FUNDAMENTAL THEOREMS
ON THE DISTRIBUTION OF PRIME NUMBERS

by 1264

ROBERT T. CURRY

B. S., Kansas State University, 1966

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY

Manhattan,Kansas

1970

Approved by:

Eckford Cohen

Major Professor

TABLE OF CONTENTS

INTRODUCTION

"I have sometimes thought that the profound mystery which envelops our conceptions relative to prime numbers depends upon the limitations of our faculties in regard to time, which like space may be in its essence poly-dimensional, and that this and such sort of truths would become self evident to a being whose mode of perception is according to superficially as distinguished from our own limitation to linearly extended time."

J. J. Sylvester

The theory of numbers is one of the oldest of the presently "respectable" mathematical disciplines, and the study of prime numbers one of its fundamental topics. While this paper culminates with Viggo Brun's theorem on the convergence of the harmonic series of the twin primes, the major portion is dedicated to a systematic development of the fundamental theorems which lead to this important result.

As might be expected, the paper begins with Euclid's theorem on the number of primes and then proceeds to a study of the harmonic series of primes. The major result here is Euler's theorem which proves that this so-called Euler series is divergent. Also developed are two O-estimates (see Appendix) of its partial sums. One,

$$\sum_{p \leq x} \frac{1}{p} \sim \log\log x,$$

is obtained from Legendre's theorem, $\pi(x) = o(x)$ as $x \to \infty$. (Theorem 6)

The other,

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + C + 0\left(\frac{1}{\log x}\right) \quad \text{for } x \geq 2$$

is derived from Chebyshev's theorem, $\pi(x) = 0 \frac{x}{(\log x)}$. (Theorem 10)

The Legendre and Chebyshev theorems are developed in turn from earlier estimates of the arithmetic function $\pi(x)$, which represents the number of positive primes $p \leq x$ where x is some real number.

Analogously, Brun's theorem is obtained from an estimate of the function

$$T * (x) < Cx \left(\frac{\log \log x}{\log x}\right)^2 \quad \text{for } x > 3$$

where $T * (x)$ represents the number of twin primes not exceeding x, and C is a positive constant.

It will be noted that a theorem concerning the number of twin primes is conspicuously absent, since it has yet to be shown if their number is infinite. (According to Landau, "One would certainly place one's bet on a 'yes' answer").

ELEMENTARY METHODS

In this section we look at several proofs of Euclid's theorem dealing with the fact that there are an infinite number of primes. A second theorem -- usually called Euler's theorem -- demonstrates that the harmonic series of the prime numbers is divergent, which is a much stronger result, and proves Euclid's theorem as a corollary.

Let us consider first the simplest and earliest proof of Euclid's theorem: Euclid's own.

<u>THEOREM 1.a</u>: There exist an infinite number of primes.

<u>Proof</u>: Suppose there exist only k primes

$$2,3,5, \ldots ,p_k,$$

and consider the number

$$(2 \cdot 3 \cdot 5 \cdot \ldots \cdot p_k) + 1.$$

This number is certainly not divisible by any of the k primes, for such a supposition would lead immediately to the conclusion that 1 was also divisible by that prime. Hence the number is either itself a prime or is divisible by some prime which is larger that p_k -- either of which results lead to a contradiction of the original assumption, thus proving the theorem.

The second proof -- due to Polya -- depends on the fact that the Fermat numbers are relatively prime in pairs. The Fermat numbers are, of course, numbers of the form

$$F_n = 2^{2n} + 1.$$

The following preliminary is necessary.

LEMMA 1: No two distinct Fermat numbers have a common divisor greater than 1.

Proof: Consider F_n and F_{n+k}, where $k > 0$. Suppose the following is true.

$$m|F_n \quad \text{and} \quad m|F_{n+k}$$

Then if we let $x = 2^{2n}$, we see that

$$\frac{F_{n+k} - 2}{F_n} = \frac{2^{2n+k} - 1}{2^{2n} + 1}$$

$$= \frac{(2^{2n})^{2k} - 1}{2^{2n} + 1}$$

$$= \frac{x^{2k} - 1}{x + 1}$$

$$= x^{2k} - 1 - x^{2k-2} + \ldots -1$$

Therefore $F_n|(F_{n+k} - 2)$ and hence

$$m|F_{n+k} \quad \text{and} \quad m|(F_{n+k} - 2)$$

which implies that $m|2$. However, since F_{n+k} is odd, it follows that $m - 1$.

The proof of Euclid's theorem now follows very easily.

THEOREM 1.b: There exist an infinite number of primes.

Proof: Since each of the numbers $F_1, F_2, \ldots, F_n$ is divisible by an odd prime (perhaps the number itself) which does not divide any of the others, it follows that there are at least n odd primes which do not exceed F_n. Thus, since the result holds for all n, the number of primes must be infinite.

Euclid's theorem can be proved in a stronger form by noting that every integer is representable as the product of a square and a square-free number.

Definition: An integer is said to be square-free if the multiplicities of the primes in its standard factorization dc not exceed one.

LEMMA 2: Every integer $n \geq 1$ can be expressed as the product of a square and a square-free number.

Proof: Recalling the fundamental theorem of arithmetic, we can represent n as follows:

$$n = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}$$

Now every exponent is either of the form 2k or 2k + 1. Thus, (reindexing if necessary)

$$n = p_1^{2k_1} p_2^{2k_2} \cdots p_j^{2k_j} p_{j+1}^{2k_{j+1}+1} \cdots p_n^{2k_n+1}$$

$$= p_1^{2k_1} \cdots p_n^{2k_n} p_{j+1} \cdots p_n$$

$$= (p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n})^2 (p_{j+1} p_{j+2} \cdots p_n)$$

THEOREM 2: $\pi(n) \geq A \log n$, where $\pi(n)$ represents the number of primes not exceeding n, and where A is a positive constant.

Proof: Let $p_1, p_2, \ldots, p_r$ be the primes less than or equal to n, i.e., $r = \pi(n)$. By lemma 2 we can write

$$n = (n_1)^2 m$$

where m is square-free. Since the square-free integers are composed of primes which appear only to the first power, the number of square-free integers $\leq n$ does not exceed

$$\binom{r}{0} + \binom{r}{1} + \cdots + \binom{r}{r} = 2^r .$$

Since $m \geq 1$, it also follows that $n_1 \leq \sqrt{n}$, and thus there are no more than $\sqrt{n}$ values that n_1 can assume. Therefore

$$n \leq 2^r \sqrt{n}$$

or

$$r \geq \frac{\log n}{2 \log 2}$$

Finally

$$\pi(n) \geq A \log n$$

which is the desired result.

It may be noted at this point that Euclid's theorem follows easily by letting n increase without bound in the final expression above.

