SOME TAUBERIAN THEOREMS CONNECTED WITH THE PRIME NUMBER THEOREM

Thesis by

Basil Gordon

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1956

ACKNOWLEDGEMENTS

I would like to express my deep indebtedness to Professor T.M. Apostol for suggesting the problem treated in this thesis, and for providing constant encouragement and guidance to me in carrying out the research leading to its solution. I wish also to thank the California Institute of Technology for its generosity in providing the assistantships and scholarships which I held during the past two years. Finally, thanks are due to Miss Rosemarie Stampfel for her excellent job of typing the manuscript.

ABSTRACT

Let $A(x)$ be a monotone non-decreasing function of x, and let

$$T(x) = \sum_{n=1}^{x} A\left(\frac{x}{n}\right) .$$

It is possible that $T(x) \sim ax \log x$, but

$$\varliminf_{x\to\infty} \frac{A(x)}{x} = 0, \qquad \varlimsup_{x\to\infty} \frac{A(x)}{x} = \infty .$$

If $T(x) = ax \log x + O(x)$, then

$$\varliminf_{x\to\infty} \frac{A(x)}{x} > 0, \qquad \varlimsup_{x\to\infty} \frac{A(x)}{x} < \infty ,$$

but $A(x) \sim ax$ is in general false. If $T(x) = ax \log x + bx + o\left(\frac{x}{\log x}\right)$, then $A(x) \sim ax$.

The prime number theorem is the special case $A(x) = \psi(x)$.

SOME TAUBERIAN THEOREMS CONNECTED WITH THE PRIME NUMBER THEOREM

§ 1. Introductory remarks.

Before beginning a discussion of the results to be proved in this thesis it is convenient to assemble a few standard definitions and notations to be used throughout the sequel.

n will always denote an arbitrary positive integer.

p denotes a prime number.

x denotes any positive real number.

$[x]$ stands for the "integer part" of x, i.e., the greatest integer not exceeding x.

A summation of the form $\sum_{n=1}^{x}$ is to be extended over all integers n not exceeding the real number x; thus

$$\sum_{n=1}^{x} = \sum_{n=1}^{[x]} .$$

$\pi(x)$ stands for the number of primes less than or equal to x.

$\Lambda(n)$ is the function defined by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n \text{ is a prime power, } \quad n = p^{a} \\ 0 & \text{otherwise} \end{cases}$$

$\mu(n)$ is the Mobius μ-function, defined by

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ has a square factor} \\ (-1)^{r} & \text{if } n = p_1\, p_2 \cdots p_r \end{cases}$$

$\psi(x)$ is the function defined by

$$\psi(x) = \sum_{n=1}^{x} \Lambda(n)$$

The symbols O, o, and $\sim$ are used in the classical sense. That is, if $g(x)$ is a positive function of x, then

$$f(x) = O(g(x))$$

means that $\frac{|f(x)|}{g(x)}$ remains bounded as $x \to \infty$,

$$f(x) = o(g(x))$$

means that $\frac{f(x)}{g(x)} \to 0$ as $x \to \infty$, and

$$f(x) \sim g(x)$$

means that $\frac{f(x)}{g(x)} \to 1$ as $x \to \infty$.

It may be remarked that $f(x) \sim g(x)$ is equivalent to $f(x) = g(x) + o(g(x))$.

The prime number theorem was conjectured by Gauss and Legendre as early as 1798, and was proved in 1896 by Hadamard and de la Vallee Poussin. It asserts that

$$\pi(x) \sim \frac{x}{\log x},$$

but is usually proved in the equivalent form

$$\psi(x) \sim x,$$

because the function $\psi(x)$ is more directly related to the classical functions of analysis than is $\pi(x)$. The known proofs of the theorem fall into three general categories according to the nature and degree of sophistication of the methods employed. The first class of proofs includes those of Hadamard and de la Vallee Poussin, which are based on the identity

$$(1) \qquad \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)},$$

valid for complex numbers $s = \sigma + it$ whose real part σ is greater than one. Here $\zeta(s)$ is the Riemann ζ-function:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad (\sigma > 1),$$

and $\zeta'(s)$ its derivative. The identity relates $\Lambda(n)$ and hence indirectly $\psi(x)$ to a function whose properties can be studied by the theory of functions of a complex variable. Using rather delicate estimates of a function-theoretical character, it can be shown from (1) that

$$\psi_1(x) = \int_0^x \psi(t)\,dt \sim \frac{x^2}{2},$$

and the monotonicity of $\psi(x)$ then implies the result

$$\psi(x) \sim x$$

with the aid of an elementary Tauberian theorem.

The second class of proofs includes those of Wiener and Ikehara, which are again based on the identity (1). This time, however, the only function-theoretical property used is the non-vanishing of $\zeta(s)$ on the line $\sigma = 1$. The main burden of the argument rests on a Tauberian theorem of far greater complexity than the one already referred to in connection with the first category of proofs. Thus a precarious balance seems to exist between the roles of function theory and Tauberian analysis, so that the easing of one carries with it a corresponding strengthening of the other.

The third and most recent group of proofs is typified by those of Selberg and Erdös. These proofs are completely elementary in the technical sense that the theory of functions of a complex variable is eliminated altogether. It might be conjectured in view of the preceding discussion that the effect of this would be to make the Tauberian element completely dominant, but at first there seemed to be no Tauberian arguments whatsoever in the Selberg-Erdös proof. It will be shown in this thesis that, first appearances to the contrary, the whole proof is purely Tauberian.

This is accomplished by studying instead of (1) another formula connecting $\psi(x)$ to standard functions of analysis, namely

$$(2) \qquad \sum_{n=1}^{x} \psi\left(\frac{x}{n}\right) = \log [x]!$$

(for a proof see for example [1], p. 76).

It is interesting to note that this relation was used by Chebyshev in his celebrated memoir on the prime number theorem in order to prove the existence of positive constants c_1 and c_2 such that

$$c_1 x < \psi(x) < c_2 x .$$

Thus he showed that $\psi(x)$ is of the order of magnitude of x, but failed to establish the existence of

$$\lim_{x \to \infty} \frac{\psi(x)}{x} ,$$

which would be needed for the proof of the prime number theorem. Accordingly the identity (2) was abandoned by later investigators in favor of (1) which, as has already been remarked, ultimately formed the basis of the first proof. In fact Landau stated in his Handbuch ([1] p. 597) that Chebyshev's failure was due to the intrinsic impossibility of deriving the prime number theorem from (2). This assertion is, however, incorrect, as will be seen in what follows.

