perform a summation by parts we can write the lemma as

$$R(x) \log x = \sum_{n \leq x} R(x/n)(1 - \Lambda)(n) + O(x). \tag{6.1}$$

The last formula can be expressed more symmetrically as the Stieltjes integral

$$R(x) \log x = \int_{1-}^{x} R(x/t)\, dR(t) + O(x). \tag{6.2}$$

The goal, of course, is to show that $R(x) = o(x)$.

The variation of R is important for making estimates. An argument akin to the one yielding (5.5) gives

$$0 \leq \Psi(x + y) - \Psi(x) \leq 2y + O(x/\log x)$$

for $0 < y < x$. In this range we thus have

$$|R(x + y) - R(x)| \leq y + O(x/\log x).$$

If we pretend that R is differentiable (it is not even continuous!), the last estimate would say that in some average sense

$$\text{“}|R'(x)| \leq 1\text{”}.$$

The preceding ideas can be made rigorous and the following inequality shown:

$$|R(x)| \log x \leq \int_{1}^{x} |R(x/u)|\, du + O(x \log\log 4x). \tag{6.3}$$

Let us define a transformation T on piecewise continuous functions F defined on $[1, \infty)$ by setting

$$TF(x) = \frac{1}{\log x} \int_{1}^{x} F\left(\frac{x}{t}\right) dt.$$

Relation (6.3) can be expressed in terms of T as

$$|R(x)| \leq T|R|(x) + O\{x(\log\log 4x)/\log x\}. \tag{6.4}$$

We hope that an estimate for $|R|$ can be inserted into the right side of the last expression to give a new and better estimate for $|R|$.

Unfortunately, the map T does not induce a contraction for functions $R(x)$ of size roughly x. Indeed, if we know that $|R(x)| < \varepsilon x$ for all sufficiently large values of x, then (6.4) will return the same estimate to us. Thus, a proof of the

P.N.T. using a recursion such as (6.4) must include another ingredient to rule out a relation of the form $R(x) \sim cx$ for some nonzero constant c.

Mertens' estimate (1.5) and an integration by parts give

$$\int_1^x R(t)t^{-2}\,dt = O(1). \tag{6.5}$$

Thus, it is impossible that $R(x) \sim cx$ should hold for a nonzero value of c. One may hope to combine (6.5) with (6.4) to prove that $R(x) = o(x)$.

Selberg's elementary proof of the P.N.T. was carried out with these ingredients along the following lines. Let the interval $[1, \infty)$ be expressed as a union of intervals of maximal length on which $R(x)$ is of one sign. For such an interval $[a, b)$ we have by (6.5)

$$\int_a^b |R(u)|\,u^{-2}\,du = \left|\int_a^b R(u)u^{-2}\,du\right| = O(1). \tag{6.6}$$

We total up the contributions to the right side of (6.4) made by R in each of the intervals $[a, b)$.

It is convenient to change variables and rewrite (6.4) as

$$\frac{|R(x)|}{x} \leqslant \frac{1}{\log x}\int_1^x \frac{|R(u)|}{u^2}\,du + O\left(\frac{\log\log 4x}{\log x}\right). \tag{6.7}$$

If an interval $[a, b)$ where R is of constant sign is "long", then its contribution to the integral in (6.7) is $O(1/\log x)$, which is an improvement over the trivial estimate.

On the other hand, suppose that an interval of constant sign is not "long". In this case we note that R is relatively small near each end of the interval and does not vary too swiftly. Thus the contribution of such an interval $[a, b)$ to the integral in (6.7) is again less than the trivial estimate.

Together these estimates show that if we assume that

$$\limsup_{x\to\infty} |R(x)|/x = \alpha > 0,$$

then there exists $c = c(\alpha) > 0$ such that

$$|R(x)|/x < (1 - c)\alpha$$

holds for all sufficiently large values of x, which is impossible. This concludes our sketch of elementary proofs of the P.N.T. □

A number of variants of this argument have been given (cf. [**HW**, **Lev**, **Nev**]). Differences include replacing (6.3) by the slightly sharper bound

$$|R(x)|\log^2 x \leqslant 2\int_1^x |R(x/u)|\log u\,du + O(x\log x)$$

and the replacement of $R(x) = [x] - \Psi(x)$ by various smoother functions, such as $\int_1^x R(t)\,dt/t$. The last change necessitates an additional, but simple, Tauberian argument to recover Ψ from the smoothed expression.

Direct elementary proofs of the estimates (4.2), (4.4), and (4.5) have also been given (cf. [**Wir 1**, **PR**, **Ami 1**]). The method in each case is based on a formula that is analogous to that of Selberg. The direct estimates of $M(x)$ or $\sum \mu(n)/n$ have the attractive feature that the trivial bound $|M(y) - M(x)| < |y - x| + 1$ can be used to measure the variation of M.

There have been other elementary proofs of the P.N.T. which proceed along rather different lines than the two arguments that we have sketched. Of particular note are the arguments of Wirsing [**Wir 1**] and Bang [**Ban**, **Sch**] which exploit connections between the term $\sum \Psi(x/n)\Lambda(n)$ in Selberg's formula and an expression occurring in certain isoperimetric estimates.

7. Connections with analytic methods. Analytic proofs of the P.N.T. are based on properties of the Riemann zeta function in the complex plane. In particular, the location of the zeros of zeta plays an important role in the analytic theory, as we shall indicate presently. In this section we shall give elementary proofs of some facts about the zeros of zeta. See also [**EMF**, pp. 93–99].

As is the custom in analytic number theory, we express a complex variable in the peculiar form $s = \sigma + it$ with σ and t real. Recall that zeta is defined for $\sigma > 1$ by $\zeta(s) = \sum n^{-s}$. (We shall suppress summation limits when the sum extends over all positive integers.) Summation by parts yields

$$\zeta(s) = \sum n\{n^{-s} - (n+1)^{-s}\} = s \sum n \int_n^{n+1} x^{-s-1}\,dx.$$

If we subtract from this

$$\frac{s}{s-1} = s\int_1^{\infty} x^{-s}\,dx = s\sum \int_n^{n+1} x x^{-s-1}\,dx,$$

we obtain

$$\zeta(s) - \frac{s}{s-1} = s\sum \int_n^{n+1} (n - x)x^{-s-1}\,dx. \tag{7.1}$$

The right side of this formula is analytic for $\sigma > 0$, and it provides an extension of zeta as an analytic function in $\{s \in \mathbf{C}\colon \sigma > 0\}$ with a simple pole at $s = 1$ as its only singularity.

Logarithmic differentiation of (1.4), the Euler product representation of zeta, yields

$$-\zeta'(s)/\zeta(s) = \sum \Lambda(n)n^{-s}, \qquad \sigma > 1. \tag{7.2}$$

(This is the analytic counterpart of the Chebyshev relation $\Lambda * 1 = L$, since

$$\sum L(n)n^{-s} = -\zeta'(s) \quad \text{and} \quad \sum 1n^{-s} = \zeta(s).)$$

The function ζ'/ζ has poles at all points at which zeta vanishes. It follows from the convergence of (7.2) or (1.4) that zeta cannot have any zeros in the open half plane $\{s \in \mathbf{C}\colon \sigma > 1\}$.

We now indicate the connection between the P.N.T. and the absence of zeros of zeta on the line $\{s = 1 + it\colon -\infty < t < \infty\}$. Assume that the P.N.T.

holds. Summation by parts yields

(7.3)

$$-\frac{\zeta'}{\zeta}(s) - \zeta(s) = \sum_{n=1}^{\infty} (\Lambda(n) - 1)n^{-s} = \sum_{n=1}^{\infty} (\Psi(n) - n)(n^{-s} - (n+1)^{-s})$$
$$= \sum_{n=1}^{\infty} o(n)s \int_n^{n+1} x^{-s-1}\,dx = |s| \sum_{n=1}^{\infty} o(1) \int_n^{n+1} x^{-\sigma}\,dx = |s|\, o\left(\frac{1}{\sigma - 1}\right)$$

for $\sigma = \operatorname{Re} s \to 1+$. If $\zeta(1 + it_0)$ were zero for some real $t_0 \neq 0$, then ζ'/ζ would have a simple pole there with positive integer residue. However, (7.3) insures that

$$\lim_{\sigma \to 1+} (\sigma - 1)\frac{\zeta'}{\zeta}(\sigma + it_0) = 0. \tag{7.4}$$

Thus ζ can have no zeros in the *closed* half plane $\{s\colon \sigma \geqslant 1\}$.