The following lemma, often called Legendre's identity, will be of use in later chapters, and will be used here to prove an interesting result.

LEMMA 3: (Legendre's Identity) Let $e(p,n)$ be the uniquely determined non-negative integer such that

$$p^e \mid n! \qquad \text{and} \qquad p^{e+1} \nmid n!$$

where p is a prime and n is any natural number. Then

$$e(p,n) = \sum_{a>0} \left[\frac{n}{p^a}\right]$$

where $[t]$ is the greatest integer $\leq t$ for all real t and the summation extends over all natural numbers a.

Proofs of this lemma can be found in most texts on elementary number theory and a detailed treatment can be found in [1].

Employing this lemma, we note the following:

$$n! = \prod_{p \mid n!} p^{e(p,n)} = \prod_{p \le n} p^{e(p,n)} \qquad \text{(for } n > 1\text{)}$$

Since $[x] \le x$, we have

$$e(p,n) = \sum_{a=1}^{\infty} \left[\frac{n}{p^a}\right] \le n \sum_{a=1}^{\infty} \frac{1}{p^a} = \frac{n}{p-1}$$

and thus

$$\sqrt[n]{n!} \le \prod_{p \le n} p^{\frac{1}{p-1}}$$

It is well known that

$$\lim_{n\to\infty} \sqrt[n]{n!} = \infty$$

hence

$$\lim_{n\to\infty} \prod_{p \le n} p^{\frac{1}{p-1}} = \infty$$

Since this product diverges to infinity, its logarithm must also, thus

$$\lim_{n\to\infty} \sum_{p\le n} \frac{\log p}{p-1} = \infty$$

But since $p \ge 2$

$$\sum_{p\le n} \frac{\log p}{p-1} \le 2 \sum_{p\le n} \frac{\log p}{p}$$

Therefore

$$\lim_{n\to\infty} \sum_{p\le n} \frac{\log p}{p} = \infty$$

We now conclude with the following result, which we state as a theorem.

THEOREM 3: $\sum_p \frac{\log p}{p} = \infty$, where the summation is taken over all prime numbers.

We now present two similar elementary proofs for Euler's theorem which were advanced by Erdos.

THEOREM 4: The series

$$\sum_p \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \dots$$

is divergent

Proof: (I): Suppose the series converges. Then we can select j so that the remainder of the series after j terms is less than $\frac{1}{2}$, i.e.,

$$\frac{1}{p_{j+1}} + \frac{1}{p_{j+2}} + \dots < \frac{1}{2}$$

Now let us define the following quantity.

$N(x)$ = number of $n \le x$ which are not divisible by any of $p_{j+1}, p_{j+2}, \dots$, where x is any positive integer.

Hence, since the number of $n \le x$ which are divisible by a prime p is $\left[\frac{x}{p}\right]$, the quantity $(x - N(x))$, which is the number of $n \le x$ divisible by at least one of the primes larger than p_j, has the following property.

$$x - N(x) \le \left[\frac{x}{p_{j+1}}\right] + \left[\frac{x}{p_{j+2}}\right] + \cdots$$

$$\le \frac{x}{p_{j+1}} + \frac{x}{p_{j+2}} + \cdots < \frac{x}{2}$$

or $N(x) > \frac{x}{2}$.

Claim: $N(x) \le 2^j \sqrt{x}$

Proof: Let $2,3,4,\dots,p_j$ be the first j primes, and let n be any number not exceeding x which is not divisible by $p > p_j$. Then by Lemma 2 we can express n in the following manner:

$$n = (n_1)^2 m$$

where m is a square-free integer. In other words

$$m = 2^{a_1} 3^{a_2} 5^{a_3} \ldots p_j^{a_j}$$

where all of the a_i are either 0 or 1. Then certainly there are at most 2^j numbers m. Furthermore,

$$n_1 \leq \sqrt{n} \leq \sqrt{x}$$

and so there do not exceed x values of n_1. Therefore

$$N(x) \leq 2^j \sqrt{x}$$

Combining this with the previous result for N(x), we obtain

$$x < 2^{2j+2}.$$

But this limitation on x is certainly contradictory. The result follows. The second proof is essentially similar.

Proof: (II): The following inequality is certainly true.

$$\sum_p \frac{1}{p^2} < \frac{1}{4} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \ldots$$

$$= \frac{1}{4} + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right) + \ldots$$

$$= \frac{3}{4}$$

Now if we assume that the Euler series converges, then we can find j such that the remainder of the series after j terms is less than $\frac{1}{8}$, i.e.

$$\sum_{p>p_j} \frac{1}{p} < \frac{1}{8}$$

Then if we consider the number of square-free integers not exceeding n which are composed only of the primes $2,3,5,\ldots,p_j$, we can establish the following inequality.

$$2^j \geq n - \sum_{p\leq p_j} \left[\frac{n}{p^2}\right] - \sum_{p>p_j} \left[\frac{n}{p}\right] \geq n - \sum_{p\leq p_j} \frac{n}{p^2} - \sum_{p>p_j} \frac{n}{p}$$

$$> n - \frac{3}{4} n - \frac{n}{8}$$

$$= \frac{n}{8}$$

But this implies $n < 2^{j+3}$ which is again a contradictory restriction, establishing the desired result.

THE SIEVE OF ERATOSTHENES AND LEGENDRE'S THEOREM

In this section we consider certain approximations to the function $\pi(x)$, which represents the number of primes not exceeding x.

The method developed will also be used to sharpen the Euler Theorem. Eratosthenes (circa 300 b.c.) found a method of detecting the primes in the sequence of natural numbers. It depended on the fact that any proper divisor of a number must precede the number, but cannot be 1. Thus 2 is the first prime. After 2, no multiple of 2 can be prime, so all even numbers are stricken out. Thus three is the next prime. After all multiples of 3 are removed, 5 is left as the next prime, and so on. This method is known as the sieve of Eratosthenes. Even more can be concluded from this procedure. If a number n is composite, at least one of its divisors must be less than or equal to the square root of n. Then if we have found all primes not exceeding $\sqrt{n}$ and deleted their multiples, all the remaining numbers up to n must be prime, i.e., the primes between $\sqrt{n}$ and n are found by eliminating the multiples of primes less than or equal to $\sqrt{n}$.

Direct application of the sieve of Eratosthenes to problems of prime number distribution has not been successful so far, bur various analytic refinements have been used with success and have yielded interesting results. In this section, sieving methods will be used to prove Legendre's theorem about the density of the prime numbers. Later they will be utilized to prove Viggo Brun's very ingenious theorem concerning the convergence of the series of the reciprocals of the twin primes.

We first establish the following theorem -- a special case of which will yield the sieve we need.