The problem, then, is to start with (2), and conclude, using only elementary methods, that $\psi(x) \sim x$. This is clearly a special case of the following question: If $A(x)$ is some function of x, and if information is known about the function

$$T(x) = \sum_{n=1}^{x} A\left(\frac{x}{n}\right) ,$$

then what can be said about $A(x)$ itself? In particular what hypothesis on $T(x)$ is needed in order to conclude that

$$A(x) \sim ax \qquad (a \text{ a constant})?$$

A necessary condition on $T(x)$ would certainly be

$$T(x) \sim ax \log x .$$

For if $A(x) \sim ax$, then

$$T(x) = \sum_{n=1}^{\frac{x}{\log x}} A\left(\frac{x}{n}\right) + \sum_{n=\frac{x}{\log x}}^{x} A\left(\frac{x}{n}\right) .$$

When n is in the range between 1 and $\frac{x}{\log x}$, $\frac{x}{n} \to \infty$ as $x \to \infty$, and therefore $A\left(\frac{x}{n}\right) \sim a\,\frac{x}{n}$ may be applied there to give

$$T(x) = \sum_{n=1}^{\frac{x}{\log x}} \left[\frac{ax}{n} + o\left(\frac{x}{n}\right)\right] + \sum_{n=\frac{x}{\log x}}^{x} O\left(\frac{x}{n}\right)$$

$$= ax \log\left(\frac{x}{\log x}\right) + o\left(x \log \frac{x}{\log x}\right) + O\left(x \log x - x \log \frac{x}{\log x}\right)$$

$$= ax \log x + o(x \log x) + O(x \log \log x)$$

$$\sim ax \log x \,.$$

An examination of the familiar Tauberian theorems of analysis might lead one to suspect that conversely the condition

(3) $\quad T(x) \sim ax \log x$

plus a Tauberian condition such as the monotonicity of $A(x)$ should be sufficient to guarantee that $A(x) \sim ax$. It turns out, however, that the situation is considerably more complicated than this. In fact an example is given in § 2 in which (3) is satisfied (with $a > 0$) and $A(x)$ is monotone non-decreasing, but where

$$0 = \varliminf_{x\to\infty} \frac{A(x)}{x} < \varlimsup_{x\to\infty} \frac{A(x)}{x} = \infty \,.$$

Thus it can be seen that something stronger than (3) is needed.

In 1950, H.N. Shapiro [3] proved that if

(4) $\quad T(x) = ax \log x + O(x),$

and if $A(x)$ is monotone non-decreasing then the Chebyshev estimates

$$c_1 x < A(x) < c_2 x$$

(where c_1 and c_2 are positive constants) hold for all large x. Actually Shapiro considered only the special case where $A(x)$ is a step function, but his method works in general, and this extension will be made in § 3, as it is needed later on.

Thus the strengthening of (3) to (4) eliminates the extreme pathology mentioned above, in which

$$\varliminf_{x\to\infty} \frac{A(x)}{x} = 0, \qquad \varlimsup_{x\to\infty} \frac{A(x)}{x} = \infty.$$

It may be interpreted as a weak Tauberian theorem in the desired direction. But unfortunately Shapiro's condition (4) still does not imply $A(x) \sim ax$, except in the case $a = 0$. The proof of this is given in § 4.

One must therefore look for an even stronger hypothesis on $T(x)$. Such a condition was given by Landau [2], who proved ([1], pp.599-604) that if $A(x)$ is monotone, and if

$$T(x) = ax \log x + bx + O(\omega(x)) \qquad (b \text{ a constant}) \tag{5}$$

with $\omega(x)$ monotone increasing and such that

$$\int_1^\infty \frac{\omega(x)}{x^2}\, dx < \infty,$$

then $A(x) \sim ax$.

Later Ingham [4] relaxed this to the requirement that $A(x)$ be monotone, and

$$T(x) = ax \log x + bx + o(x). \tag{6}$$

Unfortunately the results of Landau and Ingham are unsuitable for the present prupose of deriving the prime number theorem in an elementary manner from (2), because their proofs rely on machinery of greater depth than the prime number theorem itself, including the theory of functions of a complex variable.

Accordingly the problem is tackled afresh in § 5, where it is shown by elementary means that if $A(x)$ is monotone, and if

$$(7) \qquad T(x) = ax \log x + bx + o\left(\frac{x}{\log x}\right),$$

then $A(x) \sim ax$.

This condition is intermediate between (5) and (6), so that the resulting theorem is better than Landau's, but not as good as Ingham's. However, it suffices for all the cases commonly encountered, including the case of the prime number theorem, where $A(x) = \psi(x)$. For then the corresponding $T(x)$, by identity (2), is $\log [x]!$, and using Stirling's formula, it is seen that

$$T(x) = x \log x - x + O(\log x).$$

Thus (7) is easily satisfied with $a = 1$, $b = -1$, and since $\psi(x)$ is monotone, the desired result $\psi(x) \sim x$ follows. This accomplishes the Tauberianization of the Selberg-Erdös proof, and also brings out its relation to the work of Chebyshev.

§ 2. A pathological example.