The Tauberian theorem of Wiener and Ikehara (cf. [**Wie 1, 2, Lnd 4, Wid, Cha 1**,...]) establishes the converse proposition: If $\zeta(1 + it) \neq 0$, $-\infty < t < \infty$, then $\Psi(x) \sim x$. Thus, we see that in the analytic theory the nonvanishing of zeta on the vertical line $\{s\colon \operatorname{Re} s = 1\}$ is "equivalent" to the P.N.T. in the sense that each can be deduced from the other (modulo properties of analytic functions and the Wiener-Ikehara theorem).

THEOREM 7.1. $\zeta(1 + it) \neq 0$, $-\infty < t < \infty$.

The most familiar proof of this theorem uses the inequality

$$-3\frac{\zeta'}{\zeta}(\sigma) - 4\operatorname{Re}\frac{\zeta'}{\zeta}(\sigma + it) - \operatorname{Re}\frac{\zeta'}{\zeta}(\sigma + 2it)$$
$$= \sum n^{-\sigma}\{3 + 4\cos(t\log n) + \cos(2t\log n)\}\Lambda(n)$$
$$= 2\sum n^{-\sigma}\{1 + \cos(t\log n)\}^2\Lambda(n) \geqslant 0,$$

which is valid for any $\sigma > 1$ and any real t. The inequality implies that if zeta were zero at $1 + it_0$, then it would have to have a pole at $1 + 2it_0$. But this is impossible, since zeta is analytic at all points of the line $\{s\colon \sigma = 1\}$ different from $s = 1$. This proof can be modified to eliminate all use of analytic functions.

We now give two different elementary proofs of Theorem 7.1. The point here is to avoid introducing any properties of analytic functions. We regard a function of a complex variable as a function of two real variables and define the derivative (when it exists) by

$$f'(z) = \lim(\{f(z + u) - f(z)\}/u)$$

as $u \to 0$ through real values. We consider zeta to be defined on the "nearly closed" half plane $\{s\colon \sigma \geqslant 1, s \neq 1\}$ by continuity. Clearly, zeta satisfies (7.1) on this region. Also, zeta has a derivative in the foregoing sense, and it can be represented by the formal differentiation of (7.1). The derivative is a continuous function on the nearly closed half plane.

The P.N.T. can be proved by an elementary argument and, as we have seen above, relation (7.4) is an elementary consequence of the P.N.T. We can complete one elementary proof of Theorem 7.1 with the following lemma.

LEMMA 7.2. *Suppose* $\zeta(1+it_0)=0$. *Then*

$$\lim_{\sigma\to 1+}(\sigma-1)\frac{\zeta'}{\zeta}(\sigma+it_0)=1.$$

PROOF. First, we claim that $|\zeta'(1+it_0)|\geqslant 1$. Indeed, we have

$$\log\{\zeta(\sigma)\,|\zeta(\sigma+it_0)|\}=\sum_{n=2}^{\infty}n^{-\sigma}\frac{\Lambda(n)}{\log n}\{1+\cos(t_0\log n)\}\geqslant 0,$$

so

$$|\zeta(\sigma+it_0)|\geqslant 1/\zeta(\sigma)\sim\sigma-1 \qquad (\sigma\to 1+).$$

Thus

$$|\zeta'(1+it_0)|=\lim_{\sigma\to 1+}\left|\frac{\zeta(\sigma+it_0)-\zeta(1+it_0)}{\sigma-1}\right|\geqslant 1.$$

Now the definition of derivative and its continuity at $1+it_0$ imply that

$$\frac{(\sigma-1)\zeta'(\sigma+it_0)}{\zeta(\sigma+it_0)}=\frac{(\sigma-1)\{\zeta'(1+it_0)+o(1)\}}{(\sigma-1)\{\zeta'(1+it_0)+o(1)\}}\to 1$$

as $\sigma\to 1+$. (Here $o(1)$ denotes a function that tends to zero as $\sigma\to 1+$.) □

Another elementary proof of Theorem 7.1. The following argument, which was suggested by an article of Wiener and Geller [**WG**], gives a more direct elementary proof of Theorem 7.1. The argument proceeds in two stages. First, we show with the aid of Lemma 7.2 that if $\zeta(1+it_0)=0$, then $t_0\log p$ is nearly congruent to π (modulo 2π) for most primes p. Then we shall use the "continuity estimate" (5.5) to show that such a concentration of primes in short intervals cannot occur.

Let $\|x\|$ denote the distance of a real number x from the nearest integer. For $0<\delta<1/2$ let

$$I_\delta=\left\{u\geqslant 1:\left\|\frac{t_0\log u}{2\pi}-\frac{1}{2}\right\|<\delta\right\},$$

i.e.

$$I_\delta=\bigcup_{k=0}^{\infty}\left(e^{(2k+1-2\delta)\pi/t_0},\,e^{(2k+1+2\delta)\pi/t_0}\right)=\bigcup_{k=0}^{\infty}(a_k,b_k),$$

say, where we have assumed without loss of generality that $t_0>0$.

We shall show that if $\zeta(1+it_0)=0$ and $0<\delta<1/5$, then

$$\lim_{\sigma\to 1+}(\sigma-1)\sum_{n\notin I_\delta}n^{-\sigma}\Lambda(n)=0. \tag{7.5}$$

By taking real parts in Lemma 7.2 we see that if $\zeta(1+it_0)=0$, then

$$(\sigma-1)\sum n^{-\sigma}\Lambda(n)\cos(t_0\log n)\to -1$$

as $\sigma \to 1+$, and hence

$$(\sigma - 1) \sum n^{-\sigma}\Lambda(n)\{1 + \cos(t_0 \log n)\} \to 0$$

as $\sigma \to 1+$. The last expression is at least

$$(\sigma - 1) \sum_{n \notin I_\delta} n^{-\sigma}\Lambda(n)\{1 + \cos(t_0 \log n)\} \geqslant (\sigma - 1)(1 - \cos 2\pi\delta) \sum_{n \notin I_\delta} n^{-\sigma}\Lambda(n),$$

which proves (7.5).

Next, we shall show that there exists an absolute constant $B > 0$ such that for $0 < \delta < 1/5$

$$\varlimsup_{\sigma \to 1+} (\sigma - 1) \sum_{n \in I_\delta} n^{-\sigma}\Lambda(n) \leqslant B\delta. \tag{7.6}$$

The sum is at most

$$\sum_{k=0}^{\infty} a_k^{-\sigma} \sum_{a_k < n < b_k} \Lambda(n).$$

By the analogue of (5.5) for Ψ, we have

$$\sum_{a_k < n \leqslant b_k} \Lambda(n) < 3(b_k - a_k) = 3a_k(e^{4\pi\delta/t_0} - 1)$$

for all sufficiently large k, since $b_k/a_k = \exp 4\pi\delta/t_0$, a positive quantity depending only on δ and t_0. (This estimate could also have been obtained by using a sieve.) Finally, we have

$$3(e^{4\pi\delta/t_0} - 1) \sum_{k=0}^{\infty} a_k^{1-\sigma} < \frac{B\delta}{\sigma - 1},$$

which establishes (7.6).