<u>THEOREM 5</u>: Let S be a set of N distinct elements, and let $S_1, \ldots, S_r$ be

arbitrary subsets of S containing $N_1, \ldots, N_r$ elements, respectively. For $1 \le i < j < \ldots < k \le r$, let $N_{ij\ldots k}$ be the number of elements in $S_{ij\ldots k}$. Then the number of elements not in <u>any</u> of $S_1, \ldots, S_r$ is

$$K = N - \sum_{1 \le i \le r} N_i + \sum_{1 \le i < j \le r} N_{ij} - \ldots + (-1)^r N_{12\ldots r}.$$

Proof: Let a certain element s of S belong to exactly m of the sets $S_1, \ldots, S_r$. If $m = 0$, i.e., if s does not belong to any of the sets S_i, then s contributes 1 to K since it is counted once in N and nowhere else.

Suppose $0 < m \le r$. Then s is counted once, or $\binom{m}{0}$ times, in N, $\binom{m}{1}$ times in the terms N_i, $\binom{m}{2}$ times in the terms N_{ij}, etc. Hence the total contribution to K by s is

$$\binom{m}{0} - \binom{m}{1} + \binom{m}{2} - \ldots + (-1)^m \binom{m}{m}$$

$$= \binom{m}{0} 1^m - \binom{m}{1} 1^{m-1} 1 + \binom{m}{2} 1^{m-2} 1^2 - \ldots + (-1)^m \binom{m}{m} 1^m$$

$$= (1 - 1)^m = 0.$$

Thus an element of S contributes 1 to K only if it is not in any of the S_i, while all other elements contribute nothing. Thus the expression for K satisfies the requirements.

Remark: To obtain the product formula for the Euler ϕ-function, that is, the formula which gives the number of integers not exceeding an integer n and relatively prime to n, take S to be the set of integers $1, \ldots, n$, and for $1 \le k \le r$, take S_k to be the set of elements of S which are divisible by the prime p_k, where

$$n = p_1^{a_1} p_2^{a_2} \ldots p_r^{a_r}.$$

If $d|n$ the number of integers not exceeding n which are divisible by d is $\frac{n}{d}$. Hence

$$\phi(n) = n - \sum_{1 \le i \le r} \frac{n}{p_i} + \sum_{1 \le i < j \le r} \frac{n}{p_i p_j} - \cdots$$

$$= n \prod_{p|n} (1 - \frac{1}{p}).$$

A result which will be most useful is the following. If $p(x,r)$ is the number of numbers not exceeding an integer x and which are not divisible by any of the first r primes $p_1, \ldots, p_r$, then

$$p(x,r) = [x] - \sum_{1 \le i \le r} \left[\frac{x}{p_i}\right] + \sum_{1 \le i < j \le r} \left[\frac{x}{p_i p_j}\right] - \cdots$$

It is these formulations of the sieve, plus the following lemma, which will be used to prove Legendre's theorem.

LEMMA 4: $\prod_{p \le x} (1 - \frac{1}{p}) < \frac{1}{\ln x}$

Proof: Let $x \ge 2$ (x integral). Then

$$\prod_{p \le x} (1 - \frac{1}{p})^{-1} = \prod_{p \le x} (1 + \frac{1}{p} + \frac{1}{p^2} + \ldots)$$

But, in accordance with the fundamental theorem of arithmetic, this is simply the sum of the reciprocals of all integers whose prime factors do not exceed x. Then certainly

$$\prod_{p \le x} (1 - \frac{1}{p})^{-1} \ge \sum_{k=1}^{x} \frac{1}{k} .$$

By regarding the sum as the area bounded by a step function, we can establish the following inequality:

$$\sum_{k=1}^{x} \frac{1}{k} > \int_{1}^{x} \frac{du}{u} = \ln x$$

Thus
$$\prod_{p \le x} \left(1 - \frac{1}{p}\right)^{-1} > \ln x,$$

or
$$\prod_{p \le x} \left(1 - \frac{1}{p}\right) < \frac{1}{\ln x} .$$

We now have the necessary means to prove Legendre's Theorem. (The theorem can be proved without using the sieve process, however the sieve is more appropriate here due to the fact that a refinement of the sieving method will be used in the proof of Brun's Theorem in a later chapter.)

<u>THEOREM 6</u>: The prime numbers have zero density, i.e.,

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

or, equivalently, $\pi(x) = o(x)$.

Proof: Recalling the function $p(x,r)$ to be the number of numbers not exceeding x which are not divisible by $p_1,\ldots,p_r$, the inequality

$$\pi(x) \le p(x,r) + r$$

holds since all primes not exceeding x are counted in $p(x,r)$ with the exception of the r primes $p_1,\ldots,p_r$. Moreover,

$$p(x,r) = [x] - \sum_{1 \le i \le r} \left[\frac{x}{p_i}\right] + \ldots .$$

Then, since the number of square brackets involved is $1 + \binom{r}{1} + \binom{r}{2} + \binom{r}{3} + \ldots + \binom{r}{r} = (1 + 1)^r = 2^r$, and since the error involved in the removal of a square bracket is less than one, we can conclude that

$$p(x,r) \leq x - \sum_{1 \leq i \leq r} \frac{x}{p_i} + \ldots + 2^r$$

$$= x \prod_{i=1}^{r} (1 - \frac{1}{p_i}) + 2^r$$

or,

$$\pi(x) \leq x \prod_{p \leq p_r} (1 - \frac{1}{p}) + 2^r + r .$$

Now by the previous lemma, for any $\epsilon > 0$ we can choose $r = r(\epsilon)$ such that

$$\prod_{p \leq p_r} (1 - \frac{1}{p}) < \frac{1}{2} \epsilon$$

and $\pi(x) < \frac{1}{2} \epsilon x + 2^r + r < x\ (\frac{\epsilon}{2} + 2^r + r) < x\ (\frac{\epsilon}{2} + \frac{\epsilon}{2}) = \epsilon x$ for sufficiently large x. In particular, since this holds for any positive ϵ, it follows that

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

This theorem can be sharpened somewhat if Lemma 4 is used. We can get - by similar methods -

THEOREM 7: $\pi(x) = 0\left(\frac{x}{\log \log x}\right)$.

Proof: Recall

$$\pi(x) \leq r + 2^r + x \prod_{p \leq p_r} \left(1 - \frac{1}{p}\right)$$

$$\leq 2^{r+1} + x \prod_{p \leq p_r} \left(1 - \frac{1}{p}\right)$$

Then by Lemma 4,

$$\pi(x) \leq 2^{r+1} + \frac{x}{\log p_r}$$

$$< 2^{r+1} + \frac{x}{\log r}$$

Now since r is an independent variable, suppose we let $r = \log x$. We get

$$\pi(x) < \frac{x}{\log \log x} + 2 \cdot 2^{\log x}$$

$$= \frac{x}{\log \log x} + O(x^{\log 2})$$

(since $2^{\log x} = x^{\log 2}$)

The last term above is $O(x^{1-\epsilon})$ for some $\epsilon > 0$, and thus is certainly $o\left(\frac{x}{\log \log x}\right)$.