In this section an example will be given of a monotone non-decreasing function $A(x)$ such that

$$T(x) = \sum_{n=1}^{x} A(\tfrac{x}{n}) \sim ax \log x \qquad (a > 0),$$

but where

$$\underline{\lim}_{x\to\infty} \frac{A(x)}{x} = 0$$

and

$$\overline{\lim}_{x\to\infty} \frac{A(x)}{x} = \infty .$$

To construct such an example, consider a sequence of points x_n such that $x_1 < x_2 < x_3 < \cdots \to \infty$, and an increasing sequence of positive real numbers $c_1 < c_2 < \cdots$. Define $A(x)$ to be 0 if $x < x_1$, and if

$$x_m \leq x < x_{m+1} ,$$

put

$$A(x) = c_m .$$

Thus $A(x)$ is a step function which jumps at the points x_k, and whose values are the c_k. The corresponding function $T(x)$ may then be handled as follows. By definition,

$$A(\tfrac{x}{n}) = c_k \qquad \text{if} \qquad x_k \leq \frac{x}{n} < x_{k+1} .$$

The condition $x_k \leq \frac{x}{n} < x_{k+1}$ is equivalent to $\frac{x}{x_{k+1}} < n \leq \frac{x}{x_k}$, and is therefore fulfilled by $[\frac{x}{x_k}] - [\frac{x}{x_{k+1}}]$ different integers n. Hence

$$T(x) = \sum_{n=1}^{x} A(\frac{x}{n}) = ([\frac{x}{x_1}] - [\frac{x}{x_2}])\, c_1 + ([\frac{x}{x_2}] - [\frac{x}{x_3}])\, c_3 + \cdots + [\frac{x}{x_m}]\, c_m$$

(the series naturally breaks off because $x < x_{m+1}$). Rearranging,

$$T(x) = [\frac{x}{x_1}]\, c_1 + [\frac{x}{x_2}]\, (c_2 - c_1) + \cdots + [\frac{x}{x_m}]\, (c_m - c_{m-1}) \,.$$

Since $[\theta] \leq \theta$, $T(x)$ can be estimated from above by writing

$$\begin{aligned} T(x) &\leq \frac{x}{x_1} c_1 + \frac{x}{x_2}(c_2 - c_1) + \cdots + \frac{x}{x_m}(c_m - c_{m-1}) \\ &= x(\frac{c_1}{x_1} + \frac{c_2 - c_1}{x_2} + \cdots + \frac{c_m - c_{m-1}}{x_m}) \end{aligned} \tag{8}$$

On the other hand, a lower estimate is gotten by using $[\theta] > \theta - 1$. Thus

$$\begin{aligned} T(x) &> (\frac{x}{x_1} - 1)\, c_1 + (\frac{x}{x_2} - 1)\, (c_2 - c_1) + \cdots + (\frac{x}{x_m} - 1)\, (c_m - c_{m-1}) \\ &= x(\frac{c_1}{x_1} + \frac{c_2 - c_1}{x_2} + \cdots + \frac{c_m - c_{m-1}}{x_m}) - c_1 - c_2 + c_1 - \cdots - c_m + c_{m-1} \\ &= x(\frac{c_1}{x_1} + \frac{c_2 - c_1}{x_2} + \cdots + \frac{c_m - c_{m-1}}{x_m}) - c_m \,. \end{aligned} \tag{9}$$

The inequalities (8) and (9) are applicable to any step function $A(x)$ of the type under consideration, and will prove useful later on.

Now choose $x_k = 2^{k^2}$, and $c_k = k \cdot 2^{k^2}$. Then the upper estimate (8) becomes

$$T(x) \leq x \left\{ 1 + \sum_{k=2}^{m} \frac{k \cdot 2^{k^2} - (k-1)\, 2^{(k-1)^2}}{2^{k^2}} \right\}$$

$$= x \left\{ 1 + \sum_{k=2}^{m} k - \sum_{k=2}^{m} \frac{k-1}{2^{2k-1}} \right\} = x \left\{ \frac{m(m+1)}{2} + O(1) \right\}$$

(since the series $\sum_{k=2}^{\infty} \frac{k-1}{2^{2k-1}}$ converges). But m was determined so that

$$2^{m^2} \leq x < 2^{(m+1)^2}$$

Hence $m^2 \leq \frac{\log x}{\log 2}$, and so

$$(10) \qquad T(x) \leq x \quad \frac{\log x}{\log 2} + O(\sqrt{\log x}) \quad .$$

Similarly, the lower estimate (9) becomes

$$T(x) > x \left\{ 1 + \sum_{k=2}^{m} k - \sum_{k=2}^{m} \frac{k-1}{2^{2k-1}} \right\} - 2^{m^2}$$

$$= x \left\{ \frac{m(m+1)}{2} + O(1) \right\} - 2^{m^2} .$$

Again using the fact that

$$2^{m^2} \leq x < 2^{(m+1)^2}$$

it follows that $(m+1)^2 > \frac{\log x}{\log 2}$, or $m > \sqrt{\frac{\log x}{\log 2}} - 1$, and so

$$(11) \qquad T(x) > x \left\{ \frac{\log x}{\log 2} + O(\sqrt{\log x}) \right\} - x \quad .$$

Inequalities (10) and (11) together imply

$$T(x) = \frac{x \log x}{\log 2} + O(x \sqrt{\log x}) .$$

On the other hand $\underline{\lim}_{x\to\infty} \frac{A(x)}{x}$ and $\overline{\lim}_{x\to\infty} \frac{A(x)}{x}$ can be calculated as follows. At the points $x = 2^{m^2}$,

$$\frac{A(x)}{x} = \frac{m \cdot 2^{m^2}}{2^{m^2}} = m$$

which is unbounded. Hence

$$\overline{\lim_{x\to\infty}} \frac{A(x)}{x} = \infty .$$

But immediately to the left of these points, say at the points $x = 2^{m^2} - \varepsilon$,

$$\frac{A(x)}{x} = \frac{(m-1)\ 2^{(m-1)^2}}{2^{m^2} - \varepsilon} \sim \frac{m-1}{2^{2m-1}} ,$$

which becomes arbitrarily small. Thus

$$\underline{\lim}_{x\to\infty} \frac{A(x)}{x} = 0 .$$

This completes the construction of the proposed example in the case $a = \frac{1}{\log 2}$. Of course a similar example can now be constructed for any $a > 0$ by multiplying this particular one by $a \log 2$.

It is worth noting that the error $O(x \sqrt{\log x})$ is better than the bound $o(x \log x)$, which is all that would be required. As a matter of fact even sharper bounds can be obtained by a modification of the above construction. It is not difficult to see that by choosing

$$x_n = 2^{n^q}$$

$$c_n = n^{q-1}\ 2^{n^q} \qquad (q > 1)$$

a function $A(x)$ with

$$\underline{\lim}_{x\to\infty} \frac{A(x)}{x} = 0, \qquad \overline{\lim}_{x\to\infty} \frac{A(x)}{x} = \infty$$

$$T(x) = \frac{x \log x}{\log 2} + O(x(\log x)^{q-1/q})$$

will result. By taking q arbitrarily close to 1, examples can therefore be created with

$$T(x) = ax \log x + O(x(\log x)^{\varepsilon})$$

It will be seen in the next section that this pathology is eliminated by imposing Shapiro's condition

$$T(x) = ax \log x + O(x),$$

and the above result makes the necessity of some such condition obvious.