Now suppose that $\zeta(1 + it_0)$ were zero. If we choose δ so that $0 < B\delta \leqslant 1/2$, then we have by (7.5) and (7.6)

$$\varlimsup_{\sigma \to 1+} (\sigma - 1) \sum n^{-\sigma}\Lambda(n) \leqslant 1/2.$$

On the other hand we know that

$$\sum n^{-\sigma}\Lambda(n) = -\frac{\zeta'}{\zeta}(\sigma) \sim \frac{1}{\sigma - 1}$$

as $\sigma \to 1+$. This contradiction shows that zeta cannot vanish on the line $\{s \in \mathbf{C}: \sigma = 1\}$. $\square$

8. Elementary error estimates. The first elementary P.N.T. error estimates had the form

$$\pi(x) - x/\log x = O(x/\log^c x)$$

with $c \in (1, 2)$. The relatively meager size of the error term is a consequence of the use of Selberg's formula. Indeed, one can verify that *any* right-continuous monotone nondecreasing function π^* on $[2, \infty)$ having the form

$$\pi^*(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + O\left(\frac{x}{\log^c x}\right),$$

with $c > 2$, will satisfy the following generalized form of Selberg's formula:

$$\int_1^x \log^2 t \, d\pi^*(t) + \iint_{1 \leq st \leq x} \log s \, d\pi^*(s) \log t \, d\pi^*(t) = 2x \log x + O(x).$$

The integrals here may be taken in the Riemann-Stieltjes or Lebesgue-Stieltjes sense.

The improved error terms of Bombieri [**Bom 1**] and Wirsing [**Wir 2**] were achieved by overcoming the limitations of the Selberg formula. Bombieri utilized a class of analogues of Selberg's formula of greater "weight"; Wirsing introduced a recursion whereby each time that an error estimate was found for $\pi(x) - \operatorname{li} x$, it was used again to obtain an improved form of Selberg's formula.

Here we shall describe the estimation method of Diamond-Steinig [**DS**]. Like earlier elementary P.N.T. proofs the argument proceeds in two stages. First a Selberg type formula is derived and then the P.N.T. is deduced from it by a Tauberian method.

The starting point is the observation that the individual terms occurring in Selberg's formula (5.2) have representations

$$L\Lambda = L^2 * \mu - L * L * \mu * \mu \quad \text{and} \quad \Lambda * \Lambda = L * L * \mu * \mu.$$

These expressions each contain two convolution factors of the troublesome Möbius μ function, but their sum happily contains only one factor of μ. This fact leads to the $O(x)$ error term on the right side of (5.2).

One could attempt to improve the error estimate in the P.N.T. by finding other expressions in place of $L\Lambda$ and $\Lambda * \Lambda$, ones which had greater "weight" but which when added still contained relatively few Möbius convolution factors. Expressions of this type had occurred in Bombieri's methods. It was conjectured by Paul J. Cohen, by consideration of logarithmic derivatives, that one should consider expressions

$$L^{2k-1}\Lambda/\Gamma(2k) + L^{k-1}\Lambda * L^{k-1}\Lambda/\Gamma(k)^2, \tag{8.1}$$

where k is any positive integer and $\Gamma(k) = (k-1)!$

Diamond-Steinig showed that this expression is represented as a complicated polynomial in the arithmetic functions $1, L, L^2, \ldots,$ and μ, where the multipication implicit in the polynomial is $*$, multiplicative convolution. The noteworthy feature of this polynomial is that no term contains more than k convolution factors of μ. The expression (8.1) is summed over the integers in $[1, x]$ and the sum is estimated by use of Lemma 2.1, the hyperbola method. The arguments are analogous with those of §5, but are of course more intricate.

For all $k \geqslant K$ and $x \geqslant 1$ we find that

$$(8.2)\quad \sum_{n \leqslant x} \frac{L^{2k-1}\Lambda(n)}{\Gamma(2k)} + \frac{L^{k-1}\Lambda}{\Gamma(k)} * \frac{L^{k-1}\Lambda(n)}{\Gamma(k)} = \frac{2}{\Gamma(2k)} \int_1^x \log^{2k-1} t\, dt + \frac{\theta (Ak^4)^k}{\Gamma(2k)} x \log^k x,$$

where θ is a number of absolute value at most 1 and A is an absolute constant. For $K = 10$, we can take $A = 697$.

This generalized Selberg formula has a number of noteworthy features. Its uniformity will be needed in the estimate of $\Psi(x) - x$, when the parameter k will be chosen depending on x. For large k the main term and error term differ significantly in size, an improvement over the Selberg formula. Of course, the O-constant grows with k. The error term in (8.2) contains one additional $\log x$ factor, which was tolerated for convenience of estimation.

If one now sets

$$r(n) = r_k(n) = (\log n)^{k-1}(1 - \Lambda(n)),$$

formula (8.2) can be rewritten as

$$(8.3)\quad \sum_{n \leqslant x} \frac{L^k r(n)}{\Gamma(2k)} = \sum_{n \leqslant x} \frac{r * r(n)}{\Gamma(k)^2} + \theta \frac{(A'k^4)^k x \log^k x}{\Gamma(2k)}$$

where A' can be taken as 698 for all $k \geqslant 10$.

The quantity

$$R(x) = R_k(x) = \sum_{n \leqslant x} r_k(n)$$

is estimated by an iterative process analogous to that used by Wirsing. The iteration is started by use of a weighted, uniform version of the following remarkable inequality of Wirsing [**Wir 2**, **Cha 2**].

LEMMA 8.1. *Let f and g be measurable real-valued functions on* $(0, \infty)$ *and assume that*

$$\limsup_{x \to \infty} x^{-1} \int_0^x f(t)^2\, dt \leqslant F, \qquad \limsup_{x \to \infty} x^{-1} \int_0^x g(t)^2\, dt \leqslant G$$

for some positive numbers F and G. Let h be defined on $(0, \infty)$ *by*

$$h(x) = x^{-1} \int_0^x f(x - t) g(t)\, dt.$$

If

$$\lim_{x \to \infty} x^{-1} \int_0^x h(t)\, dt = 0,$$

then

$$\limsup_{x \to \infty} x^{-1} \int_{x_0}^x h(t)^2\, dt \leqslant \frac{1}{2} FG$$

for any fixed $x_0 > 0$.

(The error estimates of Lavrik-Sobirov **[LS]** and Sokolovski **[Sok]** derive from improvements at this stage. They observed that the weighted, uniform version of Lemma 8.1 could be proved under less restrictive hypotheses. This change increases the range of applicability of the estimate and leads to the better error term.)

By means of the iteration Diamond and Steinig found that

$$|R_k(x)| < e^{-ck} x \log^{k-1} x$$

for some $c \in (0, 1)$ and all $x \geqslant \exp\{5 \cdot 10^{13} k^{7+\varepsilon}\}$. A simple estimate is made for $|R_k(x)|$ for smaller values of x. An estimate of $|\Psi(x) - x|$ can now be deduced. If one integrates by parts the identity

$$[x] - \Psi(x) = 1 + \int_{3/2}^{x} (\log t)^{1-k}\, dR_k(t)$$

and uses the preceding bounds for R_k, the estimate

$$|\Psi(x) - x| < x \exp\{-ck\}$$

is obtained for all $x \geqslant \exp\{7.5 \cdot 10^{13} k^{7+\varepsilon}\}$. For given x the parameter k is chosen as large as possible, subject to the last condition. This yields the estimate

$$\Psi(x) = x + O\big(x \exp\{-c(\log x)^{\alpha}\}\big),$$

with $\alpha = 1/7 - \varepsilon$.