Hence
$$\pi(x) = O\left(\frac{x}{\log \log x}\right) + o\left(\frac{x}{\log \log x}\right)$$

$$= O\left(\frac{x}{\log \log x}\right)$$

Euler's theorem can now be strengthened by finding an asymptotic relationship for the partial sums of the Euler series, which will be shown to be

$$\sum_{p \leq x} \frac{1}{p} \sim \log \log x$$

We begin with the following theorem.

<u>THEOREM 8</u>: $\sum_{p \leq x} \frac{\log p}{p} \sim \log x$

or, more specifically,

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O\left(\frac{\log x}{\log \log x}\right)$$

<u>Proof</u>: It is known for real x and for all natural numbers k that

$$\left[\frac{[x]}{k}\right] = \left[\frac{x}{k}\right]$$

A proof of this can be found in [7] p.76 or in most other elementary texts. Thus by Legendre's identity, we know that

$$[x]! = \prod_{p \le x} p^{\left[\frac{x}{p}\right] + \left[\frac{x}{p^2}\right] + \cdots}$$

and so

$$\log [x]! = \sum_{p\le x} \left[\frac{x}{p}\right] \log p + \sum_{p\le x} \left(\left[\frac{x}{p^2}\right] + \left[\frac{x}{p^3}\right] + \cdots\right) \log p$$

Now

$$\sum_{p\le x} \left[\frac{x}{p}\right] \log p \le \sum_{p\le x} \left(\frac{x}{p}\right) \log p$$

and likewise

$$\begin{aligned}\sum_{p\le x} \left[\frac{x}{p}\right] \log p &\ge \sum_{p\le x} \left(\frac{x}{p} - 1\right) \log p \\ &= \sum_{p\le x} \frac{x}{p} \log p - \sum_{p\le x} \log p \\ &\ge \sum_{p\le x} \frac{x}{p} \log p - \log x \sum_{p\le x} 1 .\end{aligned}$$

Moreover

$$0 \le \sum_{p\le x} \left(\left[\frac{x}{p^2}\right] + \left[\frac{x}{p^3}\right] + \cdots\right) \log p \le \sum_{p\le x} \left(\frac{x}{p^2} + \frac{x}{p^3} + \cdots\right) \log p.$$

Thus

$$\log [x]! = \sum_{p\le x} \frac{x}{p} \log p + O(\pi(x) \log x) + O(x \sum_{p\le x} \left(\frac{1}{p^2} + \frac{1}{p^3} + \cdots\right)\log p)$$

$$= x \sum_{p\le x} \frac{\log p}{p} + O(\pi(x)\ \log x) + O\left(x \sum_{p\le x} \frac{\log p}{p(p-1)}\right)$$

Now $\sum_{x} \frac{\log x}{x(x-1)} = O(1)$ since it is a convergent series, so

$$\log [x]! = x \sum_{p\le x} \frac{\log p}{p} + O\left(\frac{x \log x}{\log \log x}\right) + O(x)$$

where the results of theorem 6 have been applied to the second term. On the other hand, it is shown in [4], §16 that

$$\log [x]! = x \log x + O(x)$$

Combining these two expressions, we obtain

$$x \sum_{p\le x} \frac{\log p}{p} + O\left(\frac{x \log x}{\log \log x}\right) = x \log x + O(x)$$

or

$$\sum_{p\le x} \frac{\log p}{p} + O\left(\frac{\log x}{\log \log x}\right) = \log x + O(1)$$

Finally

$$\sum_{p\le x} \frac{\log p}{p} = \log x + O\left(\frac{\log x}{\log \log x}\right)$$

In order to prove the next theorem, we will need the following partial summation lemma, a proof of which may be found in [5] p.103.

<u>LEMMA 5</u>: Let $\lambda_1, \lambda_2, \ldots$ be a non-decreasing, unbounded sequence, and $c_1, c_2, \ldots$ be an arbitrary sequence of real or complex numbers. If $f(x)$ is a function having a continuous derivative for $x \ge \lambda_1$, and if we define

$$C(x) = \sum_{\substack{n \\ (\lambda_n \le x)}} c_n ,$$

then for $x \ge \lambda_1$

$$\sum_{\substack{n \\ (\lambda_n \le x)}} c_n \; f(\lambda_n) = C(x) \; f(x) - \int_{\lambda_1}^{x} C(t) \; f'(t) \; dt.$$

Having these results, we can now prove the following theorem:

<u>THEOREM 9</u>:
$$\sum_{p \le x} \frac{1}{p} \sim \log \log x$$

<u>Proof</u>:
$$\sum_{p \le x} \frac{1}{p} = \frac{1}{2} + \sum_{2 < p \le x} \left(\frac{\log p}{p} \cdot \frac{1}{\log p} \right)$$

If we utilize the lemma by letting $\lambda_n = p_n$, $c_n = \dfrac{\log p_n}{p_n}$, and $f(t) = \dfrac{1}{\log t}$,

we obtain

$$\sum_{p \le x} \frac{1}{p} = \frac{1}{\log x} \sum_{2 < p \le x} \frac{\log p}{p} - \int_3^x \left(\sum_{p \le t} \frac{\log p}{p} \right) \frac{-dt}{t(\log t)^2} \cdot + \frac{1}{2}$$

Then by Theorem 6,

$$\sum_{p \le x} \frac{1}{p} = \frac{1}{\log x} \log x + O\left(\frac{\log x}{\log \log x}\right) + \int_3^x \log t + O\left(\frac{\log t}{\log \log t}\right) \frac{dt}{t(\log t)^2} + \frac{1}{2}$$

$$= O(1) + \int_3^x \frac{dt}{t \log t} + \int_3^x O\left(\frac{1}{t \log t \log \log t}\right) dt$$

$$= \log \log x + O(1) + \int_3^x O\left(\frac{1}{t \log t \log \log t}\right) dt$$

But

$$\left| \int_3^x O\left(\frac{1}{t \log t \log \log t}\right) dt \right| < K \int_3^x \frac{dt}{t \log t \log \log t}$$

By definition of the O-symbol. Thus

$$\sum_{p \le x} \frac{1}{p} = \log \log x + O(\log \log \log x)$$

and the proof is complete.

We conclude this section by noting that Theorems 6 and 7 both yield results that are more than sufficient to prove Euler's theorem.