§ 3. The Shapiro condition.

In this section the results of H.N. Shapiro [3] are extended from the case of a step function to that of an arbitrary non-decreasing function $A(x)$. Only the case $a > 0$ is considered here, since in § 4 a stronger result will be proved for $a = 0$.

THEOREM 1. If $A(x)$ is monotone non-decreasing, and if

$$T(x) = \sum_{n=1}^{x} A\left(\frac{x}{n}\right) = ax \log x + O(x),$$

then there is a (positive) constant c_2 such that

$$A(x) < c_2 x \quad .$$

<u>Proof</u>. Assume without loss of generality that $A(x) = 0$ for $x < 1$ (this affects neither the hypothesis nor the conclusion). Using Stieltjes integration,

$$T(x) = \sum_{n=1}^{x} A\left(\frac{x}{n}\right) = \int_1^x A\left(\frac{x}{t}\right) d[t] \quad .$$

Integrating by parts, the integrated terms vanish, and the result is

$$T(x) = - \int_1^x [t] \, dA\left(\frac{x}{t}\right) \quad .$$

Making the change of variables $t = \frac{x}{u}$, this becomes

$$T(x) = \int_1^x \left[\frac{x}{u}\right] dA(u)$$

Hence

$$T(x) - 2T(\tfrac{x}{2}) = \int_1^x [\tfrac{x}{u}]\, dA(u) - 2 \int_1^{x/2} [\tfrac{x}{2u}]\, dA(u)$$

$$= \int_1^{x/2} ([\tfrac{x}{u}] - 2[\tfrac{x}{2u}])\, dA(u) + \int_{x/2}^x [\tfrac{x}{u}]\, dA(u) .$$

Since $[\theta] - 2[\frac{\theta}{2}] \geq 0$, and since $A(u)$ is non-decreasing, it follows that

$$T(x) - 2T(\tfrac{x}{2}) \geq \int_{x/2}^x [\tfrac{x}{u}]\, dA(u)$$

$$= \int_{x/2}^x dA(u) = A(x) - A(\tfrac{x}{2}) .$$

But by hypothesis,

$$T(x) - 2T(\tfrac{x}{2}) = x \log x + O(x) - \tfrac{2x}{2} \log \tfrac{x}{2} + O(x) = O(x) .$$

Hence $A(x) - A(\frac{x}{2}) < Kx$ for some constant K. This implies that

$$A(\tfrac{x}{2}) - A(\tfrac{x}{4}) < K\, \tfrac{x}{2}$$

$$A(\tfrac{x}{4}) - A(\tfrac{x}{8}) < K\, \tfrac{x}{4}\ , \text{ etc.}$$

By summation,

$$A(x) < 2Kx = c_2 x \qquad \text{q.e.d.}$$

THEOREM 2. Under the hypothesis of Theorem 1,

$$\int_1^x \frac{dA(t)}{t} = a \log x + O(1)$$

Proof. As in the previous theorem,

$$T(x) = \int_1^x [\tfrac{x}{t}]\, dA(t) .$$

Using the fact that $[\frac{x}{t}] = \frac{x}{t} + O(1)$, this becomes

$$T(x) = x \int_1^x \frac{dA(t)}{t} + \int_1^x O(1)\, dA(t)$$

By the monotone character of $A(t)$, the second term on the right is

$$O\left(\int_1^x dA(t)\right) = O(A(x)) = O(x),$$

using theorem 1 for the last step. Hence

$$T(x) = x \int_1^x \frac{dA(t)}{t} + O(x).$$

But by hypothesis

$$T(x) = ax \log x + O(x) .$$

Comparing the two expressions for $T(x)$, it is seen that

$$x \int_1^x \frac{dA(t)}{t} = ax \log x + O(x),$$

or

$$\int_1^x \frac{dA(t)}{t} = a \log x + O(1) .$$

THEOREM 3. Again let the hypothesis be the same as in theorem 1. Then there is a positive constant c_1 such that

$$A(x) > c_1 x .$$

<u>Proof</u>: Write

$$\int_1^x \frac{dA(t)}{t} = a \log x + R(x) .$$

By theorem 2, there is some constant M such that

$$|R(x)| < M .$$

Choose $c_1 = e^{-\frac{2M+1}{a}}$. Then

$$\int_{c_1x}^{x} \frac{dA(t)}{t} = \int_{1}^{x} \frac{dA(t)}{t} - \int_{1}^{c_1x} \frac{dA(t)}{t}$$

$$= a \log x + R(x) - a \log c_1x - R(c_1x)$$

$$> - a \log c_1 - 2M = 1 .$$

Hence

$$\frac{A(x)}{c_1x} = \frac{1}{c_1x} \int_{1}^{x} dA(t) \geq \frac{1}{c_1x} \int_{c_1x}^{x} dA(t)$$

$$\geq \int_{c_1x}^{x} \frac{dA(t)}{t} > 1 .$$

Thus

$$A(x) > c_1x \qquad\qquad \text{q.e.d.}$$

Combining theorems 1 and 3, there result the Chebyshev inequalities

$$c_1x < A(x) < c_2x .$$

§ 4. Limitations of the Shapiro condition.

It has been shown that if

$$T(x) = ax \log x + O(x),$$

then inequalities of the form

$$c_1 x < A(x) < c_2 x$$

hold with c_1 and c_2 positive, so that $A(x)$ is of the order of magnitude of x. The question of whether or not

$$A(x) \sim ax$$

can be asserted is dealt with in this section, and is answered in the affirmative for $a = 0$, but in the negative for $a > 0$.

THEOREM 4. Let $A(x)$ be a non-decreasing function of x, and suppose the Shapiro condition is satisfied with $a = 0$, i.e.,

$$T(x) = O(x)$$

Then $A(x) \sim ax$, i.e., $A(x) = o(x)$.