The size of the error estimate of $\Psi(x) - x$ is affected by the error estimates in the generalized Selberg formula and by the efficiency of the Tauberian theorem. Let us speculate briefly on the effects of a possible improvement at each stage. First, suppose that some remarkable argument could show that

$$\sum_{n \leqslant x} \frac{L^{2k-1}\Lambda(n)}{\Gamma(2k)} \quad \text{and} \quad \sum_{n \leqslant x} \frac{L^{k-1}\Lambda}{\Gamma(k)} * \frac{L^{k-1}\Lambda(n)}{\Gamma(k)}$$

were almost exactly the same size. It would follow from (8.2) that

$$\sum_{n \leqslant x} L^{2k-1}\Lambda(n) = \int_1^x (\log t)^{2k-1}\, dt + \theta(Ak^4)^k x \log^k x,$$

and summation by parts would yield

$$\Psi(x) = x + \theta(A'k^4)^k x / (\log x)^{k-1}.$$

The choice $k \approx e^{-1}(A'^{-1} \log x)^{1/4}$ would give

$$\Psi(x) = x + O\{x e^{-4k} \log x\} = x + O\big\{x \exp\big(-A''(\log x)^{1/4}\big)\big\}.$$

In order to reach de la Vallée Poussin's error term (exponent $= 1/2$) by use of a generalized Selberg formula (and a very efficient Tauberian method), we would need an analogue of (8.2) with an error term $\theta(Ak^{\alpha})^k x (\log x)^k / \Gamma(2k)$ with $\alpha \leqslant 2$.

9. Improved Chebyshev-type estimates. In §2 we discussed the upper and lower estimates for $\Psi(x)/x$ obtained by Chebyshev and by Sylvester with the aid of the identity

$$L * \nu = \Lambda * 1 * \nu, \tag{9.1}$$

where ν is a suitable approximation to the Möbius function. More accurate estimates by such a method evidently require still more complicated functions ν.

Given a positive number ε, is it possible to find a function ν, depending on ε, such that the method of §3 yields

$$\limsup_{x\to\infty} | \Psi(x)/x - 1 |< \varepsilon, \tag{9.2}$$

or is there some natural limitation to the method? In this section we shall show that the first alternative holds: For any $\varepsilon > 0$ we shall show how to construct a function $\nu = \nu_\varepsilon$ for which (9.2) holds.

Relation (9.2) with arbitrary $\varepsilon > 0$ is, of course, a form of the P.N.T. Any direct construction of an appropriate function ν for each $\varepsilon > 0$ would give another type of elementary proof of the P.N.T. We are going to establish the existence of a function ν by using the P.N.T. Thus we are *not* giving another proof of the P.N.T. However, our argument guarantees that a constructive search will turn up a desired function ν.

This result has a curious history. It was first discovered by J. B. Rosser and independently by P. Erdös and L. Kalmár in the late 1930's. Because of various misadventures, neither article ever was published **[Erd 3]**. A proof was reconstructed by H. Diamond and Erdös **[DE]** and another one was given by Diamond and K. McCurley **[DM]**.

We shall describe the construction of **[DE]**. We would like to take a truncated version of the Möbius function as ν. Our function ν should satisfy the condition (cf. §3)

$$\sum_{n=1}^{\infty} \nu(n)/n = 0. \tag{9.3}$$

However, it is easy to show, e.g. with the aid of Bertrand's postulate, that no partial sum $\Sigma_{n<T}\mu(n)/n$ is zero.

For T an integer exceeding 1 we define an arithmetic function μ_T by setting

$$\mu_T(n) = \begin{cases} \mu(n), & 1 \leqslant n < T, \\ -T \sum\limits_{i<T} \mu(i)/i, & n = T, \\ 0, & n > T. \end{cases}$$

We are going to apply μ_T for some appropriate T as the function ν in (9.1). By construction μ_T satisfies (9.3).

One can find a sequence of integers $T \to \infty$ such that each satisfies

$$\left| \sum_{n\leqslant x} L*\mu_T(n) - x \right| < \varepsilon x$$

for all x exceeding some X (depending on T). This result is proved without appeal to the P.N.T. by showing that

$$\sum_{n\leqslant x} L*\mu_T(n) \sim x m_1(T),$$

where

$$m_1(T) = \sum_{n<T} \left(\log \frac{T}{n}\right) \frac{\mu(n)}{n}$$

is a function which tends to 1 for a sequence of T's tending to infinity. Indeed, the integral transformation

$$s \int_1^\infty u^{-s-1} m_1(u)\, du = \frac{1}{s\zeta(s+1)} \qquad (s > 0)$$

provides an average of m_1, and it follows from (7.1) that $s\zeta(s+1) \to 1$ as $s \to 0+$.

It remains to show that the summatory function of the right side of (9.1) is close to $\Psi(x)$. We set

$$g = g_T = 1 * \mu_T \quad \text{and} \quad E(x) = E_T(x) = \sum_{n \leq x} g(n).$$

Then the Dirichlet hyperbola method (Lemma 2.1) yields

$$\begin{aligned} \sum_{n \leq x} \Lambda * g(n) &= \sum_{j \leq T-1} \Psi\left(\frac{x}{j}\right) g(j) \\ &\quad + \sum_{i \leq x/(T-1)} E(x/i)\Lambda(i) - E(T-1)\Psi\left(\frac{x}{T-1}\right) \\ &= \mathrm{I} + \mathrm{II} - \mathrm{III}, \quad \text{say.} \end{aligned}$$

Since $\mu_T(n) = \mu(n)$ for $1 \leq n < T$, we have

$$g(j) = 1 * \mu(j) = \begin{cases} 1, & \text{if } j = 1, \\ 0, & \text{if } 1 < j < T. \end{cases}$$

Thus $\mathrm{I} = \Psi(x)$ and $\mathrm{III} = \Psi(x/(T-1)) = O(x/T)$.

The estimate of II is made with the aid of bounds on E. The preceding remarks about g imply that $E(y) = 1$ for $1 \leq y < T$. Also, it is not hard to verify that $E(y) = O(T)$ for all values of y. One can show that $|E(y)|$ is in fact much smaller than T for a long interval. It is here that a form of the P.N.T. ($M(y) = o(y)$) is used. We deduce that $|\mathrm{II}| + |\mathrm{III}| < \varepsilon x/2$ and hence that (9.2) holds.

The problem with an argument such as this is that a finite amount of arithmetical information appears sufficient to give only upper and lower estimates for $\Psi(x)/x$, but no asymptote. In contrast, the knowledge that zeta is nonvanishing at all points of the line $\{s\colon \operatorname{Re} s = 1\}$ provides—in one stroke—enough information to deduce the P.N.T.

An interesting variant on the Chebyshev method, involving a diophantine type argument, was put forth by A. O. Gelfond and L. G. Schnirelmann **[Gel 2]** and was recently rediscovered by M. Nair **[Nai]**. The argument is based on the following observations:

(a) $d_n \overset{\text{def}}{=} \text{l.c.m.}\{1, 2, \ldots, n\} = e^{\Psi(n)}$;

(b) if $P(x)$ is a polynomial of degree $\leqslant n - 1$ and has integer coefficients, then

$$d_n \int_0^1 P(x)\, dx \in \mathbf{Z};$$

and

(c) the smallest positive integer is 1. Thus, if $\int P(x)\, dx > 0$, we have

$$\Psi(n) \geqslant \log\left\{1 \Big/ \int_0^1 P(x)\, dx\right\}.$$

The choice $P(x) = x^k(1 - x)^k$ and some simple upper estimates of $P(x)$ yield the lower bound $\underline{\lim}_{x\to\infty} \Psi(x)/x \geqslant \log 2$. There are other polynomials that yield somewhat better lower estimates for $\Psi(x)/x$, but it is known that polynomials in one variable that are nonnegative on $[0, 1]$ cannot achieve the bound $\underline{\lim}_{x\to\infty} \Psi(x)/x > .9$ [**Apa**]. Perhaps better estimates can be obtained by using polynomials in several variables.

10. Some other elementary results. The years since the discovery of an elementary proof of the P.N.T. have seen a general renaissance in number theory. We shall note briefly some of the areas in which elementary prime number theoretic methods have played a role.