Chebyshev's Theorem and a Refinement of $\sum_{p \leqq x} \frac{1}{p}$

In order to prove Chebyshev's theorem -- the strongest result concerning $\pi(x)$with which we shall deal -- the following lemmas are necessary.

LEMMA 6: $[x] - 2\left[\frac{x}{2}\right] \le 1$ for all real x.

Proof: We consider two cases:

A. $x = 2n + a$, n an integer, $0 \le a < 1$

$$[x] = 2n$$

$$\left[\frac{x}{2}\right] = \left[n + \frac{a}{2}\right] \qquad 0 \le \frac{a}{2} < \frac{1}{2}$$

$$= [n]$$

$$2\left[\frac{x}{2}\right] = 2n$$

$$[x] - 2\left[\frac{x}{2}\right] = 0$$

B. $x = 2n + 1 + a$, n an integer, $0 \le a < 1$

$$[x] = 2n + 1$$

$$\left[\frac{x}{2}\right] = \left[n + \frac{1}{2} + \frac{a}{2}\right], \qquad 0 \le \frac{a}{2} < \frac{1}{2}$$

$$= n$$

$$2\left[\frac{x}{2}\right] = 2n$$

$$[x] - 2\left[\frac{x}{2}\right] = 1$$

The proof is complete.

LEMMA 7: $\binom{2n}{n} \ge 2^n$, n a positive integer.

Proof: $$\binom{2n}{n} = \frac{(n+1)\ (n+2)\ (n+3)\ \ldots\ (n+n)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot n}$$

$$= \prod_{a=1} \frac{n+a}{a}$$

$$\geq \prod_{a=1} 2 = 2^n$$

LEMMA 8: $\binom{2n}{n} \leq 2^{2n}$, n a positive integer.

Proof: $\binom{2n}{n} \leq \sum_{a=0}^{2n} \binom{2n}{a} = (1+1)^{2n} = 2^{2n}$

THEOREM 10: (Chebyshev's Theorem) There exist positive constants c_1, c_2, such that, for $x \geqq 2$,

$$c_1 < \frac{\pi(x)}{\left(\frac{x}{\log x}\right)} < c_2$$

Proof: (Note: The c_i's contained in the proof are suitably chosen positive constants.) Let $n \geq 2$ be an integer. For each $p \leq 2n$ there exists a unique integer r_p such that

$$p^{r_p} \leq 2n < p^{r_p + 1} .$$

It is evident that

$$\prod_{n<p\leq 2n} p \;\Big|\; \frac{(2n)!}{n!n!}$$

since any prime between n and 2n must be a factor of (2n)! but cannot be a factor of n!

Now by Legendre's Identity, the exponent of the highest integral power of p that divides (2n)! is

$$\sum_{m=1}^{r_p} \left[\frac{2n}{p^m}\right]$$

while the corresponding exponent for $(n!)^2$ is

$$2\sum_{m=1}^{r_p} \left[\frac{n}{p^m}\right] .$$

Hence the highest power of p which divides $\binom{2n}{n} = \frac{(2n)!}{(n!)^2}$ is

$$\sum_{m=1}^{r_p} \left[\frac{2n}{p^m}\right] - \sum_{m=1}^{r_p} \left[\frac{n}{p^m}\right] = \sum_{m=1}^{r_p} \left[\frac{2n}{p^m}\right] - 2\left[\frac{n}{p^m}\right]$$

But by Lemma 6,

$$\sum_{m=1}^{r_p} \left[\frac{2n}{p^m}\right] - 2\left[\frac{n}{p^m}\right] \le \sum_{m=1}^{r_p} 1 = r_p$$

and, therefore

$$\binom{2n}{n} \Bigm| \prod_{p \le 2n} p^{r_p} .$$

We can now state the following;

$$\prod_{n<p\le 2n} p \le \binom{2n}{n} \le \prod_{p \le 2n} p^{r_p}$$

But it is evident that

$$n^{\{\pi(2n)-\pi(n)\}} \le \prod_{n<p\le 2n} p$$

and that $\prod_{p \le 2n} p^{r_p} \le (2n)^{\pi(2n)}$ (recalling that $p^{r_p} \le 2n$).

It then follows that

$$(\pi(2n) - \pi(n)) \log n \le \log \binom{2n}{n} \le \pi(2n) \log 2n,$$

and we can now apply the results of lemmas 7 and 8.

Thus $\{\pi(2n) - \pi(n)\} \log n \le 2n \log 2$,

which implies $\pi(2n) - \pi(n) \le c_3 \frac{n}{\log n}$. (1)

Likewise, $\pi(2n) \log 2n \ge n \log 2$

or $$\pi(2n) > c_4 \frac{n}{\log n} \tag{2}$$

Letting $x \geq 4$ and applying (2), we get

$$\pi(x) \geq \pi\left(2\left[\frac{x}{2}\right]\right) > c_4 \frac{\left[\frac{x}{2}\right]}{\log\left[\frac{x}{2}\right]} > c_5 \frac{x}{\log x}$$

But since $\pi(x) \geq 1$ for $2 \leq x < 4$, the result is also true for $x \geq 2$. Thus

$$\frac{\pi(x)}{\left(\frac{x}{\log x}\right)} > c_1 \quad .$$

(The constants c_i are positive. Aside from that, their values are of no concern in this paper.)

Letting $y \geq 4$ and applying (1), we get

$$\begin{aligned}
\pi(y) - \pi\left(\frac{y}{2}\right) &= \pi(y) - \pi\left(\left[\frac{y}{2}\right]\right) \\
&\leq \pi\left(2\left[\frac{y}{2}\right]\right) + 1 - \pi\left(\left[\frac{y}{2}\right]\right) \\
&< c_3 \frac{\left[\frac{y}{2}\right]}{\log\left[\frac{y}{2}\right]} + 1 < c_6 \frac{y}{\log y}
\end{aligned}$$

Thus, for $y \geq 2$, $\pi(y) - \pi\left(\frac{y}{2}\right) < c_7 \frac{y}{\log y}$.