Proof. The proof is by contradiction. Suppose that $A(x) \neq o(x)$. Then there is some $\varepsilon > 0$ such that

$$\frac{A(x)}{x} \geq \varepsilon \qquad \text{infinitely often} . \tag{12}$$

Suppose that this inequality occurs at a sequence of points $x_1 < x_2 < x_3 < \cdots \to \infty$. Now let $B(x)$ be the step function defined by

$$B(x) = \begin{cases} A(0) & \text{if} \quad x < x_1 \\ \varepsilon x_m & \text{if} \quad x_m \leq x < x_{m+1} \end{cases}$$

Then inequality (12) asserts that at the points x_m,

$$A(x) \geq B(x).$$

By the monotonicity of $A(x)$, and the fact that $B(x)$ is a step function, it follows that

$$A(x) \geq B(x) \qquad \text{for all } x.$$

Hence also

$$T(x) = \sum_{n=1}^{x} A(\tfrac{x}{n}) \geq \sum_{n=1}^{x} B(\tfrac{x}{n}) .$$

The latter sum can be estimated by the general inequality (9) which was developed in § 2 for step functions. The steps c_m are given by

$$c_m = \varepsilon x_m ,$$

and hence (9) yields

$$\sum_{n=1}^{x} B(\tfrac{x}{n}) \geq \varepsilon x \left(\frac{x_1}{x_1} + \frac{x_2 - x_1}{x_2} + \cdots + \frac{x_m - x_{m-1}}{x_m} \right) - \varepsilon x_m$$

$$= \varepsilon x \left(1 - \frac{x_1}{x_2} + 1 - \frac{x_2}{x_3} + \cdots + 1 - \frac{x_{m-1}}{x_m} \right) + \varepsilon(x - x_m)$$

$$\geq \varepsilon x \sum_{k=1}^{m-1} \left(1 - \frac{x_k}{x_{k+1}} \right) .$$

The proof can be completed by showing that

$$\sum_{k=1}^{\infty} \left(1 - \frac{x_k}{x_{k+1}} \right) = \infty ,$$

because the hypothesis $T(x) = O(x)$ is then violated, giving a contradiction. But an infinite series of the form

$$\sum_{k=1}^{\infty} c_k$$

with $c_k > 0$ diverges if and only if the infinite product

$$\prod_{k=1}^{\infty} (1 - c_k)$$

diverges to 0. In this case the corresponding product is

$$\prod_{k=1}^{\infty} \frac{x_k}{x_{k+1}},$$

and its partial products telescope to $\frac{x_1}{x_{N+1}}$, which tends to 0 as $N \to \infty$. q.e.d.

THEOREM 5. For any $a > 0$, there is a monotone non-decreasing function $A(x)$ such that

$$T(x) = \sum_{n=1}^{x} A(\tfrac{x}{n}) = ax \log x + O(x),$$

but

$$\varliminf_{x\to\infty} \frac{A(x)}{x} < \varlimsup_{x\to\infty} \frac{A(x)}{x} .$$

<u>Proof</u>. Choose a fixed number $c > 1$, and define the function $A(x)$ by

$$A(x) = \begin{cases} c^m & \text{if} \quad c^m \leq x < c^{m+1} \\ \\ 0 & \text{if} \quad x < c . \end{cases}$$

$A(x)$ is a step function whose jumps occur at the points $x_m = c^m$, and whose steps are of heights $c_m = c^m$. Hence inequalities (8) and (9) of § 2 may be applied in order to estimate $T(x)$ from above and below.

From (8) it follows that

$$T(x) \leq x \left(\frac{c}{c} + \frac{c^2 - c}{c^2} + \cdots + \frac{c^m - c^{m-1}}{c^m} \right)$$

$$= x + (c - 1)\, x \left(\frac{c}{c^2} + \frac{c^2}{c^3} + \cdots + \frac{c^{m-1}}{c^m} \right)$$

$$= x + m(1 - \frac{1}{c})\, x \, .$$

Since $c^m \leq x < c^{m+1}$, m is at most $\frac{\log x}{\log c}$. Hence

$$(13) \qquad T(x) \leq x + \frac{x \log x}{\log c} \;(1 - \frac{1}{c}) \, .$$

On the other hand (9) asserts that

$$T(x) \geq x \left(\frac{c}{c} + \frac{c^2 - c}{c^2} + \cdots + \frac{c^m - c^{m-1}}{c^m} \right) - c^m$$

$$= x + m(1 - \frac{1}{c})\, x - c^m \, .$$

Again using $c^m \leq x < c^{m+1}$, it follows that

$$(14) \qquad T(x) \geq x + (\frac{\log x}{\log c} - 1)\;(1 - \frac{1}{c})\, x - x \, .$$

Combining inequalities (13) and (14), the result is

$$(15) \qquad T(x) = \frac{1 - \frac{1}{c}}{\log c}\; x \log x + O(x) \, .$$

Thus the Shapiro condition is satisfied with $a = \dfrac{1 - \frac{1}{c}}{\log c}$. But it is immediately seen from the definition of $A(x)$ that

$$\frac{1}{c} = \underline{\lim}_{x\to\infty} \frac{A(x)}{x} < \overline{\lim}_{x\to\infty} \frac{A(x)}{x} = 1 \ . \tag{16}$$

To get an example for any $a > 0$, multiply this particular example by

$$a \frac{\log c}{1 - \frac{1}{c}}$$

Then equations (15) and (16) become

$$T(x) = ax \log x + O(x)$$

$$\underline{\lim}_{x\to\infty} \frac{A(x)}{x} = \frac{a \log c}{c - 1}$$

$$\overline{\lim}_{x\to\infty} \frac{A(x)}{x} = \frac{a \log c}{1 - \frac{1}{c}}$$

This shows somewhat more, for by choosing c large, examples are obtained with $\underline{\lim}_{x\to\infty} \frac{A(x)}{x}$ arbitrarily small and $\overline{\lim}_{x\to\infty} \frac{A(x)}{x}$ arbitrarily large. q.e.d.

§ 5. A sufficient condition.