Sieve theory provided the impetus for Selberg's elementary prime number investigation [**Sel 1**]. Indeed the arithmetic function $\{L\Lambda + \Lambda * \Lambda\}(n)$ occurring in Selberg's formula has the sieve property of being nonzero precisely when n is a number having at most two distinct prime divisors. There have been further developments in sieve theory along the lines pioneered by Brun and Selberg [**HaRi, Iwa**].

During this time the radically different large sieve was developed [**Dav, Bom 3, Mon, Ric**]. With the aid of the large sieve the remarkable theorem of Bombieri [**Bom 2**] and A. I. Vinogradov [**ViA**] was proved. This gives good simultaneous estimates on the distribution of the primes in all eligible residue classes with respect to many different moduli. One important application of this result was to Brun or Selberg type sieves involving primes.

Sieve theory is an active area today utilizing a combination of elementary and analytic ideas. One of its most significant recent accomplishments is Chen's theorem [**Chn, HaRi, Ros**] that every sufficiently large even number can be expressed as a sum of a prime and a number having at most two prime factors. This is an approximation to the famous unsolved Goldbach problem of showing that every even number exceeding 2 is representable as a sum of two primes.

Elementary methods occur in many "almost all" results. We say that a proposition P holds for almost all integers if

$$\#\{n \leqslant x: P \text{ does not hold for } n\} = o(x)$$

as $x \to \infty$. One identifies a certain set of "bad" integers and shows them to be sparse (usually by an elementary argument, often involving a sieve estimate). Then the complementary set is shown to have the desired property. For example, P. Erdös has used such arguments to estimate the modulus of continuity of the distribution function of the arithmetic function $n \mapsto \sigma(n)/n$, where σ denotes the sum of divisors function [**Erd 2**].

It had been asserted by A. Wintner [**Win**] that any multiplicative function f which assumes only the values $+1$ and -1 has a mean value, i.e. there exists a number c, the mean value, such that

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\leq x} f(n) = c. \tag{10.1}$$

(An arithmetic function is called *multiplicative* if $f(mn) = f(m)f(n)$ for all pairs of relatively prime positive integers m and n. It is called *completely multiplicative* if $f(mn) = f(m)f(n)$ for all positive integers m and n.)

Suppose that we apply Wintner's assertion to Liouville's arithmetic function λ, which is defined by $\lambda(n) = (-1)^r$, where r is the total number of prime factors of n (counting multiplicity). A brief calculation shows that in this case the mean value must be 0, i.e.

$$L(x) = \sum_{n\leq x} \lambda(n) = o(x), \qquad x \to \infty.$$

Now, we can represent μ, the Möbius function, in terms of λ by the convolution formula $\mu = \lambda * g$, where g is a multiplicative function which satisfies

$$g(1) = 1, \qquad g(p^2) = -1, \qquad \text{and} \qquad g(p^\alpha) = 0 \quad \text{for } \alpha = 1, 3, 4, 5, \ldots$$

for each prime p. Then $\Sigma\,|\,g(n)\,|/n < \infty$ and so

$$M(x) = \sum_{n\leq x} \mu(n) = \sum_{n\leq x} L\left(\frac{x}{n}\right)g(n) = \sum_{n\leq x} o\left(\frac{x}{n}\right)|\,g(n)\,| = o(x),$$

an assertion equivalent to the P.N.T., as we noted in §4. Thus the mean value assertion is not superficial.

In 1967 K. Corrádi [**Crd**] gave a conditional elementary proof of Wintner's assertion. The condition, that all partial sums of the series $f(1)/1 + f(2)/2 + f(3)/3 + \ldots$ be bounded, is easily verified for λ or μ. At the same time, E. Wirsing [**Wir 3**] gave an unconditional proof of Wintner's claim. Wirsing's argument, based on an analogue of Selberg's formula, was elementary but hardly simple. Shortly afterward, there appeared the analytic argument of G. Halász [**Hls**], which has attracted wide interest.

J. Pintz [**Pin**] published a series of articles on the use of elementary methods in studying L functions of real Dirichlet characters. He considered the location of zeros of L functions, their size at the special argument $s = 1$, and the consequent estimates of the class number of the associated imaginary quadratic field. In the last article in the series [**Pin**, VIII], a proof of the Siegel-Walfisz Theorem is given which is inspired by elementary considerations but also contains some complex analysis.

The following striking theorem is established in [**Pin**, VII] for a class of arithmetic functions reminiscent of the foregoing functions of Wintner. Let f denote a completely multiplicative function which assumes only the values $+1$, 0, and -1, and let x be some positive number for which the one-sided bound

$$f(1) + f(2) + \cdots + f([x]) \leq \varepsilon x$$

holds. Then, for the same x, the inequality

$$\sum_{d \leqslant x} f(d)/d \leqslant 2(1 - e^{-1/2} + \delta)\log x$$

holds, where $\delta = \delta(\varepsilon, x) \to 0$ as $x \to \infty$ and $\varepsilon \to 0$. This inequality is sharp, as shown by the example $f(p) = 1$ for $p \leqslant x^{1/\sqrt{e}}$, $f(p) = -1$ for $p > x^{1/\sqrt{e}}$. This theorem is combined with character sum estimates of D. A. Burgess **[Bur]** (which are not elementary) to obtain new upper bounds for $L(1, \chi)$ in case χ is a real primitive character or a real nonprincipal character.

Hardy and Littlewood **[HL]** conjectured that

(10.2) "$\pi(x + y) \leqslant \pi(x) + \pi(y)$ for all $x, y \geqslant 2$",

i.e. "no interval $(y, x + y]$ contains more primes than the initial interval $(0, x]$." This is a plausible conjecture, because the P.N.T. implies that for each x, (10.2) holds for "most" values of y. Also, no numerical counterexample to (10.2) has been found.

We preface another of their conjectures with two examples. Are there infinitely many different triples of prime numbers, each having the form n, $n + 2$, $n + 10$? The answer is No, because $0, 2, 10 \equiv 0, 2, 1 \pmod 3$ and so for each n, the numbers n, $n + 2$, and $n + 10$ are congruent to 0, 2, and 1 (mod 3) in some order. It follows that one of n, $n + 2$, and $n + 10$ must be a multiple of 3.

We say that a sequence $b_1, b_2, \dots, b_k$ is *admissible* if for each prime p there is some residue class (modulo p) which contains none of $b_1, \dots, b_k$. For example, 0, 2, 6 is an admissible triple. There is no congruential barrier to the possibility that n, $n + 2$, and $n + 6$ are all primes for infinitely many integers n. These examples suggest

(10.3) *The prime k-tuples conjecture*. Let $b_1, b_2, \dots, b_k$ be an admissible sequence. Then there exist infinitely many integers n for which all of the numbers $n + b_1, \dots, n + b_k$ are primes.

This is a natural generalization of the twin prime conjecture that n and $n + 2$ are each primes for infinitely many integers n. Hardy and Littlewood offered a heuristic argument that if $b_1, \dots, b_k$ is an admissible sequence, then

$$\#\{n \leqslant x : n + b_1, \dots, n + b_k \text{ are all primes}\} \sim Cx/\log^k x,$$

where C is a positive number depending on the arithmetic properties of the particular admissible sequence. This heuristic estimate has been well supported by numerical evidence **[Pol]**.

D. Hensley and I. Richards **[HeR 1, 2]** showed that it is impossible that both of these conjectures be true. Their argument, however, does not allow one to decide which of the conjectures is false. It proceeds along the following lines. The prime k-tuples conjecture is assumed to hold. A sieve argument and the P.N.T. with a modest error term are used to show that (10.2) must fail to hold for an infinite number of integers y. It seems most likely that (10.3) is in fact true and (10.2) is false.

11. Comparison with analytic methods. We conclude with some remarks on the relative efficiency of elementary and analytic methods for various problems in prime number theory. It is well to note here that many results have been achieved by a mixture of elementary and analytic methods. We shall describe a method as being "mainly elementary" if its main ideas are elementary and some details are carried out by analytic means. A number of sieve estimates are of this type.