Noting the obvious fact that $\pi\left(\frac{y}{2}\right) \leq \frac{y}{2}$, we get

$$\pi(y) \log y - \pi\left(\frac{y}{2}\right) \log y \leq \pi(y) \log y - \pi\left(\frac{y}{2}\right) \log\left(\frac{y}{2}\right)$$

and

$$\begin{aligned}
\pi(y) \log y - \pi\left(\frac{y}{2}\right) \log\left(\frac{y}{2}\right) &= \left\{\pi(y) - \pi\left(\frac{y}{2}\right)\right\} \log y + \pi\left(\frac{y}{2}\right) \log 2 \\
&< c_j \frac{y}{\log y} \cdot \log y + \frac{y}{2} \\
&< c_8\, y
\end{aligned}$$

If we now let $y = \frac{x}{2^m}$ · , where $m \geq 0$ and $2^m \leq \frac{x}{2}$, certainly $y \geq 2$ is satisfied. Furthermore

$$\pi\left(\frac{x}{2^m}\right) \log \frac{x}{2^m} - \pi\left(\frac{x}{2^{m+1}}\right) \log \frac{x}{2^{m+1}} < c_8 \frac{x}{2^m} .$$

Summing over all permissible values of m, the left side of the inequality "telescopes" in such a manner that only the terms

$$\pi(x) \log x - \pi\left(\frac{x}{2^{\mu+1}}\right) \log \frac{x}{2^{\mu+1}}$$

remain (where $2^{\mu} \leq \frac{x}{2} < 2^{\mu+1}$), while on the right side we obtain a term of the form x times a convergent series. i.e.,

$$\pi(x) \log x - \pi\left(\frac{x}{2^{\mu+1}}\right) \log \frac{x}{2^{\mu+1}} < c_2 x$$

But $\frac{x}{2^{\mu+1}} < 2$, so $\pi\left(\frac{x}{2^{\mu+1}}\right) = 0$.

Therefore $\pi(x) \log x < c_2 x$.

Hence we conclude

$$c_1 < \frac{\pi(x)}{\left(\frac{x}{\log x}\right)} < c_2 ,$$

and the proof is complete.

The results of this theorem allow us to improve on the results of theorems 6 and 7 in a straightforward and simple manner. These results are due to Mertens.

THEOREM 9:

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1).$$

Proof: By re-examining the proof of theorem 6, we note, (in the first expression for log n!) an error term that is $O(\pi(n) \log n)$. By the theorem just proven, this can now be replaced by $O(n)$ and the result of this theorem follows immediately.

THEOREM 10: There exists a constant C such that

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + C + O\left(\frac{1}{\log x}\right) .$$

Proof: By applying the results of theorem 9 to the proof of theorem 7, we get the following:

$$\sum_{p \leq x} \frac{1}{p} = \frac{1}{\log x} \left(\log x + O(1)\right) + \int_2^x \log t \frac{dt}{t(\log t)^2}$$

$$+ \int_2^x \left(\sum_{p \leq t} \frac{\log p}{p} - \log t\right) \frac{dt}{t(\log t)^2}$$

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

$$+ \int_2^\infty \left(\sum_{p \leq t} \frac{\log p}{p} - \log t\right) \frac{dt}{t(\log t)^2} - \int_x^\infty \frac{O(1)\, dt}{t(\log t)^2}$$

where the first integral is convergent and the second is $O\left(\frac{1}{\log x}\right)$. Thus the result follows.

In the next sections we will see a very interesting analogue to these theorems.

BRUN'S THEOREM

In earlier sections we proved that the primes occurred frequently enough that their harmonic series diverged, while in the last sections it was shown by various methods that they are scarce enough that their density is zero. In this section we prove a theorem due to the Norwegian Mathematician Viggo Brun which is analagous to Chebyshev and Euler's theorems except that it deals not with all the primes, but with the twin primes - that is pairs of integers of the form n, n + 2, such that both n and n + 2 are prime. It is not known whether the number of such primes is finite or infinite, but the following theorem can be established.

<u>THEOREM</u>: <u>The series</u>

$$\frac{1}{3} + \frac{1}{5} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{29} + \frac{1}{31} + \cdots$$

of the reciprocals of the twin primes converges.

Remark: 3,5,7 is the only prime triplet, since one of any three successive odd numbers must be divisible by 3.

A. <u>The sieve</u>.

In proving this theorem, we will develop a refinement of the sieve used earlier. This sieve will exclude numbers n where either n or n + 2 is composite.

Definition: Let $T(x)$ be the number of odd primes $p \le x$ such that $p + 2$ is prime as well.

Definition: Let $U(x;p_1,p_2,\cdots p_r)$ be the number of odd numbers $n \le x$ for which $n(n + 2)$ is not divisible by any of the odd primes $p_1,\cdots,p_r$.

If $\{p_1,\cdots,p_r\}$ is the set of all primes $p_i \le \sqrt{x+2}$, then $U(x;p_1,\cdots,p_r)$ counts only the first members p of the twin primes lying between $\sqrt{x+2}$ and

and $x + 2$. This follows by a simple argument: unless both n and $n + 2$ are prime, $n > \sqrt{x+2}$, then one of them (and hence their product) must have some prime divisor $p \le \sqrt{x+2}$ by the argument in the previous section. Since some twin primes may be found among $p_1, \cdots, p_r$, then the following inequality holds.

$$T(x) \le U(x;p_1,\cdots,p_r) + r.$$

Note: If we take for the sieving process only odd primes $p_i \le y < \sqrt{x+2}$, the inequality still holds. This follows from the fact that $U(x;p_1,\cdots,p_r)$ may now contain some odd numbers which are not the first members of twin primes.

Let us simplify the notation as follows: for $p_1,\cdots,p_r$ all odd primes $p_i \le y$, where $y \le \sqrt{x+2}$, let

$$U(x;p_1,\cdots,p_r) = U(x;y)$$

(We observe for later use that $2r \le y$).

Now let $B(x;p_i p_j \cdots p_k)$ be the number of odd number $n \le x$ for which $n(n+2)$ is divisible by the product $p_i p_j \cdots p_k$. Then

$$U(x;y) = \left[\frac{x+1}{2}\right] - \sum_{1\le i\le r} B(x;p_i) + \sum_{1\le i<j\le r} B(x;p_i p_j)$$

$$- \sum_{1\le i<j<k\le r} B(x;p_i p_j p_k) + \cdots + (-1)^r \; B(x;p_1 \cdots p_r).$$

This is established in a manner exactly analagous to that in section 2 once we note that $\left[\frac{x+1}{2}\right]$ is simply the number of odd numbers $n \le x$.

Suppose we let $\rho^{(f)}$ represent a number which is the product of f different prime numbers. Then

$$U(x;y) = \left[\frac{x+1}{2}\right] + \sum_{f=1}^{r} (-1)^f \sum_{\rho^{(f)}} B(x;\rho^{(f)})$$

where the inner sum runs over all possible products of f of the primes 3,5, $\ldots, p_r$.