In this section a sufficient condition for concluding $A(x) \sim ax$ will finally be established, namely

$$T(x) = ax \log x + bx + o\left(\frac{x}{\log x}\right),$$

where b is a constant.

LEMMA 1. If $F(x)$ is any function defined for $x \geq 1$, and if

$$G(x) = \log x \sum_{n=1}^{x} F\left(\frac{x}{n}\right),$$

then

$$F(x) \log x + \sum_{n=1}^{x} F\left(\frac{x}{n}\right) \Lambda(n) = \sum_{n=1}^{x} \mu(n)\, G\left(\frac{x}{n}\right), \tag{16a}$$

<u>Proof.</u> Making use of the identities

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d}$$

and

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if} \quad n = 1 \\ 0 & \text{if} \quad n > 1, \end{cases}$$

the left hand side of (16a) may be written in the form

$$\sum_{n=1}^{x} F\left(\frac{x}{n}\right) \log \frac{x}{n} \sum_{d|n} \mu(d) + \sum_{n=1}^{x} F\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{n}{d} \quad .$$

Since $\log \frac{x}{n} + \log \frac{n}{d} = \log \frac{x}{d}$, this becomes

$$\sum_{n=1}^{x} F(\frac{x}{n}) \sum_{d|n} \mu(d) \log \frac{x}{d}$$

Now set $n = md$, and the double series can be rearranged to give

$$\sum_{d=1}^{x} \mu(d) \log \frac{x}{d} \sum_{m=1}^{\frac{x}{d}} F(\frac{x}{md}) = \sum_{d=1}^{x} \mu(d)\ G(\frac{x}{d})$$

by the definition of G. q.e.d.

THEOREM 6. Let $A(x)$ be any function of the positive real variable x, and suppose that

$$T(x) = \sum_{n=1}^{x} A(\frac{x}{n}) = ax \log x + bx + o(\frac{x}{\log x}) ,$$

Then

$$A(x) \log x + \sum_{n=1}^{x} A(\frac{x}{n})\ A(n) = 2ax \log x + o(x \log x).$$

Proof. Apply lemma 1 to the function

$$F_1(x) = A(x) .$$

The corresponding G is

$$G_1(x) = \log x \sum_{n=1}^{x} A(\frac{x}{n}) = T(x) \log x$$

$$= ax(\log x)^2 + bx \log x + o(x),$$

using the hypothesis on $T(x)$. By lemma 1,

$$A(x) \log x + \sum_{n=1}^{x} A(\frac{x}{n}) A(n)$$

$$= a \sum_{n=1}^{x} \mu(n) \frac{x}{n} (\log \frac{x}{n})^2 + \sum_{n=1}^{x} \mu(n)\ b \frac{x}{n} \log \frac{x}{n} + \sum_{n=1}^{x} \mu(n)\ o(\frac{x}{n})$$

Now $\sum_{n=1}^{x} \mu(n)\, o(\frac{x}{n}) = o(x \log x)$. Because

$$\sum_{n=1}^{x} \mu(n)\, o(\tfrac{x}{n}) = \sum_{n=1}^{\frac{x}{\log x}} \mu(n)\, o(\tfrac{x}{n}) + \sum_{n=\frac{x}{\log x}}^{x} \mu(n)\, O(\tfrac{x}{n}) .$$

Since $|\mu(n)| \leq 1$, the first term is

$$o\left(\sum_{n=1}^{\frac{x}{\log x}} \frac{x}{n} \right) = o(x \log x),$$

and the second term is

$$O\left(\sum_{n=\frac{x}{\log x}}^{x} \frac{x}{n} \right) = O(x \log x - x \log \frac{x}{\log x})$$

$$= O(x \log \log x) = o(x \log x) .$$

Therefore

$$(17) \quad A(x) \log x + \sum_{n=1}^{x} A(\tfrac{x}{n}) \;\; \Lambda(n)$$

$$= a \sum_{n=1}^{x} \mu(n)\, \frac{x}{n} (\log \frac{x}{n})^2 + b \sum_{n=1}^{x} \mu(n)\, \frac{x}{n} \log \frac{x}{n} + o(x \log x).$$

Next apply lemma 1 to the function

$$F_2(x) = ax + b - \gamma ,$$

where γ is Euler's constant. The corresponding G is

$$G_2(x) = \log x \sum_{n=1}^{x} (a\frac{x}{n} + b - \gamma) = \log x(ax \log x + \gamma x + O(1) + bx - \gamma x)$$

$$= ax(\log x)^2 + bx \log x + O(\log x).$$

Hence by lemma 1,

$$(ax + b - \gamma) \log x + \sum_{n=1}^{x} (\frac{ax}{n} + b - \gamma) \Lambda(n)$$

$$= \sum_{n=1}^{x} \mu(n) \frac{ax}{n} (\log \frac{x}{n})^2 + \sum_{n=1}^{x} \mu(n)\, b \frac{x}{n} \log \frac{x}{n} + \sum_{n=1}^{x} \mu(n)\, O(\log \frac{x}{n})$$

Again the last term is $o(x \log x)$. Because $|\mu(n)| \leq 1$, and

$$O(\sum_{n=1}^{x} \log \frac{x}{n}) = O(\int_1^x \log \frac{x}{t}\, dt) .$$

Making the change of variables $t = \frac{x}{u}$, this becomes

$$O(\int_1^x \frac{x \log u\, du}{u^2}) = O(x) = o(x \log x)$$

Hence

$$(ax + b - \gamma) \log x + \sum_{n=1}^{x} (\frac{ax}{n} + b - \gamma) \Lambda(n)$$

(18)
$$= a \sum_{n=1}^{x} \mu(n) \frac{x}{n} (\log \frac{x}{n})^2 + b \sum_{n=1}^{x} \mu(n) \frac{x}{n} \log \frac{x}{n} + o(x \log x).$$

In equations (17) and (18) the right hand sides are identical. Therefore the left hand sides differ by $o(x \log x)$, i.e.,

$$A(x) \log x + \sum_{n=1}^{x} A(\frac{x}{n}) \Lambda(n) = (ax + b - \gamma) \log x$$

$$+ \sum_{n=1}^{x} (\frac{ax}{n} + b - \gamma) \Lambda(n) + o(x \log x).$$

But
$$\sum_{n=1}^{x} \Lambda(n) = O(x) = o(x \log x),$$
and
$$\sum_{n=1}^{x} \frac{\Lambda(n)}{n} = \log x + O(1),$$

as is seen by applying theorems 1 and 2 to the special case $A(x) = \psi(x)$. Hence

$$(19) \quad A(x) \log x + \sum_{n=1}^{x} A(\tfrac{x}{n}) \Lambda(n) = 2ax \log x + o(x \log x) \qquad \text{q.e.d.}$$

Formula (19) is very suggestive of Selberg's identity, which is the key to the elementary proof of the prime number theorem. In fact, it would be possible to proceed from here along much the same lines as those of Selberg himself. However, a more direct line of attack will be taken.