The choice of a preferred approach to a problem depends on a number of considerations: the nature of the data; what we are seeking; what representations and techniques are available and simple to use. For example, it would be generally deemed inappropriate to prove analytically the assertion that if $n > 3$, then n, $n + 2$, and $n + 4$ are not all primes.

On the other hand, analytic methods often organize arithmetic data in a form that is ideally suited for the proof of "oscillation theorems". For example, either of the assertions

$$\text{“}\Psi(x) = x + o\left(\sqrt{x}\right)\text{”} \quad \text{or} \quad \text{“}M(x) = o\left(\sqrt{x}\right)\text{”}$$

can be shown to be false by an easy analytic argument based on known information about the location of zeros of the Riemann zeta function. The author knows of no elementary argument that will show even that $M(x)$ is unbounded.

A problem is a likely candidate for an analytic method if it contains a relatively high degree of structure and the result sought is a gross count. For example, nontrivial elementary and analytic estimates are known for $\pi(x)$ and $\Sigma_{n<x} 1*1(n)$ (the "divisor problem"), but in both cases the most accurate estimates have been achieved by analytic methods using information known about the Riemann zeta function.

If a problem does not have a high degree of structure that we can exploit or if a fine count is sought, then a (mainly) elementary approach is more likely to succeed. For example, no simple mechanism for detecting twin primes is presently known, and most results concerning twin primes have been achieved by mainly elementary methods. Also, the best upper estimates of the number of primes in short intervals have been made by elementary methods [**Sel 3**, **HaRi**].

There has been a revolution in the treatment of problems of numerical approximation and explicit bounds, brought on by the advent of computers. In earlier times virtually all explicit estimates (e.g. the proof of Bertrand's Postulate in §3) were made with elementary methods. The reason for this was the need to obtain expressions that were simple enough to estimate. This consideration has waned in importance with the availability of fast and economical computers. Today it is also possible, and sometimes perferable, to make numerical estimates of quantities which are represented in terms of Fourier transforms and other analytic expressions. The important prime number theoretic estimates of Rosser and Schoenfeld [**RS 1, 2, 3**], for example, have been achieved in this way.

ACKNOWLEDGMENTS. The author has benefited from the ideas of many people in the preparation of this article. Particular thanks are extended to Professors P. T. Bateman, H. L. Montgomery, R. Narasimhan, and J. Steinig for their many helpful comments.

REFERENCES

[Ami] S. A. Amitsur, [1] *On arithmetic functions*, J. Anal. Math. **5** (1956/57), 273–314. [2] *Arithmetic linear transformations and abstract prime number theorems*, Canadian J. Math. **13** (1961), 83–109; Corrigendum, *ibid.* **21** (1969), 1–5.

[Apa] E. Aparicio Bernardo, *Métodos para el cálculo aproximado de la desviación diofántica uniforme mínima a cero en un segmento*, Rev. Math. Hispanico-Americana **38** (1978), 259–270.

[Axe] A. Axer, *Über einige Grenzwertsätze*, Sitzber. Akad. Wiss. Wien, Math.-nat. Kl. **120** (1911), Abt. IIa, 1253–1298.

[Ayo] R. Ayoub, [1] *On Selberg's lemma for algebraic fields*, Canad. J. Math. **7** (1955), 138–143. [2] *Euler and the zeta function*, Amer. Math. Monthly **81** (1974), 1067–1086.

[Ban] Th. Bang, *An inequality for real functions of a real variable and its application to the prime number theorem*, Proc. Conf. (Oberwolfach, 1963), Birkhauser, Basel, 1964, pp. 155–160.

[Ber] J. Bertrand, *Mémoire sur le nombre de valeurs que peut prendre une fonction quand on y permute les lettres qu'elle renferme*, J. École Roy. Poly. **18** (1845), 123–140.

[Boh] H. Bohr, *Address of Professor Harald Bohr*, Proc. Internat. Congr. Math. (Cambridge, 1950) vol. 1, Amer. Math. Soc., Providence, R.I., 1952, pp. 127–134.

[Bom] E. Bombieri, [1] *Sulle formule di A. Selberg generalizzate per classi di funzioni aritmetiche e le applicazioni al problema del resto nel "Primzahlsatz"*, Riv. Math. Univ. Parma (2) **3** (1962), 393–440. [2] *On the large sieve*, Mathematika **12** (1965), 201–225. [3] *Le grand crible dans la théorie analytique des nombres*, Astérisque No. 18, Soc. Math. de France, Paris, 1974.

[Bre] R. Breusch, [1] *Another proof of the prime number theorem*, Duke Math. J. **21** (1954), 49–53. [2] *An elementary proof of the prime number theorem with remainder term*, Pacific J. Math. **10** (1960), 487–497.

[Bru] V. Brun, [1] *La série* $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \cdots$ *où les dénominateurs sont "nombres premiers jumeaux" est convergente ou finie*, Bull. Sci. Math. (2) **43** (1919), 100–104 and 124–128. [2] *Le crible d'Eratosthène et le théorème de Goldbach*, Vid. Sel. Skr. mat. naturw. I, No. 3 (1920).

[Buc] A. A. Buchstab, [1] *Asymptotic estimates of a general number theoretic function*, Mat. Sb. (N.S.) **2** (44) (1937), 1239–1246. (Russian) [2] *Combinatorial strengthening of the sieve of Eratosthenes method*, Uspehi Mat. Nauk **22** (1967), 199–226 (Russian); English transl. Russian Math. Surveys **22** (1967), 205–233.

[Bur] D. A. Burgess, *On character sums and L series*. II, Proc. London Math. Soc. **12** (1962), 193–206.

[Cha] K. Chandrasekharan, [1] *Introduction to analytic number theory*, Die Grundlehren der Math. Wiss., Bd. 148, Springer, New York, 1968. [2] *Arithmetic functions*, Die Grundlehren der Math. Wiss., Bd. 167, Springer, New York, 1970.

[Chb] P. L. Chebyshev, *Mémoire sur les nombres premiers*, J. de Math. Pures Appl. (1) **17** (1852), 366–390. Also in Mémoires présentés à l'Académie Impériale des sciences de St.-Pétersbourg par divers savants **7** (1854), 15–33. Also in Oeuvres **1** (1899), 49–70.

[Chn] J. R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157–176.

[Coh] L. W. Cohen, *The annual meeting of the society*, Bull. Amer. Math. Soc. **58** (1952), 159–160.

[Crp] J. G. van der Corput, [1] *Démonstration élémentaire du théorème sur la distribution des nombres premiers*, Scriptum No. 1, Math. Centrum, Amsterdam, 1948. [2] *Sur la reste dans la démonstration élémentaire du théorème des nombres premiers*, Colloq. sur la Théorie des Nombres (Bruxelles, 1955), Thone, Liège, 1956, 163–182.

[Crd] K. Corrádi, *A remark on the theory of multiplicative functions*, Acta Sci. Math. (Szeged) **28** (1967), 83–92.

[Dav] H. Davenport, *Multiplicative number theory*, Markham, Chicago, 1967; 2nd ed. (revised by H. L. Montgomery), Springer, New York, 1980.

[Dia] H. Diamond, [1] *Changes of sign of* $\pi(x) - \mathrm{li}(x)$, L'Enseignement Mathématique **21** (1975), 1–14. [2] *Chebyshev type estimates in prime number theory*, Sem. Théorie des Nombres, Univ. de Bordeaux I, Exp. No. 24, 1973–74, 11 pp.

[DE] H. Diamond and P. Erdös, *On sharp elementary prime number estimates*, L'Enseignement Mathématique **26** (1980), 313–321.