The following lemmas are now stated without proof:

LEMMA 4:

$$\sum_{\lambda=0}^{n} (-1)^{\lambda} \binom{f}{\lambda} \quad \begin{aligned} &= 0 \text{ for } m \geq f > 0 \\ &> 0 \text{ for } m < f,\ m \text{ even} \\ &< 0 \text{ for } m < f,\ m \text{ odd.} \end{aligned}$$

LEMMA 5:

Let ρ be an odd number and $\nu(\rho)$ the number of its different prime factors. Then the number $B(x;\rho)$ of odd numbers $n \leq x$ for which $n(n + 2)$ is divisible by ρ is

$$B(x;\rho) = 2^{\nu(\rho)} \quad \frac{x}{2\rho} + \theta \quad , \ |\theta| \leq 1$$

Note: For large ρ, the term $\frac{x}{2\rho}$ will be dominated by the magnitude of θ, so it is advisable to stop the sum in the expression for $U(x;y)$ at some $m < r$ which, for our purposes, we choose to be even. (Landau proved that failure to restrict the summation results in a situation that cannot possibly lead to the desired proof.)

We now obtain

$$T(x) \le r + \left[\frac{x+1}{2}\right] + \sum_{f=1}^{m} (-1)^f \sum_{\rho^{(f)}} B(x;\rho^{(f)})$$

$$= r + \left[\frac{x+1}{2}\right] + \sum_{f=1}^{m} (-1)^f \sum_{\rho^{(f)}} 2^f \left\{\frac{x}{2p} + \theta\right\}$$

$$= r + \left[\frac{x+1}{2}\right] + \frac{x}{2}\sum_{f=1}^{m} (-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} + \theta \sum_{f=1}^{m} (-1)^f \sum_{\rho^{(f)}} 2^f$$

$$= r + \left[\frac{x+1}{2}\right] + \frac{x}{2}\sum_{f=0}^{m} (-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} + \theta \sum_{f=0}^{m} (-1)^f \sum_{\rho^{(f)}} 2^f - x - \theta$$

$$\le r + \frac{x}{2}\sum_{f=0}^{m} (-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} + \sum_{f=0}^{m} \sum_{\rho^{(f)}} 2^f$$

where $\rho^{(0)} = 1$ and $\nu(\rho^{(f)}) = f$

B. First estimates.

Unlike the problem in Chapter 2, this formulation is not in a readily usable form. Thus we wish to derive certain estimates of the terms under the summations which will allow us to complete the proof.

I. Let us consider the third term first. Since $\rho^{(f)}$ is made up of all possible products of f distinct prime factors taken from the first r odd primes,

$$\sum_{f=0}^{m} \sum_{\rho^{(f)}} 2^f = \sum_{f=0}^{m} \binom{r}{f} 2^r .$$

But by induction,

$$\sum_{f=0}^{m} \binom{r}{f} 2^{f} < \sum_{f=0}^{m} (2r)^{f} < 2^{m+1} \sum_{f=0}^{m} r^{f}$$

and

$$2^{m+1} \sum_{f=0}^{m} (r)^{f} = 2^{m+1} \frac{r^{m+1} - 1}{r - 1} = \frac{(2r)^{m+1} - 2^{m+1}}{r - 1} < (2r)^{m+1} .$$

II. $$\sum_{f=0}^{m} (-1)^{f} \sum_{\rho(f)} \frac{2^{f}}{\rho^{(f)}} = \sum_{f=0}^{r} (-1)^{f} \sum_{\rho(f)} \frac{2^{f}}{\rho^{(f)}} - \sum_{f=m+1}^{r} (-1)^{f} \sum_{\rho(f)} \frac{2^{f}}{\rho^{(f)}}$$

By direct multiplication, we see that

$$\sum_{f=0}^{r} (-1)^{f} \sum_{\rho(f)} \frac{2^{f}}{\rho^{(f)}} = \prod_{i=1} \left(1 - \frac{2}{p_i}\right)$$

For convenience, let us define

$$S_f = \sum_{\rho(f)} \frac{2^{f}}{\rho^{(f)}}$$

Therefore

$$T(x) \le r + \frac{x}{2} \prod_{i=1}^{r} \left(1 - \frac{2}{p_i}\right) + \frac{x}{2} \sum_{f=m+1}^{r} (-1)^{f-1} S_f + (2r)^{m+1}.$$

Note: We observe that S_f is nothing more than the f-th elementary symmetric function of the r rational numbers $\frac{2}{3}, \frac{2}{5}, \cdots, \frac{2}{p_r}$. It can be established simply by multiplying that

$$S_1 \cdot S_f \ge (f + 1) S_{f+1} \qquad f = 1,2,\cdots$$

(where $S_{r+1} = S_{r+2} = \cdots = 0$).

Moreover, the following iterative results can be established

$$S_2 \le \frac{S_1^{\,2}}{2},\ S_3 \le \frac{S_1^{\,3}}{3!},\ \cdot\ \cdot\ \cdot\ , S_f \le \frac{S_1^{\,f}}{f!}\ .$$

We also deduce that

$$S_f \ge S_{f+1}$$

provided

$$f + 1 \ge S_1$$

We also note that

$$S_1 = \frac{2}{3} + \frac{2}{5} + \cdot\ \cdot\ \cdot + \frac{2}{p_r} = 2 \sum_{3 \le p \le y} \frac{1}{p}$$

and that S_1 therefore depends on our choice of y. In light of the above conclusions, we pick m such that

$$m + 1 \ge S_1\ .$$

Thus it follows that

$$\sum_{f=m+1}^{r} (-1)^{f-1}\ S_f \le S_{m+1} \le \frac{S_1^{\,m+1}}{(m+1)!}$$

Claim: $$\frac{S_1^{\,m+1}}{(m+1)!} \le \left(\frac{eS_1}{m+1}\right)^{m+1}$$

Proof: $$e^n = 1 + n + \frac{n^2}{2!} + \frac{n^3}{3!} + \cdot\ \cdot\ \cdot + \frac{n^n}{n!} + \cdot\ \cdot\ \cdot > \frac{n^n}{n!}$$

or $$n! < \left(\frac{n}{e}\right)^n$$

Thus the claim is proven.

C. <u>Second estimates</u>.

It is now evident that estimates for the terms

$$\sum_{3\le p\le y} \frac{1}{p} \qquad \text{and} \qquad \prod_{i=1}^{r} \left(1 - \frac{2}{p_i}\right)$$

are needed, since the first will give an estimate for S_1 and the second is the only term in our sieve formulation that remains indefinite.

By Theorem 10, we see that

$$S_1 = 2 \sum_{3\le p\le y} \frac{1}{p} = 2 \log \log y + 0(1).$$

Let us further refine m + 1 so that

$$e^2 S_1 < m + 1 < 9S_1 ,$$

which is certainly in keeping with our earlier bound.