LEMMA 2. Let $A(x)$ be a function of x such that $A(x) = O(x)$. Then
$$\sum_{n=1}^{x} A(\tfrac{x}{n}) \Lambda(n) = \int_1^{\sqrt{x}} A(\tfrac{x}{t})\, d\psi(t) + \int_1^{\sqrt{x}} \psi(\tfrac{x}{t})\, dA(t) + O(x).$$

<u>Proof.</u> $\displaystyle\sum_{n=1}^{x} A(\tfrac{x}{n}) \Lambda(n) = \int_1^{x} A(\tfrac{x}{t})\, d\psi(t)$

$$= \int_1^{\sqrt{x}} A(\tfrac{x}{t})\, d\psi(t) + \int_{\sqrt{x}}^{x} A(\tfrac{x}{t})\, d\psi(t).$$

In the second integral, make the change of variables $u = \frac{x}{t}$. Then it becomes
$$- \int_1^{\sqrt{x}} A(u)\, d\psi(\tfrac{x}{u}) .$$

Integrating by parts, this can be written as

$$-A(\sqrt{x})\,\psi(\sqrt{x}) + \int_1^{\sqrt{x}} \psi\left(\frac{x}{u}\right) dA(u) = \int_1^{\sqrt{x}} \psi\left(\frac{x}{t}\right) dA(t) + O(x) .$$

It is now convenient to introduce the following notation. Let

$$\alpha = \underline{\lim}_{x\to\infty} \frac{A(x)}{x} \qquad \beta = \overline{\lim}_{x\to\infty} \frac{A(x)}{x}$$

$$\gamma = \underline{\lim}_{x\to\infty} \frac{\psi(x)}{x} \qquad \delta = \overline{\lim}_{x\to\infty} \frac{\psi(x)}{x}$$

THEOREM 7. If $A(x)$ is a monotone non-decreasing function of x such that

$$T(x) = \sum_{n=1}^{x} A\left(\frac{x}{n}\right) = ax \log x + bx + o\left(\frac{x}{\log x}\right),$$

then

$$a\gamma \le \alpha \le \beta \le a\delta .$$

<u>Proof</u>. By theorem 6,

$$(20) \qquad A(x) \log x + \sum_{n=1}^{x} A\left(\frac{x}{n}\right) \Lambda(n) = 2ax \log x + o(x \log x).$$

By theorem 1, $A(x) = O(x)$, and hence lemma 2 can be applied to the second term of (20), giving

$$A(x) \log x + \int_1^{\sqrt{x}} A\left(\frac{x}{t}\right) d\psi(t) + \int_1^{\sqrt{x}} \psi\left(\frac{x}{t}\right) dA(t) = 2ax \log x + o(x\log x)$$

Dividing by $x \log x$, this becomes

$$(21) \qquad \frac{A(x)}{x} + \frac{1}{x \log x} \int_1^{\sqrt{x}} A(\tfrac{x}{t})\, d\psi(t) + \frac{1}{x \log x} \int_1^{\sqrt{x}} \psi(\tfrac{x}{t})\, dA(t)$$

$$= 2a + o(1) .$$

Now let $x \to \infty$ in such a way that

$$\frac{A(x)}{x} \to \alpha .$$

Then the right hand side of (21) tends to $2a$, and the left hand side can be bounded as follows. For any $\varepsilon > 0$,

$$(22) \qquad A(\tfrac{x}{t}) < (\beta + \varepsilon) \frac{x}{t}$$

$$\psi(\tfrac{x}{t}) < (\delta + \varepsilon) \frac{x}{t}$$

for sufficiently large $\frac{x}{t}$. In (21) t runs from 1 to $\sqrt{x}$, and so $\frac{x}{t}$ runs from $\sqrt{x}$ to x. Hence as $x \to \infty$, inequalities (22) may be applied, giving

$$\frac{1}{x \log x} \int_1^{\sqrt{x}} A(\tfrac{x}{t})\, d\psi(t) + \frac{1}{x \log x} \int_1^{\sqrt{x}} \psi(\tfrac{x}{t})\, dA(t)$$

$$< \frac{1}{x \log x} \left[\int_1^{\sqrt{x}} (\beta + \varepsilon) \frac{x}{t}\, d\psi(t) + \int_1^{\sqrt{x}} (\delta + \varepsilon) \frac{x}{t}\, dA(t) \right]$$

$$= \frac{1}{\log x} \left[(\beta + \varepsilon) \log \sqrt{x} + O(1) + a(\delta + \varepsilon) \log \sqrt{x} + O(1) \right]$$

$$= \frac{\beta}{2} + \frac{a\delta}{2} + \frac{\varepsilon + a\varepsilon}{2} + O(\frac{1}{\log x}) ,$$

Theorem 2 has been used here in carrying out the integration.