[DM] H. Diamond and K. McCurley, *Constructive elementary estimates for* $M(x)$, Proc. Conf. Number Theory (Temple Univ., 1980), Lecture Notes in Math., vol. 899, Springer-Verlag, Berlin and New York, 1981, pp. 239–253.

[DS] H. Diamond and J. Steinig, *An elementary proof of the prime number theorem with a remainder term*, Invent. Math. **11** (1970), 199–258.

[Dic] L. E. Dickson, *History of the theory of numbers*, Carnegie Inst., Washington, D.C., 1919; reprinted by Chelsea, New York, 1966.

[Dir] G. Lejeune Dirichlet, *Über die Bestimmung der mittleren Werte in der Zahlentheorie*, Abh. Akad. Wiss. Berlin, 1849; 1851, 69–83. Also in Werke, vol. 2, 1897, 49–66.

[Eda] Y. Eda, *On the prime number theorem*, Sci. Rep. Kanazawa Univ. **2** (1953), 23–33.

[Edw] H. M. Edwards, *Riemann's zeta function*, Academic Press, New York, 1974.

[EMF] W. J. Ellison (in collaboration with M. Mendès-France), *Les nombres premiers*, Publ. Inst. Math. Univ. Nancago, No. IX, Hermann, Paris, 1975.

[Erd] P. Erdös, [1] *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. [2] *On the distribution of numbers of the form* $\sigma(n)/n$ *and on some related questions*, Pacific J. Math. **52** (1974), 59–65. [3] Personal communication.

[Euc] Euclid, *Elements*, vol. 2, IX, 20 (T. Heath, ed.), Dover, New York, 1956.

[Eul] L. Euler, *Varias observationes circa series infinitas*, Comment. Acad. Sci. Imp. Petropolitanae **9** (1737; 1744), 160–188. Also in Opera omnia (1) **14**, 216–244.

[Fog] È. Fogels, *On an elementary proof of the prime number theorem*, Latvijas PSR Zin. Akad. Fiz. Mat. Inst. Raksti **2** (1950), 14–45. (Russian)

[FS] W. Forman and H. N. Shapiro, *Abstract prime number theorems*, Comm. Pure Appl. Math. **7** (1954), 587–619.

[Gau] C. F. Gauss, *Letter to Encke*, 24 *Dec*. 1849, Werke, vol. 2, Kng. Ges. Wiss., Göttingen, 1863, pp. 444–447.

[Gel] A. O. Gelfond, [1] *On the arithmetic equivalent of analyticity of the Dirichlet L-series on the line* Re $s = 1$, Izv. Akad. Nauk SSSR Ser. Mat. **20** (1956), 145–166 (Russian); English transl. Amer. Math. Soc. Transl. (2) **19** (1962), 87–108. [2] *Commentary on the papers "On the estimation of the number of primes not exceeding a given value" and "On prime numbers"*, Collected Works of P. L. Chebyshev, vol. 1, Akad. Nauk SSSR, Moscow-Leningrad, 1946, pp. 285–288. (Russian)

[GL] A. O. Gelfond and Ju. V. Linnik, *Elementary methods in the analytic theory of numbers*, Fizmatgiz, Moscow, 1962 (Russian); English transl. (L. J. Mordell, ed.), Rand McNally, Chicago, 1965. Also (I. Snedden, ed.), M.I.T., Cambridge, Mass., 1966.

[Gio] A. Gioia, *The theory of numbers. An introduction*, Markham, Chicago, 1970.

[Had] J. Hadamard, [1] *Étude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann*, J. de Math. Pures Appl. (4) **9** (1893), 171–215; reprinted in *Oeuvres de Jacques Hadamard*, C.N.R.S., Paris, 1968, vol. 1, pp. 103–147. [2] *Sur la distribution des zéros de la fonction* $\zeta(s)$ *et ses conséquences arithmétiques*, Bull. Soc. Math. France **24** (1896), 199–220; reprinted in Oeuvres, vol. 1, pp. 189–210.

[Hls] G. Halász, *Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen*, Acta Math. Acad. Sci. Hungar. **19** (1968), 365–403.

[HaRi] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press, London, 1974.

[HaRo] H. Halberstam and K. Roth, *Sequences*, Oxford, London, 1966.

[Har] G. H. Hardy, *Prime numbers*, Brit. Assn. Rep., 1915, pp. 350–354; also in *Collected papers*, vol. 2, Oxford Univ. Press, London, 1967, pp. 14–18.

[HL] G. H. Hardy and J. E. Littlewood, *Some problems of 'partitio numerorum'. III. On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[HW] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., Oxford, London, 1979.

[HeRi] D. Hensley and I. Richards, [1] *On the incompatibility of two conjectures concerning primes*, Analytic Number Theory, Proc. Sympos. Pure Math., vol. 24, Amer. Math. Soc., Providence, R.I., 1973, pp. 123–127. [2] *Primes in intervals*, Acta Arith. **25** (1974), 375–391.

[Ing] A. E. Ingham, [1] *The distribution of prime numbers*, Cambridge Tracts in Math., No. 30, Cambridge, 1932; reprinted by Hafner, New York, 1971. [2] *Note on the distribution of primes*, Acta Arith. **1** (1936), 201–211. [3] *Review of Selberg and Erdös elementary proofs of P.N.T.*, Math. Reviews **10** (1949), 595–596; Reprinted in Reviews in Number Theory (W. J. LeVeque, ed.), vol. 4, N 20-3, Amer. Math. Soc., Providence, R.I., 1974.

[Iwa] H. Iwaniec, *Rosser's sieve*, Acta Arith. **36** (1980), 171–202.

[Jur] W. B. Jurkat, *Abstracts of short communications*, Proc. Internat. Congr. Math. (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, Sweden, 1963, p. 35.

[Kuh] P. Kuhn, *Eine Verbesserung des Restglieds beim elementaren Beweis des Primzahlsatzes*, Math. Scand. **3** (1955), 75–89.

[laV] C. J. de la Vallée Poussin, [1] *Recherches analytiques sur la théorie des nombres premiers*, Ann. Soc. Sci. Bruxelles **20** (1896), 183–256. [2] *Sur la fonction $\zeta(s)$ de Riemann et le nombre des nombres premiers inférieurs à une limite donnée*, Memoires Couronnés de l'Acad. Roy des Sciences, Belgique **59** (1899–1900); reprinted in Colloque sur la Théorie des Nombres (Bruxelles, 1955), Thone, Liège, 1956, pp. 9–66.

[Lnd] E. Landau, [1] *Handbuch der Lehre von der Verteilung der Primzahlen*, Teubner, Leipzig, 1909; reprinted (with an appendix by P. T. Bateman), Chelsea, New York, 1953. [2] *Über einige neuere Grenzwertsätze*, Rend. Circ. Mat. Palermo **34** (1912), 121–131. [3] *Vorlesungen über Zahlentheorie*, Hirzel, Leipzig, 1927; reprinted by Chelsea, New York, 1947. [4] *Über den Wienerschen neuen Weg zum Primzahlsatz*, Sitzber. Preuss. Akad. Wiss., 1932, pp. 514–521; reprinted in *Handbuch*, 1953, pp. 917–924.

[Lng] M. Langevin, *Méthodes élèmentaires en vue du théorème de Sylvester* (Sem. Delange-Pisot-Poitou, 17e année, 1975/76), Théorie des nombres: Fasc. 2, Exp. No. G2, Secrétariat Math., Paris, 1977.

[LS] A. F. Lavrik and A. Š. Sobirov, *On the remainder term in the elementary proof of the prime number theorem*, Dokl. Akad. Nauk SSSR **211** (1973), 534–536 (Russian); English transl. Soviet Math. Dokl. **14** (1973), 1063–1066.

[Leg] A.-M. Legendre, *Essai sur la théorie des Nombres*, Duprat, Paris, 1798.

[Leh] R. S. Lehman, *On the difference* $\pi(x) - \mathrm{li}(x)$, Acta Arith. **11** (1966), 397–410.