Recalling

$$\sum_{f=m+1}^{r} (-1)^f \; S_1 \le \left(\frac{eS_1}{m+1}\right)^{m+1}$$

it now becomes evident that

$$\sum_{f=m+1}^{r} (-1)^f \; S_1 \le \left(\frac{1}{e}\right)^{m+1} < e^{-e^2 S_1} < e^{-S_1} .$$

Furthermore, by observing the series expansion for e^{-x}, we see that the following relation is true for real x.

$$1 - x \le e^{-x}$$

Therefore

$$\prod_{i=1}^{r} \left(1 - \frac{2}{p_i}\right) < e^{-2 \sum_{i=1}^{r} \frac{1}{p_i}} = e^{-S_1} .$$

Combining these results, and using our earlier observation that $2r \le y$, we conclude that

$$T(x) \le y + x\,e^{-S_1} + y^{9S_1} .$$

D. Choice of the parameter.

Now, for y sufficiently large, by our expression for S_1 we have

$$2 \log\log y - M < S_1 < 3 \log\log y,$$

for some positive constant M. Thus the above expression for T(x) can be replaced by

$$T(x) \le y + e^M \frac{x}{(\log y)^2} + y^{27 \log\log y} .$$

Now suppose we let $y = x^\gamma$ where $0 < \gamma \le \frac{1}{2}$. This implies $y \le \sqrt{x}$ which certainly agrees with our earlier restriction on y, i.e., $y \le \sqrt{x+2}$. Thus

$$T(x) < x^{1/2} + e^M \frac{x}{(\gamma \log x)^2} + x^{27\gamma \log\log x} .$$

It is evident that if we allow γ to be a constant, we will not obtain a convergent bound for T(x), for the third term in the above equation will increase faster than will x. Let us choose

$$\gamma = \frac{1}{30 \log\log x} \qquad \text{for } x \ge 3 ,$$

which is always positive and has $\frac{1}{2}$ as an upper bound. Thus

$$T(x) < x^{1/2} + 900\, e^M\, x \left(\frac{\log\log x}{\log x}\right)^2 + x^{9/10} ,$$

where, for sufficiently large x, the second term will predominate.

Now let $T^*(x)$ be the number of all twin primes not exceeding x. Then certainly

$$T^*(x) \le 2\,T(x).$$

These results can be combined in the following theorem.

THEOREM: There exists a positive constant C such that $T^*(x)$, the number of twin primes not exceeding x, satisfies

$$T^*(x) < Cx\left(\frac{\log\log x}{\log x}\right)^2 ,$$

for $x > 3$.

E. Completion of the proof of Brun's Theorem.

We now proceed to the final step - the convergence of the harmonic series of the twin primes - which follows easily from the preceding theorem. Let

$$S(x) = \sum_{\substack{p \text{ twin prime} \\ p \le x}} \frac{1}{p} .$$

Then

$$S(x) = \sum_{3 \le n \le x} \frac{1}{n}\left(T^*(n) - T^*(n-2)\right) \qquad (n \text{ odd}) .$$

This relation follows easily if we note that $T^*(n) - T^*(n-2)$ is 1 if n is either the first or second member of a pair of twin primes, and 0 for all other odd numbers. Now, therefore,

$$S(x) = \sum_{3 \le n \le x} \frac{1}{n} T^*(n) - \sum_{3 \le n \le x-2} \frac{1}{n+2} T^*(n) ,$$

$$= \sum_{3\le n\le x} \frac{1}{n} T^*(n) - \left[\sum_{3\le n\le x} \frac{1}{n+2} T^*(n) - \frac{T^*(\eta)}{(\eta+2)} \right] ,$$

$$= \sum_{3\le n\le x} T^*(n) \left(\frac{1}{n} - \frac{1}{n+2} \right) + \frac{T^*(\eta)}{(\eta+2)}$$

where $\eta = 2\left[\frac{x-1}{2}\right] + 1$ is the largest odd number which does not exceed x. Thus, by the theorem,

$$S(x) = 2\sum_{3\le n\le x} T^*(n) \left(\frac{1}{n(n+2)}\right) + O\left(\frac{\log \log^2 x}{\log^2 x}\right)$$

$$= O\left(\sum_{3\le n\le x} \frac{(\log \log n)^2}{n(\log n)^2}\right)$$

But it is easily shown by the integral test that

$$\sum_{2\le n<\infty} \frac{1}{n(\log n)^{1+\epsilon}} \quad .$$

is convergent for $\epsilon > 0$, so it follows that

$$\sum_{3\le n<\infty} \frac{(\log \log n)^2}{n(\log n)^2} \quad ,$$

is also.

BIBLIOGRAPHY

1. Cohen, Eckford, Legendre's Identity, to appear.

2. Erdos, P., "Uber Die Reihe $\sum 1/p$", Mathematica, B 7 (1938).

3. Hardy, G.H., and E. M. Wright, An Introduction to the Theory of Numbers, London: Oxford University Press, 1954

4. Landau, Edmund, Elementary Number Theory, New York: Chelsea Publishing Company, 1958.

5. Landau, Edmund, Handbuch der Lehre von der Verteilung der Primzahlen, New York: Chelsea Publishing Company, 1953.

6. Leveque, Wm. J., Topics in Number Theory, Reading, Massachusetts: Addison-Wesley Publishing Co., 1956.

7. Rademacher, Hans., Lectures on Elementary Number Theory, New York: Blaisdell Publishing Co., 1964.

8. Stewart, B. M., Theory of Numbers, New York: The MacMillan Company, 1964.

9. Titchmarsh, E. C., The Theory of the Riemann Zeta-Function, London: Oxford University Press, 1951.

APPENDIX

DEFINITION: Given function f,g, where g is always positive, we say that:

1) $f = O(g)$ if $|f| < Ag$ for some constant A,

2) $f = o(g)$ if $\frac{f}{g} \to 0$,

3) $f \sim g$ if $\frac{f}{g} \to 1$ (read "f is asymptotic to g.")

ACKNOWLEDGEMENT

The writer wishes to express his sincere appreciation to Dr. Eckford Cohen for his patient guidance and supervision given during the preparation of this report.

FUNDAMENTAL THEOREMS
ON THE DISTRIBUTION OF PRIME NUMBERS

by

ROBERT T. CURRY

B. S., Kansas State University, 1966

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1970

ABSTRACT

The purpose of this paper is to present certain fundamental theorems regarding the distribution of prime numbers. The paper begins by proving the existence of an infinite number of primes, and then sharpens this result by proving Euler's theorem of the divergence of the harmonic series of the prime numbers. In the second section, sieve methods are used to show that the primes have density zero, and to give estimates of the order of $\pi(x)$ and $\sum_p 1/p$. Section three deals with Chebychev's theorem, and further refines the estimates of section two. The final section is devoted to the proof of Brun's theorem and uses results and methods developed in earlier sections.