Together with (21) this inequality gives

$$\alpha + \frac{\beta + a\delta + \varepsilon + a\varepsilon}{2} > 2a$$

Since this holds for any $\varepsilon > 0$, it follows that

$$(23) \qquad \alpha + \frac{\beta}{2} + \frac{a\delta}{2} \geq 2a$$

Similarly, by letting $x \to \infty$ in such a way that

$$\frac{A(x)}{x} \to \beta ,$$

and bounding the integrals in (21) from below, it can be seen that

$$(24) \qquad \beta + \frac{\alpha}{2} + \frac{a\gamma}{2} \leq 2a .$$

Inequalities (23) and (24) can now be applied to the special case $A(x) = \psi(x)$, where $\alpha = \gamma$, $\beta = \delta$, and $a = 1$ to obtain

$$\gamma + \delta \geq 2$$

and

$$\gamma + \delta \leq 2 .$$

Thus it follows that $\gamma + \delta = 2$. Next multiply (23) by 2, (24) by -4, and add, getting

$$\begin{array}{r} 2\alpha + \beta + a\delta \geq 4a \\ -4\beta - 2\alpha - 2a\gamma \geq -8a \\ \hline -3\beta + a\delta - 2a\gamma \geq -4a \end{array}$$

Using $\gamma + \delta = 2$, this becomes

$$\begin{aligned} -3\beta + a\delta - 2a(2 - \delta) &\geq -4a \\ -3\beta + a\delta - 4a + 2a\delta &\geq -4a \\ -3\beta + 3a\delta &\geq 0, \end{aligned}$$

or finally,

$$\beta \leq a\delta .$$

Similarly, if equation (23) is multiplied by 4 and (24) by -2, there results

$$\begin{aligned} 4\alpha + 2\beta + 2a\delta &\geq 8a \\ -2\beta - \alpha - a\gamma &\geq -4a \end{aligned}$$

Adding,

$$3\alpha + 2a\delta - a\gamma \geq 4a.$$

From the fact that $\gamma + \delta = 2$, one gets

$$\begin{aligned} 3\alpha + 2a(2 - \gamma) - a\gamma &\geq 4a \\ 3\alpha + 4a - 2a\gamma - a\gamma &\geq 4a \\ 3\alpha - 3a\gamma &\geq 0 \\ \alpha &\geq a\gamma . \end{aligned}$$

Combining these inequalities, it follows that

$$a\gamma \leq \alpha \leq \beta \leq a\delta \qquad \text{q.e.d.}$$

As an application of theorem 7, let

$$M(x) = \sum_{n=1}^{x} \mu(n),$$

and let

$$A(x) = \sum_{n=1}^{x} (\mu(n)+1) = M(x) + [x]$$

Since $\mu(n) + 1 \geq 0$, $A(x)$ is a non-decreasing function of x. The corresponding $T(x)$ is

$$\sum_{n=1}^{x} \left\{ M(\tfrac{x}{n}) + [\tfrac{x}{n}] \right\} = 1 + \sum_{n=1}^{x} [\tfrac{x}{n}]$$

$$= x \log x + (2\gamma - 1)\, x + O(\sqrt{x}).$$

The hypotheses of theorem 7 are therefore satisfied with $a = 1$, $b = 2\gamma - 1$, and so

$$\gamma \leq \underline{\lim}_{x\to\infty} \frac{M(x) + [x]}{x} \leq \overline{\lim}_{x\to\infty} \frac{M(x) + [x]}{x} \leq \delta ,$$

or

$$\gamma - 1 \leq \underline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \overline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \delta - 1 .$$

The prime number theorem asserts that $\gamma = \delta = 1$, and hence that

$$\lim_{x\to\infty} \frac{M(x)}{x} = 0 .$$

This gives an alternative derivation of the result of Landau's dissertation, that $\psi(x) \sim x$ implies $M(x) = o(x)$ in an elementary way.

Actually, somewhat sharper inequalities can be obtained by the same method. Since

$$\mu(n) = 0 \quad \text{if} \quad n \equiv 0 \pmod 4).$$

the function $A(x) = M(x) + [x] - [\frac{x}{4}]$ is monotone non-decreasing, and the corresponding $T(x)$ is

$$T(x) = \frac{3}{4}\, x \log x + \frac{3}{4}(2\gamma - 1)x + O(\sqrt{x}).$$

Hence by theorem 7,

$$\frac{3}{4}\gamma \leq \underline{\lim}_{x\to\infty} \frac{M(x) + [x] - [\frac{x}{4}]}{x} \leq \overline{\lim}_{x\to\infty} \frac{M(x) + [x] - [\frac{x}{4}]}{x} \leq \frac{3}{4}\delta,$$

or

$$\frac{3}{4}(\gamma - 1) \leq \underline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \overline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \frac{3}{4}(\delta - 1).$$

By continuing the process the inequalities

$$\frac{6}{\pi^2}(\gamma - 1) \leq \underline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \overline{\lim}_{x\to\infty} \frac{M(x)}{x} \leq \frac{6}{\pi^2}(\delta - 1)$$

will finally result, since the square-free numbers form a set of measure $\frac{6}{\pi^2}$.

Theorem 7 asserts that of all non-decreasing functions $A(x)$ satisfying

$$(25) \qquad T(x) = ax \log x + bx + o\left(\frac{x}{\log x}\right)$$

with a fixed a, the one having the most widely separated limits of indetermination

$$\varliminf_{x\to\infty} \frac{A(x)}{x} \qquad \text{and} \qquad \varlimsup_{x\to\infty} \frac{A(x)}{x},$$

is a $\psi(x)$. Hence in carrying out the deduction of $A(x) \sim ax$ from (25) it is sufficient to work only with this special case. One may therefore follow the Selberg-Erdös line of attack, or more appropriately the modification of it due to E.M. Wright [5]. Thus we have

THEOREM 8. If $A(x)$ is non-decreasing, and

$$T(x) = ax \log x + bx + o\left(\frac{x}{\log x}\right),$$

then $A(x) \sim ax$.

In conclusion it should be mentioned that this elementary method will give Tauberian theorems other than the one worked out in detail. For example, it is not difficult to see that if $A(x)$ is monotone non-decreasing, and if

$$T(x) = ax^2 + bx + o\left(\frac{x}{\log x}\right),$$

then $A(x) \sim \frac{6}{\pi^2} x^2$.

REFERENCES

[1] E. Landau, Handbuch der Lehre von der Verteilung der Primzahlen, Teubner, Leipzig (1909).

[2] E. Landau, Wiener Sitzungsberichte, 120, 973-988 (1911).

[3] H.N. Shapiro, Proceedings of the American Mathematical Society, 1, 346-348 (1950).

[4] A.E. Ingham, Journal of the London Mathematical Society, 20, 171-180 (1945).

[5] E.M. Wright, Duke Math. J., 19, 695-704 (1952).