[LF] B. V. Levin and A. S. Feinleib, *Applications of certain integral equations to questions of the theory of numbers*, Uspehi Mat. Nauk **22** (1967), no. 3 (135), 119–197 (Russian); English transl., Russian Math. Surveys **22** (1967), 119–204.

[Lev] N. Levinson, *A motivated account of an elementary proof of the prime number theorem*, Amer. Math. Monthly **76** (1969), 225–245.

[Lit] J. E. Littlewood, *Sur la distribution des nombres premiers*, C. R. Acad. Sci. Paris **158** (1914), 1869–1872.

[Man] H. von Mangoldt, [1] *Zu Riemanns Abhandlung "Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse"*, J. Reine Angew. Math. **114** (1895), 255–305. [2] *Zur Verteilung der Nullstellen der Riemannschen Funktion $\xi(t)$*, Math. Ann. **60** (1905), 1–19.

[Mat] G. B. Mathews, *Theory of numbers*. I, Deighton Bell, Cambridge, England, 1892; reprinted by Stechert, New York, 1927.

[Mei] E. Meissel, *Ueber die Bestimmung der Primzahlmenge innerhalb gegebener Grenzen*, Math. Ann. **2** (1870), 636–642.

[Mer] F. Mertens, *Ein Beitrag zur analytischen Zahlentheorie*, J. Reine Angew. Math. **78** (1874), 46–62.

[Mon] H. L. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Math., vol. 227, Springer-Verlag, Berlin and New York, 1971.

[Nag] T. Nagell, *Introduction to number theory*, Wiley, New York, 1951.

[Nai] M. Nair, *On Chebyshev-type inequalities for primes*, Amer. Math. Monthly **89** (1982), 126–129.

[Nev] V. Nevanlinna, *Über die elementaren Beweise der Primzahlsätze und deren äquivalente Fassungen*, Ann. Acad. Sci. Fenn. Sér. AI, No. 343, 1964.

[Pin] J. Pintz, *Elementary methods in the theory of L-functions*. I–VIII, Acta Arith. **31** (1976), 53–60; **31** (1976), 273–289; **31** (1976), 295–306; **31** (1976), 419–429; **32** (1977), 163–171; **32** (1977), 173–178; **32** (1977), 397–406; Corrigendum **33** (1977), 293–295; **33** (1977), 89–98.

[Pit] H. R. Pitt, *Tauberian theorems*, Oxford, London, 1958.

[Pol] G. Pólya, *Heuristic reasoning in the theory of numbers*, Amer. Math. Monthly **66** (1959), 375–384.

[Pop] J. Popken, *On convolutions in number theory*, Indag. Math. **17** (1955), 10–15.

[PR] A. G. Postnikov and N. P. Romanov, *A simplification of Selberg's elementary proof of the asymptotic law of distribution of primes*, Uspehi. Mat. Nauk **10** 4(66), (1955), 75–87 (Russian); Corrigendum, *ibid.* **24**, no. 5 (149), (1969), 263.

[Pra] K. Prachar, *Primzahlverteilung*, Die Grundlehren der Math. Wiss., Bd. 91, Springer, Berlin, 1957.

[Ric] H.-E. Richert, *Lectures on sieve methods*, Tata Inst., Bombay, 1976.

[Rie] B. Riemann, *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Kgl. Preuss. Akad. Wiss. Berlin, 1860, pp. 671–680; also in Werke, 2nd ed., Teubner, Leipzig, 1892, pp. 145–155. Reprinted by Dover, New York, 1953.

[Ros] P. M. Ross, *On Chen's theorem that each large even integer has the form $p_1 + p_2$ or $p_1 + p_2 p_3$*, J. London Math. Soc. (2) **10** (1975), 500–506.

[RS] J. B. Rosser and L. Schoenfeld, [1] *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94. [2] *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\Psi(x)$*, Math. Comp. **29** (1975), 243–269. [3] *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\Psi(x)$*. II, Math. Comp. **30** (1976), 337–360; Corrigendum, *ibid.* **30** (1976), 900.

[Sch] W. Schwarz, *Einführung in Methoden und Ergebnisse der Primzahltheorie*, Bibliogr. Inst. **278/278a**, Mannheim, 1969.

[Seg] S. L. Segal, *Prime number theorem analogues without primes*, J. Reine Angew. Math. **265** (1974), 1–22.

[Sel] A. Selberg, [1] *An elementary proof of the prime number theorem*, Ann. of Math. (2) **50** (1949), 305–313. [2] *An elementary proof of the prime number theorem for arithmetic progressions*, Canad. J. Math. **2** (1950), 66–78. [3] *On elementary methods in prime number theory and their limitations*, Den 11te Skand. Mat. Kongress (Trondheim, 1949), Tanums Forlag, Oslo, 1952, pp. 13–22.

[Sha] H. N. Shapiro, [1] *On a theorem of Selberg and generalizations*, Ann. of Math. (2) **51** (1950), 485–497. [2] *On primes in arithmetic progression*. II, Ann. of Math. (2) **52** (1952), 231–243.

[Sok] A. V. Sokolovski, Letter to H. Diamond and J. Steinig, 17 February, 1973.

[Spe] W. Specht, *Elementare Beweise der Primzahlsätze*, Hochschulbücher, Bd. 30, VEB Deutcher Verlag Wiss., Berlin, 1956.

[Syl] J. J. Sylvester, [1] *On Tchebycheff's theorem of the totality of prime numbers comprised within given limits*, Amer. J. Math. **4** (1881), 230–247. [2] *On arithmetical series*, Messenger of Math. (2) **21** (1892), 1–19 and 87–120.

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

[Tit] E. C. Titchmarsh, *The theory of the Riemann zeta function*, Oxford, London, 1951.

[ViA] A. I. Vinogradov, *The density hypothesis for Dirichlet L series*, Izv. Akad. Nauk SSSR Ser. Mat. **29** (1965), 903–934. (Russian)

[ViI] I. M. Vinogradov, *The method of trigonometrical sums in the theory of numbers* (Translated, revised, and annotated by K. F. Roth and Anne Davenport), Interscience, London, 1954.

[Wal] A. Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie*, Math. Forschungsber. XV, VEB Deutscher Verlag Wiss., Berlin, 1963.

[Wid] D. V. Widder, *The Laplace transform*, Princeton Univ. Press, Princeton, N.J.. 1941.

[**Wie**] N. Wiener, [1] *A new method in Tauberian theorems*, J. Math. Phys. M.I.T. **7** (1927–28), 161–184. [2] *Tauberian theorems*, Ann. of Math. (2) **33** (1932), 1–100.

[**WG**] N. Wiener and L. Geller, *Some prime number consequences of the Ikehara theorem*, Acta Sci. Math. Szeged **12B** (1950), 25–28.

[**Win**] A. Wintner, *The theory of measure in arithmetical semigroups*, Waverly Press, Baltimore, 1944.

[**Wir**] E. Wirsing, [1] *Elementare Beweise des Primzahlsatzes mit Restglied*. I, J. Reine Angew. Math. **211** (1962), 205–214. [2] *Elementare Beweise des Primzahlsatzes mit Restglied*. II, J. Reine Angew. Math. **214/215** (1963), 1–18. [3] *Des asymptotische Verhalten von Summen über multipikative Funktionen*. II, Acta Math. Acad. Sci. Hungar. **18** (1967), 411–467.

[**Wri**] E. M. Wright, *The elementary proof of the prime number theorem*, Proc. Roy. Soc. Edinburgh. A **63** (1952), 257–267.

[**Zag**] D. Zagier, *Die ersten* 50 *Millionen Primzahlen*, Elemente der Math., Beihefte No. 15, Birkhäuser, Basel, 1977; English transl., Math. Intelligencer **0** (1977), 7–19.

Department of Mathematics, University of Illinois, Urbana, Illinois 61801

Department of Mathematics, The University of Texas, Austin, Texas 78712